SELF-PACED STUDY AND WORD DIMENSIONALITY IN METACOGNITION

by Ryan Lee, A thesis submitted to the School of Graduate Studies in partial fulfillment

of the requirements for the degree of

Master of Science in Experimental Psychology

Memorial University of Newfoundland

January 2022

St. John's Newfoundland and Labrador

Abstract

The mechanisms underlying metacognitive monitoring-control relationships for emotional information are unclear in the existent literature. We conducted two online studies of university students measuring metacognition for emotional words across three related dependent variables: study time, judgements of learning (JOLs) and recall. We used a repeated measures design with free-recall testing for both experiments. In Experiment 1, words were distinctly categorized based on emotional valence (negative, neutral, and positive). In Experiment 2, words were categorized based on arousal (low, medium, and high). In the first experiment we found that both negative- and positive-valenced words were higher than neutral-valenced words in both JOL rating and recall. In the second experiment we did not find an effect of arousal on JOL ratings or recall. These data suggest that valence and arousal have conceptually distinct roles in metacognition within the mixed-list, three-level design we used. In the second experiment only, we also found that study time was reduced between blocks. Participants were surveyed on their explicit metacognitive beliefs underlying their study habits. Participants tended to explicitly notice the valence manipulation but did not notice the arousal manipulation. Participants also tended to qualify emotional words as being more relatable to themselves, perhaps hinting at how our memory systems engage with emotional information. Further qualifications about how individuals' metacognitive strategies vary based on word quality are discussed.

General Summary

We conducted a memory study to further understand how individuals interact with emotional words. Over the course of two experiments, we manipulated how positive, neutral, or negative in connotation words were (valence) in the first experiment and how exciting, neutral, or mundane words were (arousal) in the second experiment. Emotional words were presented in a mixed-list design where all three levels were presented within the same wordlist. We found that individuals tended to engage with words differently based on their valence than they do based on arousal. Participants believed positive and negative valence words would be remembered better than neutral words and they were correct in that belief; this was not true in the case of arousing words. This suggests that valence is more important than arousal when it comes to remembering emotional information and beliefs about remembering emotional information. Participants were also likely to specify that they engaged with words because of their valences, and one of the highlighted features of positive or negative emotional words was that they were more relatable to participants.

Acknowledgements

I would like to thank everyone who helped with this thesis research. The very foundations of this research are based on my primary supervisor *Dr. Kathleen Hourihan*'s exploration of metamemory for emotional information. We were lucky to have *Dr. Jonathan Fawcett* as a co-supervisor on this project. He was a great help in the development and design of our experiments. I would like to thank our third committee member, *Dr. Darcy Hallett* who taught and provided guidance on statistical matters and provided a fresh set of eyes in the review process. The technical and programming side of things have been managed by *Jason Fawcett* who developed and implemented the web application for the study. The last person I would like to acknowledge is *Alex Randell* who helped code data and was excellent at his job. I am proud of each and every one of these people, this project would not have been possible without them, and I am immensely grateful for their support.

Thank you to everyone who helped with this thesis idea. It has been a long road developing and refining this work. The idea for this research has changed several times over the course of the last two years. This project started out as an in-lab study with a pupillometry component, this proved difficult given the resources and time available. We then moved to a regular in-lab experiment design shortly before the world changed with Covid-19. The experiment was then redesigned to be an online study. This final version is the result of a lot of talented individuals coming together to overcome obstacles that were often out of our hands. I am thankful to those who helped with this project and proud of my peers who pulled through these unprecedented times.

Contents

List of Figures

Introduction

The work conducted in this thesis is intended to explore and understand how the emotionality of information interacts with individuals' metacognitive processes. Broadly, metacognition can be understood as how an individual interacts with their own cognitive processes as was first described and detailed by Flavell (1979) as an early concept. Modern metacognition is generally believed to consist of a relationship between two types of processes: control processes and monitoring processes. Control processes are behaviours that pertain to a cognitive task, whereas monitoring processes refer to the cognition involved when engaged in a task (e.g., Nelson, 1990). It would be reasonable to state that metacognition has been studied for as long as cognition itself, since, realistically, the two are inseparable. However, it has only been in the last half century that metacognition has been considered independently and has been gaining in popularity through recent years as witnessed via the steadily growing body of literature. The growth and development of metacognition as a field is discussed in a seminal paper on metacognition by Georgiahades (2004).

Metamemory is a subset of metacognition dealing with how individuals interact with their own memory processes, and the possible control-monitoring relationships that arise when one is tasked with studying and remembering emotional words. Recent metamemory work related to emotion places emphasis on monitoring and the influence emotion has on judgements of learning (JOLs, e.g., Hourihan & Bursey, 2017; Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010). There is a comparatively smaller body of work focused on control processes for neutral information (Hoffman et al. 2010; Koriat & Ackerman 2010; Tullis & Benjamin 2011) and until very recently, no known work that explores metacognitive control of emotional information (Witherby 2019; Witherby et al. 2021). The study of metamemory and how it relates to learning investigates the learning process in such a way that relates both to an individual's changing cognition and perspective as well as the nature and presentation of information. In short, metamemory work is important for linking learning and teaching.

Monitoring

Monitoring is the metacognitive process of assessment. In metamemory, monitoring is used to gauge the learning process and the memorability of information. A JOL is a measure that is used to assess how well individuals feel they have learned something. In one of the earliest studies of JOLs, Arbuckle and Cuddy (1969) conducted an experiment where they asked participants to evaluate how memorable letter-number and word-word paired associates would be. They found that participants were able to accurately predict which pairs would be remembered, and the accuracy of these predictions was associated with the difficulty of the pairs being assessed, demonstrating that JOLs are shaped in response to the information that is being studied. JOLs require individuals to consciously assess and rate how well they think they will remember information. This judgement is based on one's perception of the memorability of information and thus should reflect how they are interpreting both the information as well as their memory formation. This measure is important as it is a prospective gauge of memory. When used in conjunction with a measure of memory performance such as recall, it can be used to determine whether one's confidence in memory undergoes adjustment in response to the information presented and if that confidence translates to performance through measuring JOL-recall accuracy.

JOL-recall accuracy can be measured in two ways (Dunlosky & Metcalfe, 2009). The first is through calibration, which is how well JOLs generally align with recall performance on average. A broad tendency for over/under confidence in JOLs would be indicative of poor calibration (for example, consistently providing low JOLs despite good recall performance, or providing high JOLs with lower recall performance), whereas JOLs which accurately reflect an individual's recall performance would have good calibration. The second measure, resolution, refers to item-wise accuracy, or how well individuals' JOLs indicate recall for the individual items being studied. For example, good resolution would be obtained through providing consistently accurate JOLs: low JOLs to words that are not recalled and high JOLs to words that are recalled. The measure of accuracy used in this work is resolution.

Metacognitive monitoring can be described using Koriat's (1997) Cue-Utilization Theory of metacognition which involves three types of cues. Cues are sources of information that are used by an individual to inform their metacognition. The first type of cue is an *intrinsic cue* which consists of factors within information itself. Intrinsic cues are "baked into" the information, like the concreteness of a word which describes the degree to which a word refers to a real thing that can be readily understood and utilized by individuals (e.g., Charbonnair & Wartena, 2019). Individuals' JOLs tend to be sensitive to intrinsic cues such as difficulty (e.g., Arbuckle & Cuddy 1969; Koriat 1997) and word frequency, which has a small effect on increasing JOLs (e.g., Fiacconi & Dollois 2020). Intrinsic cues are called as such because they are a part of information itself and are thus intrinsic to that information.

The second form of cue is an *extrinsic cue* which has to do with the external or contextual factors related to information such as juxtaposition, learning environment, presentation medium, etc. Extrinsic cues do not necessarily relate to the information directly but associatively. An example of an extrinsic cue would be the study instructions provided to an individual, which influence their behaviour during a learning experiment. For example, in Arbuckle and Cuddy (1969) participants were asked to remember *all* of the presented associative pairs. Extrinsic cues tend to have a lesser influence on JOLs such as in the case of list and item repetition (e.g., Koriat 1997) or list composition (e.g., Laursen & Fiacconi, 2021). Where an intrinsic cue is a part of any given piece of information, an extrinsic cue exists separately from the information itself and can be changed without affecting the intrinsic properties of the information.

The third type of cue is called a *mnemonic cue*, which is a subjective, cognitive cue that is produced by the individual processing the information. A mnemonic cue can be a product of either of the other types of cues or can be in the form of a higher-level assessment related to the information. Mnemonic cues are shaped by individuals' subjective experience as well as the nature of a given task and can be used to inform JOLs through means such as processing fluency (see Fiacconi et al., 2020 for a review and meta-analysis). A mnemonic cue exists outside of the object realm of the information and only exists within an individual's cognition. An example of a mnemonic cue would be in the context of one’s goal of study for an upcoming test: You could ask yourself how well you “feel” you know the material to determine how much time you spend studying a certain term or chapter.

The Cue-Utilization Theory (Koriat, 1997) allows us to discuss metacognition in a context that emphasizes not only the role of the information itself but also the cognitive structures that are involved in the behaviour and decisions driving memory formation. Metacognitive monitoring can also be related to metacognitive control, though the exact relationship between the two processes has not been fully specified and varies across conditions. In a chapter by Nelson and Narens (1990), the relationship between monitoring and control is explored. Monitoring refers to the cognitive processes that represent information and control refers to the behaviour related to those cognitive processes. In this model, it is accepted that metacognition can be separated into two distinct levels. The first level is solely cognitive at what is called the “metacognitive level” which contains a mental representation of the real word. The second level contains the real world (that is, one’s environment outside of cognition) called the “object level”. You cannot directly interact with or change information at the metacognitive level, but it is possible to interact with the object level. Monitoring describes the one-way flow of information from the object level to the metacognitive level. That is, information from the real world is *monitored* and used to change and adjust one’s mental representation or beliefs which reside at the metacognitive level. When this metacognitive information is used to act upon or change something in the real world, it does so through exerting *control*. Control refers to behaviour that engages with the object level. This behaviour can only indirectly change the metacognitive level by altering the real-world information being monitored. Metacognition is thus able to be described as the cyclic transfer of information between the object and the metacognitive levels. The upward flow of information which is gathered through monitoring and the downward flow of information used to inform behaviour and alter the object level. In Nelson and Narens' model, the two levels (metacognitive and object) as well as the two processes (monitoring and control) exist to describe the direction and flow of information between an individual's cognition and the world around them. It also allows us to describe where different types of information are situated in metacognition research.

To link the three types of cues directly to this model of metacognition, intrinsic and extrinsic cues exist in the real world and can be used to inform monitoring and to create mnemonic cues. These can in turn be used to influence objects at the object level through exerting control on things at the object level. For example, if you were to study a page in a textbook, the world around you, including the information on the open page, would exist at the object level. As you read the page, the text would provide intrinsic cues related to the topic you are studying. The study environment around you would be full of extrinsic cues, including the textbook itself which contextualizes the chapter and page you are currently reading. Your prior knowledge of the topic exists as a mnemonic cue. As you read, you are monitoring information from the page in the book and adding it to your mental representation of the topic alongside what you already know about the topic. You can then make an informed decision about when to turn the page, altering the real world and the information available to you through exerting control. The decision to turn the page could be because you have decided that you know enough about the topic on this particular page, or because you have decided to read the page in full and would like to continue to the next. You can turn the page, skip pages, or even close the book entirely but you cannot directly change your mental representation of the topic.

Control

Control in metacognition is enacting behaviour at the object level which is informed by information gathered from monitoring. An important control process related to metamemory is study-time allocation (e.g., Thiede & Dunlosky, 1999). By understanding how individuals allocate their study time we can understand how the qualities of information themselves can be related to the metacognitive processes and behaviour strategies used in learning. A self-paced study design can be used to measure how individuals exert control over information. Conventional fixed-paced studies use equal, fixed durations for the presentation of stimuli which allows for researchers to exert experimental control over both the pace of the study as well as the duration in which participants have access to study materials. Self-paced study designs differ in that the participant manipulates the flow of information instead of the researcher. By allowing participants to control the pacing of the study, it is possible to make inferences about some of their metacognitive control strategies.

Nelson and Leonesio (1988) pioneered the roots of our modern understanding of self-paced study. They designed experiments where either accuracy or speed of learning were prioritized. In the first experiment, participants studied trigrams for a free-recall task, in the second, they studied paired associates (paired trigrams) and, in the third, they studied general-information items. They found that participants' metacognition influenced study time allocation based on the perceived difficulty of items (as assessed with Ease-of-Learning – EOL judgements) and individuals' Feeling-of-Knowing (FOK) judgements of the material; more study time tended to be given to harder items in general. One interesting quirk found in this research is that participants tended to be ineffective with their self-pacing, often terminating study before fully learning an item even when instructed to prioritize learning it. In addition, self-paced study follows the law of diminishing returns in that large increases in study time dedicated to an item often yielded little to no increase in recall: an effect dubbed the "Labour-In-Vain" effect.

A decade later, Dunlosky and Thiede (1998) proposed a theory of how individuals self-pace their studies backed by three experiments which involved studying Swahili-English word pairs. In the first experiment participants were given instructions prioritizing speed or accuracy and correct answers were awarded 1 or 10 points. When the instructions emphasized accuracy, items were studied for longer times which was associated with better recall performance. The second experiment involved assigning higher point values for recall (8-, 16-, and 64-point items), added a point cost per second of study time condition that would emphasize speed (1 point cost per second of study time) or accuracy (0-point cost per second), and additionally participants were directed to use one of two specific strategies. The result was that participants did not allocate their study time proportionate to the point value of each item; when there was no cost associated with study time, study time increased as did recall compared to when there was a cost associated with studying. For the third and final experiment, the cost or no cost conditions were kept though this time each item studied was presented alongside a likelihood that the item would appear on the recall test (10%, 50%, and 90% chance of appearing). The result was that participants still allocated their study time inefficiently and spent a similar amount of time studying all items, even those with only a 10% chance of being on the test.

Dunlosky and Theide (1998) proposed the *Norm-Affects-Allocation* model to explain their findings. This model posits that people adjust their desired degree of learning for an item, and study to meet that goal. The degree of learning sought for any given item is called a *norm of study* and is derived from the specific goal(s) of the study task. That *norm* is then used to *affect* study time *allocation*. That is, participants' study time strategies were not based solely on item difficulty (an intrinsic or mnemonic cue), but by the overall context of the goals for that specific task. Participants will often attempt to optimize their study time allocation whether that be to maximize points on a test, seek mastery, or minimize the total study time required. An example of this would be cramming for an exam: one might allocate more time to learn the answers to many low difficulty questions (e.g., multiple choice) instead of spending time learning fewer, more difficult concepts (such as long answer questions). However, Thiede and Dunlosky (1999) continued this line of research and again found that study time allocation was sub-optimal in many cases. From this it can be understood that metacognitive control is goal-oriented, and the way to orient toward that goal is making a metacognition-informed decision that considers the intrinsic and extrinsic cues related to the available information.

There have been a wide variety of studies examining self-paced study and study time, which apply to metamemory. Several studies take a broad approach toward study strategy. For example, Hoffman-Biencourt et al. (2010) used a self-paced study design of picture pairs and found that children allocated study time based on a memorization effort heuristic: The less time something takes to study, the more memorable it feels. Koriat and Ackerman (2010) conducted metacognition and “mindreading” experiments using Hebrew word pairs. In these experiments participants performed a paired-associate learning task themselves and watched another person do the same task. Their findings in this study reiterated that the memorization effort heuristic is present in adults, that is, shorter study times yielded higher JOLs; however, this heuristic does not appear to be based on an explicit belief. It seems as though this heuristic is rooted in the subjective experience of studying. When individuals must assess how others should allocate their study time, they believe that items that take less time to study should be less memorable.

In a set of word study experiments, Tullis and Benjamin (2011) found that individuals who were able to control their study time allocation tended to outperform those who did not control their study time, even when study time was applied per-item based on difficulty or study time was yoked between self-pacers and fixed-pace learners. The benefit of self-pacing was most evident for those who used study time in a discrepancy-reduction strategy. That is, when participants strategized and used additional study time to learn difficult items, they showed improved memory.

Control is not only a means toward approaching a goal but is a part of the integrated experience of a task, providing feedback that is useful in the metacognitive component as well. Memory performance requires well informed control and control is inseparable from metacognition. Study time is an excellent measure of control as it can be measured precisely and naturally during a study task. If an individual systematically alters their applied study behaviour in response the material being studied, then a change in observed study time should be a strong indicator. Participants may alter their study behaviour in response to intrinsic cues pertaining to the information being studied, such as the emotional qualities of words.

Emotion and Memory

Emotion has long played a role in memory research, and a lot of work has been dedicated to unravelling and understanding emotion. Early interest in emotion began as work on describing facial representation of emotional states and reactions with work such as Ruckmick (1921). This work led, for example, to the quantification of facial expressions by Borgatta (1961). Borgatta's quantification of emotion relied on many different scales that were only beginning to be refined. In memory research today, emotion is considered to fall on the two primary axes of valence and arousal, although some have proposed an additional element called dominance (Russell, 1980) which is how dominant or submissive an emotion is. This concept stems from models such as the circumplex model of affect posited by Russell (1980) in which different qualitative emotions fall on a circular plane with two dimensions: pleasure-displeasure and degree-of-arousal which can be clearly linked to the aforementioned valence and arousal dimensions.

In our current understanding, valence is the emotional charge of something—whether it is positive, negative, or neutral—and arousal is how exciting a word is or what level of activity it implies. In memory research there has been some debate as to relatedness of valence and arousal and the degree to which each independently influences memory for emotional items. For example, Mather and Sutherland (2009) argued that valence and arousal are linked, and that arousal drives the effect of valence in memory. In this view, the effects of both valence and arousal are all driven by arousal. However, in the same year as Mather and Sutherland's paper, Mickley Steinmetz and Kensinger (2009) found evidence that disputed the arousal model. They used Functional Magnetic Resonance Imaging (fMRI) to observe participants in a memory study and found that the study of negative valence as well as high arousal images was associated with occipital and temporal lobe activity (most strongly with temporal) and the study of positive valence or low arousal images was associated with frontal lobe activity. In this study, the brain activity associated with emotional information could be used to predict recall performance. This suggests that there are at least two types of encoding used between the two affective qualities.

Adelman and Estes (2013) had participants rate words on their emotional qualities and then conducted an immediate recognition test. They conducted a regression analysis and found that both positive and negative valence were predictive of memory performance, whereas arousal was not. Their results suggested that arousal acted neither independently nor interactively with valence. The authors posit that valence likely plays a facilitative role in memory. At this point in the literature, it is suspected that valence and arousal are distinct in concept but are practically related (for a review see Kuppens et al., 2013), though how precisely our memory systems and metacognition are influenced by each individually is not yet well understood.

Metamemory for Emotional Information

Work on emotion in metamemory suggests that polar (i.e., positive, or negative) valence can be associated with higher JOLs than neutral words in different contexts. For example, Zimmerman and Kelley (2010) found that in both cued-recall and free-recall tasks, participants predicted that both positive *and* negative words would be more memorable than neutral words; the emotional content however had a more pronounced recall effect in the free-recall design. Adelman and Estes (2013) found similar results: Participants predicted higher recognition of both negative and positive words than of neutral words. In work done by Tauber and Dunlosky (2012), JOLs for emotional words differed between age groups. When comparing young adults with older adults the researchers found that both young and older adults gave higher JOLs to negative items however the two groups differed when it came to positive items. Positive items received higher JOLs from young adults, but not from older adults. This work shows that metacognition can change over the course of one's lifetime and that positive and negative items may be remembered differently depending on the nature of the study task itself and the presentation medium. For example, Hourihan and Bursey (2017) found that positive pictures were given higher JOLs than neutral pictures, though, unlike the aforementioned studies involving words, this did not relate to a memory advantage.

Emotional valence demonstrably influences JOLs consistently and often improves recall performance compared to neutral information. As stated previously, some researchers have attributed this effect to arousal (Mather & Sutherland 2009) which explains the JOL effect as intuition or fluency resulting from a bodily response to arousal. The opposing theory is that these effects are driven by individuals' beliefs and recognition of emotion (this is tested in Hourihan et al. 2017; see below). There is even some evidence to suggest that valence and arousal utilize distinct neural pathways (Mickley Steinmetz & Kensinger 2009). However, a lot of research tends to find that arousal does not influence metacognition and memory performance to the degree that valence does, particularly for words (Adelman & Estes 2013; Hourihan et al. 2017).

Current Study

To date, few studies have examined metacognitive control of the study of emotional words. The present study is intended to analyze and compare metacognition (both monitoring and control) of emotional words. We collected individuals' immediate JOLs and study time in a self-paced study and free-recall test. In a free recall task, participants study a stimulus (a word in this case) and when they are asked to recall it, they are not given any clues or aides to help recall the stimuli. In a cued-recall design, however, stimuli are often presented alongside a 'cue' such as something logically or arbitrarily paired with it during the initial study phase. When asked to recall stimuli only one part of the stimuli pairing is presented. For our design, we would like to understand broadly how metamemory for emotion works specifically in the context of how a student might learn. This is resembled more closely by a free-recall design than a cued-recall design.

This work is intended to be an extension of Hourihan et al. (2017), where the authors ran three free-recall experiments and participants provided immediate JOLs. The first experiment manipulated arousal while controlling valence and the second manipulated valence while controlling arousal. Both of these experiments used discrete-list designs, and a third experiment used a mixed-list design which included words that varied in both arousal and valence. Like Hourihan et al. (2017), the present study also entails manipulating valence and arousal separately. Essentially, the current experiments replicate and extend the first two experiments that used discrete-list designs while addressing one of the authors' specific limitations through the inclusion of positively valenced words (in Experiment 1) and moderately arousing words (in Experiment 2). The critical addition in this study is an investigation of the role of control in metacognition for emotional words, measured by study time in a self-paced design, and a survey of beliefs about metacognition and memory for both valenced and arousing emotional information.

By using a self-paced design, we will be able to further understand how the control element of metamemory is related to the study of emotional information. There are presently few examples in the literature that attempt to measure and discuss control in a context similar to this study. According to the theories of study time allocation and control by Dunlosky and Theide (1998) and Theide and Dunlosky (1999), individuals set a norm-of-study and choose to terminate study when they feel that they have met their study goals and have satisfactorily met their norm-of-study. From previous work on emotional words, it is expected that participants will expect valenced words to be more memorable as those emotional words tend to be given higher JOLs than neutral words (Hourihan et al. 2017). It follows from this that we should expect participants in the current study who utilize emotion as an intrinsic cue to meet their norm-of-study sooner when studying emotional words when compared to neutral words resulting in positive and negative words being given less study time than neutral words.

The instructions for our study prioritize the memorization of all words presented— or at least as many as possible. What we expect to find would likely be the use of a memorization effort heuristic (ease-of-learning) in which the types of words provided with a high JOL were studied for less time. Accordingly, we predict a negative relationship between JOLs and study time. This study will explore not only how the valence and arousal dimensions of emotion affect JOLs and recall but also if there is any indication that participants decide to exert metacognitive control by allocating their study time in relation to the emotionality of information or as informed by their metacognitive monitoring. In addition to the free-recall task, this study contains a survey of metacognitive beliefs intended to assess how participants informed their metacognitive decisions. This will provide insight into how the processes of monitoring and control are related. Importantly, the survey responses will provide an opportunity to explore the depth to which individuals have access to their metacognitive intuitions.

Interestingly, Witherby's (2019) dissertation used a similar design for their first experiment and is, to date, the only work similar in design to our experiment. Witherby's work was intended to explore how students allocate study time when learning emotionally valenced material. Their first two experiments involved emotional words in a free-recall task design; the latter two of their experiments involved emotional pictures. For the scope of this work, we will discuss only the first two experiments in detail here. In their first experiment students provided immediate JOLs after self-paced study of words that were positive, negative, or neutral in valence, and then completed a free-recall test. In their second experiment, the same procedure was repeated but the study task was divided into two phases, one where the experimenter determined the study pace and a second where the participants were again in control of the study pace.

In both Experiments 1 and 2, there was no discernable difference in median study time for emotional words. However, in both experiments, emotional words were given higher JOLs than neutral words. In the first experiment, neither positive nor negative words were associated with a recall advantage when compared to neutral words. In the second experiment, the only significant difference was between negative words and neutral words, with negative words more likely to be recalled. Witherby's work is the only extant work with a similar design to our own with emphasis on study time allocation. In Witherby's wordlist design however, valenced words were higher in arousal than neutral words due to the co-variance of valence and arousal. In our design we will be examining valence and arousal separately, thus it is unclear as to how our findings may differ from those found by Witherby. As iterated above, there are several studies that suggest that valence and arousal influence memory and metamemory differently. It is possible that our findings will differ considerably. The purpose of our study is to understand how the emotional properties of valence and arousal independently influence the metamemory and learning process.

Experiment 1

The first experiment we conducted was focused on the manipulation of valence. Participants studied a mixed list of positive, negative, and neutral valenced words which all had a neutral level of arousal, at their own pace. They provided immediate JOLs after studying each word and them performed a free-recall task before repeating the task again with the same words randomized in different order. We expected findings similar to those of other free-recall studies of metacognition for emotional information: positive and negative valenced words would be given higher JOLs than neutral words while also providing a memory advantage (Hourihan et al. 2017; Zimmerman & Kelley 2010). Using a sample of university students, which tend to be young adults, we would expect participants to be sensitive to both positive and negative valenced words (Tauber & Dunlosky 2012).

From our expectations regarding JOLs, we would predict positive and negative words to be given less study time than neutral words according to the ease-of-learning heuristic (Nelson & Leonesio, 1988). We thus expect lower study times for valenced words and a negative relationship with JOLs. However, we also expect increases in study time to lead to improved recall, with excess study time providing diminishing returns. We would thus expect a moderate relationship between study time and recall. This relationship should, however, be stronger for emotional words compared to neutral ones.

We expect participants to consciously recognize the presence of valenced words and hold beliefs regarding them differently than neutral words given that JOLs are sensitive to valence.

Methods

Participants

Participants were recruited through the Memorial Psychology Research Experience Pool (PREP), in which Memorial psychology students can opt to participate in psychology studies for course credit. Before participation, participants were asked to make sure that they were fluent or native English speakers and were comfortable with visual presentation of words for the study duration (30-60 minutes). Participants were also asked not to participate from a mobile device. Recruitment for the first experiment was open from Memorial University's spring 2020 semester to the end of the summer 2020 semester. The number of sign ups in this time frame determined the sample sized to be used for subsequent analysis. There were a total of 41 participants who signed up within the recruitment window (32 Female, 8 Male, and 1 who declined to provide a gender, Average age 20.29, Average education 14.22 years) and completed the full study procedure, however there were 18 additional cases excluded for reasons ranging from blank or incomplete study data files (n = 11), participants who self-reported being distracted during the study procedure (n = 7) or even exclusion due to cheating (n = 1) in the study procedure (evidenced by perfect recall, output in the same serial order as study presentation). It should be noted however that some of the blank or incomplete datafile exclusions may be due to artifacts where datafiles were accidentally created (for example if the study link was clicked twice). All participants and their data were handled in accordance with Memorial University's psychology ethics policy concerning human participants and online studies. This study's ethics approval letter can be found in Appendix A.

Materials

For this study, a variety of resources were used to assemble the study program. The wordlist for this experiment was designed to manipulate emotional valence while keeping arousal at a moderate level. First, words were selected from the Affective Norms for English Words (ANEW) database (Bradley & Lang 2010) general population norms which were exactly five characters in length. Next, words were selected based on their position on the mean rating scale for valence. Within the ANEW, word qualities are given means based on a Likert-style scale ranging from 1-9. This allowed words to be categorized as high-scoring (positive), medium-scoring (neutral), and low-scoring (negative) for valence. Next 36-word lists were assembled that consisted of 12 high-scoring, 12 medium-scoring and 12 low-scoring words on the variable of interest while controlling other variables, notably by keeping the other important emotion variable near the median of the scale. Words were selected to be exemplars of their category and words were picked to maximize their mean valence score in the case of the positive words or to minimize it in the case of negative valence words. The neutral words were selected to maintain the scores around the median of the valence scale such that they were representative of words that were not strongly associated with either pole of valence. The ANEW database has been used in other studies which utilize emotional words. The 1999 version of the ANEW was used by Witherby et al. (2019), Zimmerman & Kelley (2010) (by extension, Tauber and Dunlosky 2012). The 2010 version of the ANEW was used by Hourihan et al. (2017), and this is the version used in this study.

In addition to the measures within the ANEW, Log_{10} frequency was added and controlled as a measure of word frequency (or how commonly it is used in English literature) from the SUBTLEXus database (Brysbaert & New, 2009). This measure was considered as Hourihan et al. (2017) uncovered a frequency effect in their study. To verify wordlist composition, t-tests were performed to ensure that the high, medium, and low scoring word categories were distinct from each other on the valence dimension but not significantly different in arousal or frequency. It should be noted that we did not control for the dominance dimension as valence and dominance strongly co-vary. This strong covariance makes separating the two difficult, especially with exemplars of positive and negative valence. Thus, valenced words also varied in dominance. The wordlist (Appendix Table B1) as well as the associated t-tests (Appendix Table B4) can be found in Appendix B.

The study program was written in JavaScript and designed to randomize and present words to study from our wordlists as well as capture response times, JOL ratings and free recall responses. Qualtrics surveys (Qualtrics, 2020) were used both before and after the study portion of the experiment. Preceding the study phase was a Qualtrics survey containing the informed consent process (mandatory) and demographics (optional). Upon completion of the informed consent process, participants were redirected to the study web application. Once the study process was completed participants were directed to a second Qualtrics survey that contained a beliefs questionnaire before being directed to a final debriefing page. The beliefs questionnaire (Appendix C) consisted of response trees that were intended to provide qualitative insight into what metacognitive strategies participants were using throughout the study process.

The surveys used before and after the study component utilized a combination of forced-choice responses as well as open response questions. Open response questions were used for both demographics and beliefs responses. In the case of the demographics process this allowed for participants to write in any gender or age they wanted including the decision not to specify. For the beliefs, this decision allowed participants to disclose their thoughts without any limitations or potential biases introduced through the presentation of forced-choice responses. The full question tree used for the beliefs survey can be found in Appendix C.

Design and Procedure

The study was a 2 (study block) x 3 (Valence levels) repeated-measures design. The two blocks presented the same words, but in a different randomized order. The reuse of the same stimuli was intended to allow individuals to use repetition as an extrinsic cue and familiarity with the material as a mnemonic cue to inform their JOLs. This approach and rationale is similar to the repeated-measures designed used in the first two experiments in Zimmerman & Kelley (2010) where participants studied the same word-pairs in the first and second trials with newly randomized presentation order in the second experiment. The three valence levels were the three separate emotional word categories: positive valence, neutral valence, and negative valence which exist in a mixed-list design.

Three dependent variables were measured: JOL ratings of words, the time spent studying those words, and recall accuracy.

When participants were directed to the study application, a short practice sequence was presented to allow them to become familiar with the study procedure and process. The practice trials consisted of five three-letter words (the practice wordlist can be found in Appendix Table B3). A word was presented on screen until the participant chose to move on by clicking a button labelled “next” (*study* screen). The length of time the participant spent at this stage was measured as the study time for the presented word. Next, participants were presented with a *ratings* screen that asked them to rate how well they believed they would remember the word they had just studied. Ratings were made by clicking one of eight checkboxes ranging from 1- “I will definitely NOT remember” to 8- “I will definitely remember”. This resulted in a JOL for each word ranging from 1 to 8. The *study* and *rating* process was repeated for each of the 36 words until no words remained.

Next, participants were presented with a distractor task that was designed to occupy a 72-second (approximate) retention interval. The distractor task in this case was assessing math problems; each problem was preceded by a 2-second fixation period and consisted of a 2-second presentation. Participants were instructed to rate whether a simple addition (or subtraction) problem was true or false (e.g., 1+7 = 10 is this true or false?). Once 18 problems were presented, the *recall* phase began. Participants were able to proceed to the free-recall task, where they were presented with a grid of textboxes. Participants were instructed to recall as many of the presented words as they could remember. The “next” button on this page had a two-minute lockout to encourage participants to engage with the task and ensured that participants spent a minimum of two minutes in the recall section. Once the participants were finished entering words, they could press “next” to proceed. If this were the first recall phase, they would be asked to repeat the procedure again beginning with a re-randomized study phase. All components of the study were repeated in the same order. Screenshot examples of the study procedure are available in Appendix D. After the second recall phase, participants were redirected to the structured beliefs questionnaire hosted on Qualtrics.

The beliefs questionnaire was structured so that information about the structure of the wordlists would start obfuscated and be revealed gradually as the participant progressed through the questions. The questions were intended to obtain participants’ intuitions on the processes underlying how they allocated their study time for each word, and whether they had made conscious decisions regarding their studying. The survey logic was set up so that participants who answered “yes” to questions regarding the structure of the wordlists (e.g., “Did you notice anything about the structure of the word list?” “Did you find some words to be more emotional than others?”) would then be presented with a textbox where they were able to elaborate. Once the beliefs survey was completed, participants were presented with a quality of data survey that asked questions to determine whether they were engaged with the tasks during the study phase. Additional questions were asked about the technical integrity of the web application to ensure that the recorded responses were valid. Finally, participants were presented with another opportunity to withdraw their consent for their data to be used at the end of the study before being presented with the feedback/debriefing form.

Results

Analytic strategy

Results were analysed using Jamovi (The jamovi project, 2021) and the JASP plugin (The JASP team, 2021) to conduct frequentist statistics as well as Bayes factors using chance-level priors. That is, for each possible outcome, the probability of randomly selecting a given outcome is used as a prior probability. For frequentist statistics, a series of three 2 (study block) x 3 (valence) analyses of variance (ANOVAs) were conducted for each primary dependent variable. For significant main effects, post-hoc *t*-tests were performed using Tukey correction.

Gamma correlations were calculated for the relationships among each dependent variable for a total of three sets of correlations which describe JOL-Recall accuracy (JOL x recall), the relationship between JOLs and study time (JOL x study time) and the relationship between recall and study time (recall x study time). Gamma correlations are non-parametric and compare the ranked association between two items. This is necessary as two of our scales—JOLs and recall—are not continuous. JOLs are a subjective ranking of word memorability and recall is a binary ranking. The gamma correlation value is a measure of how well two items share their ordinal ranking. For example, if the highest JOLs were given for words with the highest study times, then the gamma correlation between the two would be large in magnitude, and positive. However, if the opposite was true (i.e., the highest JOLs were given for words with the shortest study times), then the relationship would be large in magnitude and negative. If JOLs were not consistently associated with a pattern of study times, then the gamma correlation between the two dependent variables would be close to zero. Through using gamma correlations as a dependent variable can observe whether this rank relationship changes in response to our independent variables. For example, if positive words have a larger, positive JOL x recall gamma correlation when compared to neutral words, that would be evidence in support of positive words having higher resolution accuracy than neutral words. A large positive JOL x recall gamma correlation means that words that are given the largest JOLs are those which are recalled most often. For a further discussion of the utility of gamma correlations see Nelson (1984).

These three gamma correlations were calculated for each participant, split by study block and valence. The gamma correlations were then used to observe whether the relationship between the dependent variables changed in relation to the independent variables. To do this, three additional 2 (study block) x 3 (valence) ANOVAs were conducted using the gamma correlations as the dependent variable.

For all frequentist statistics, alpha = .05 was used as the threshold for significance. All frequentist statistics are paired with a Bayes factor inclusion score (Rouder et al. 2012) to provide additional (non-frequentist) evidence for effects. Bayes inclusion factors indicate the relative change in the strength of support within the data in favour of an effect when a model term is included through evaluation of the odds of an effect existing with (or without) the term included in the model. The Bayes factor calculated here includes both matched and unmatched models. That is, essentially, all possible models that include the term of interest are calculated and compared. The change in in odds from prior to posterior gives rise to the inclusion score. This score is discussed in a JASP tutorial paper by van den Bergh et al. (2020). The Bayes factor inclusion score is also a measure of effect size, with larger scores (greater than one) indicating support for the presence of an effect (support for the alternative hypothesis) and lower scores (less than one, or a fraction) indicating support for the null hypothesis, or lack of an effect. A Bayes factor score of less than one will never be negative but can be interpreted as an inverse fraction which shows the degree of support for the null hypothesis. Here the Bayes factor inclusion scores are presented as decimal numbers rather than fractions for ease of interpretation. Anaszewicz et al. (2015) provide a useful table (Table 1 in their work) of categories used to interpret the relative strength of evidence obtained using Bayes Factors (BF). A BF score of 1 is considered to show no evidence, 3 is anecdotal evidence for the alternative hypothesis, 10 is moderate, 30 is strong, 100 is very strong, and a number greater than 100 is considered extreme evidence. Of course, the same categories hold true for the inverse where 1/3 is anecdotal evidence in favour of the null hypothesis and so-forth. For a description of the utility of Bayes factors, see Jarosz and Wiley (2014).

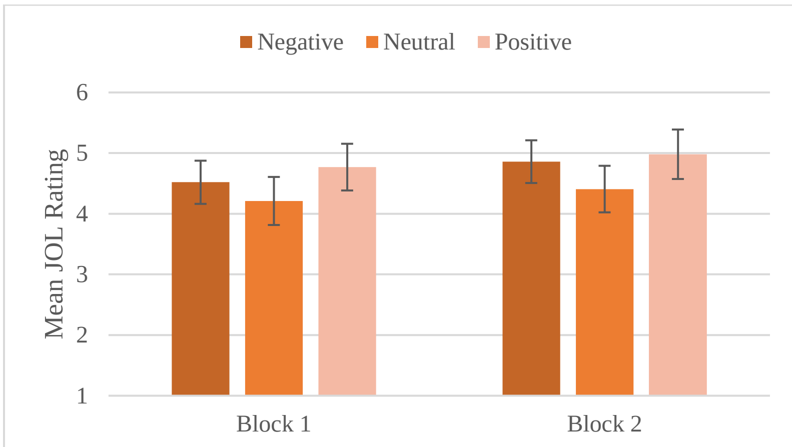
Due to extensive variability in the recorded study times (likely due to the online study format; see below), study time data were trimmed on a per-subject basis. To do so, a recursive trimming procedure was used (Van Selst & Jolicouer 1994; Whelan 2008). This trimming process was used to exclude response times that fell above or below 2.5 standard deviations of the mean study time for each participant. This mean was recalculated with each case removal. The result of this trimming process was a much smaller spread of study times within and across participants at the cost of word by participant case exclusions. However, this trimming process helps to eliminate cases that skew the study time data but also study times that are unrealistically short/long due to inattentiveness or distraction. Because of the nature of this study being administered online, recursive trimming helps preserve the quality of data of the final analysis (Van Selst & Jolicoeur 1994). In total 6.74% (199 of 2952) word study blocks (41 participants x 36 words x 2 study blocks) were excluded on this basis; items that were trimmed on the basis of the above procedure were excluded from analysis entirely (i.e., a word trial with an outlying study time was also excluded from JOL and recall analysis)

JOLs

The first analysis used JOLs as a dependent variable. Mean JOLs are shown in Figure 1. The ANOVA revealed a main effect of valence $F(2,80) = 17.74$, $p < .001$, $\eta^2_p = 0.307$, $BF_{Incl} = 11209.01$. Neither the main effect of block ($F(1,40) = 2.66$, $p = .111$, $BF_{Incl} = 5.49$ nor the valence x block interaction ($F(2,80) = 0.74$, $p = .479$, $BF_{Incl} = 0.33$) were significant using frequentist statistics, though the Bayes factor analysis shows only anecdotal evidence for the block effect. Planned comparisons showed that negative words were given higher JOLs than neutral words ($t(40) = 3.91$, $p < .001$, $d = 0.61$, $BF_{10} = 1034.10$) as were positive words ($t(40) = 5.36$, $p < .001$, $d = 0.84$, $BF_{10} = 1.34e+6$). This evidence suggests that participants' metacognitive monitoring was informed by the valence of the studied word, with both positive and negative valenced words receiving higher JOLs than neutral words. There was no difference between positive and negative words, however, ($t(40) = 2.14$, $p = .095$, $d = 0.33$, $BF_{10} = 2.14$.

Figure 1

Experiment 1 JOL Ratings for Words Across Valence Categories and Blocks



Note. Error bars represent 95% confidence interval around the means

Recall

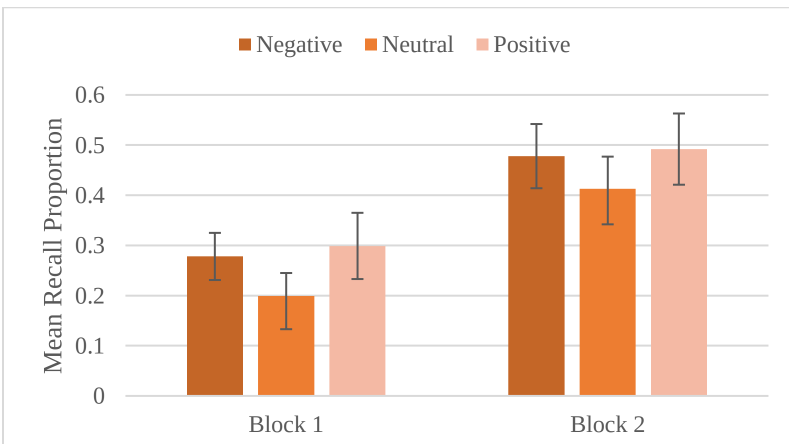
The second analysis used recall proportion as the dependent variable. The ANOVA revealed a significant main effect of valence $F(2,80) = 6.88$, $p = .002$, $\eta^2 p = .147$, $BF_{Incl} = 170.79$, as well as a significant main effect for block $F(1,40) = 102.57$, $p < .001$, $\eta^2 p = .719$, $BF_{Incl} =$ *inf*[1] . The valence x block interaction, however, was not significant ($F(2,80) = 0.24$, $p = .788$, $BF_{Incl} = .37$) . Negative words were recalled better than neutral words ($t(40) = 2.63$, $p = .012$, $d = 0.41$, $BF_{10} = 16.46$) as were positive words ($t(40) = 3.81$, $p < .001$ $d = 0.60$, $BF_{10} = 1137.77$).There was no difference between positive and negative words ($t(40) = 0.69$, $p = .497$, $d = 0.11$, $BF_{10} = 0.17$). This suggests

[1] This *inf* value is representative of a software overflow error. The output number is so high that it cannot be displayed fully.

that recall improved between valence categories and blocks independently. Words were more memorable based on their valence, and memory scores improved from first block to the second ($t(40) = 10.13$, $p < .001$, $d = 1.58$, $BF_{10} = 2.58e+22$ as would be expected on any memory task (a practice effect). The pattern of results can be seen in Figure 2.

Figure 2

Experiment 1 Mean (correct) Recall Proportion for Words Across Valence Categories and Blocks



Note. Error bars represent 95% confidence interval around the mean

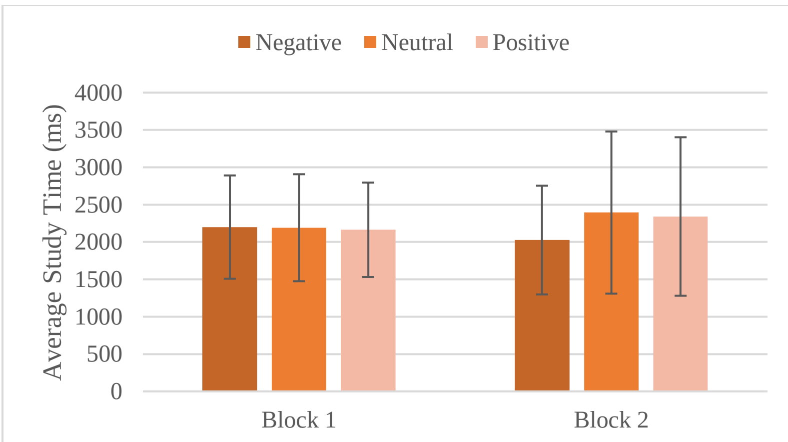
Study Time

The third analysis used study time as a dependent variable. There were no significant findings in the ANOVA and the Bayes factor analysis favoured the null model for all terms; for valence $F(2,80) = 1.54$, $p = .221$, $BF_{Incl} = .04$, block $F(1,40) = 0.02$, $p = .881$, $BF_{Incl} = .10$, and the valence x block interaction $F(2,80) = 1.68$, $p = .192$, $BF_{Incl} <$

.01. The pattern of these results can be seen in Figure 3. From these results it is likely that neither word valence nor block influenced study time allocation.

Figure 3

Experiment 1 Mean Study Times for Words Across Valence Categories and Blocks



Note. Error bars represent 95% confidence interval around the mean

JOL x Recall Gamma

For the fourth analysis, we calculated JOL by recall gamma correlations which show the strength of the relationship between JOL and recall and compared them in a 2 (study block) x 3 (valence) ANOVA (Appendix Table E1) to understand how the relationships between JOLs and recall may have changed across our independent variables. In this case there were no significant effects and the Bayes analysis tended to strongly favour the null hypothesis; for valence $F(2,58) = 0.46$, $p = .636$, $BF_{Incl} = .05$, block $F(1,29) = 2.34$, $p = .137$, $BF_{Incl} = .22$, and the valence x block interaction $F(2,58) =$

1.05, p = .357, BF_{Incl}= .02. This suggests that the relation between JOLs and performance (resolution) did not change dependent on word valence or between blocks.

JOL x Study Time Gamma

For the fifth analysis of Experiment 1 we conducted an analysis of the gamma correlation between JOLs and study time (Appendix Table E1). There were no significant main effects for valence $F(2,74) = 0.18$, p = .837, BF_{Incl}= .04, or block $F(1,37) = 0.34$. p = .561, BF_{Incl}= .14, however there was a significant valence x block interaction $F(2,74)$ =3.47, p = .036, η^2p = .086. Post-hoc testing (using Tukey correction) showed no significant effects, however, and according to the Bayesian analysis the interaction is not supported as BF_{Incl}= .04. Given the current evidence, despite the p-value less than .05 found in the frequentist analysis, the Bayes factor inclusion value suggests that the effect is not strong enough to conclude that the interaction is meaningful.

Recall x Study Time Gamma

For the sixth and final analysis we repeated the previous analysis using recall x study time gamma correlations (Appendix Table E1) and again found no significant effects with the Bayesian analysis strongly favouring the null model; for valence $F(2,60)$ = 0.12, p = .890, BF_{Incl}= .05, block $F(1,30) = 0.97$, p = .332, BF_{Incl}= .15, and the valence x block interaction $F(2,60) = 1.45$, p = .224, BF_{Incl}= .01. These results suggest that the relationship between study time and recall performance did not change over the course of the study, nor was it influenced by valence. A summary of all Gamma correlation descriptive statistics can be found in Appendix E.

Beliefs

Participants' responses to the metacognitive beliefs survey were coded according to an index of relevant response categories. For a summary of codes and responses for each question refer to Appendix C. Responses were coded by two independent raters, one of whom was the author. There was a $\kappa = .61$ level of agreement averaged across questions (Kappas for individual questions are presented in Appendix Table F1). Discrepancies were resolved by a third independent rater. Due to the non-standardized and open-ended nature of the question categories, some responses did not fall neatly into a single category. As such, Kappa values should be interpreted as a level of reasonable agreement amongst raters. It should be noted that there were 41 valid surveys, and that for questions where the participant was asked to explain their reasoning, it was possible for their responses to fall under multiple response categories, thus some response counts may exceed 41. All yes/no question items were assessed for difference from chance using a goodness of fit Chi-square test compared to a 50-50 split in responses. All forced-choice responses were significantly different from chance except for the responses to question 4.

<u>Question 1</u>

The first question was asked to generally assess if participants found some words easier to remember throughout the study and if so, why. Forty (97.56%) of the respondents indicated that they did, only 1 (2.44%) did not (χ^2 (1, N = 41) = 37.10, p < .001, W = .95). Of those 40, the most frequent explanation was that words that were relatable were easier to remember (13 individuals, 31.71%).

Question 2

The second question was asked to assess participants' control beliefs and whether they believed that studying words for longer made them more memorable. A second part of the question asked how participants decided to finish studying any given word. Twenty-eight (68.29%) participants believed that study time helped them remember, 13 (31.71%) did not (χ^2 (1, N = 41) = 5.49, p = .019, W = .37). The explanations were relatively evenly distributed across categories. The most observed explanation was that participants were using a rehearsal or goal-oriented strategy that increased memorability as they studied. Ten (35.71%) participants stated that they used rehearsal or other goal-oriented strategy while studying. For question 2c the most frequent method used to decide when to conclude studying was after applying their set strategy (18 individuals, 43.90%) followed by when they judged or felt that a word was remembered (14 individuals, 34.15%).

These results show that participants varied in their control beliefs, though their control strategies—the goals they set for studying—were more consolidated. A large proportion of individuals adhered to a set learning strategy, whereas just as many used metacognitive monitoring to inform their control process. This might suggest that the relationship between metacognitive monitoring and control operates in both directions (Nelson & Narens 1990).

Question 3

This question was asked to get a broad understanding of how participants' outcome confidence was decided. Thirty (73.17%) participants were not confident in their predictions of recall, 11 (26.83%) were (χ^2 (1, N = 41) = 8.81, p = .003, W = .46). Most participants cited their confidence or skill level for why their predictions were accurate (6, 54.55%) or inaccurate (21, 70.00%)

It is very likely that this assessment was influenced by participants' perceived accuracy or the ease of recall when the recall phases occurred. This might suggest that participants were using their estimated performance to inform their monitoring processes. This performance estimate may involve participants estimating their performance on the first block and using that information to inform their metacognition during the second block.

Question 4

This question was asked to assess participants' sensitivity to word qualities, and if they would spontaneously offer explanations as to what manipulations may have occurred within the wordlist. Eighteen (43.90%) participants did not notice anything, 23 (56.10%) did (χ^2 (1, N = 41) = 0.61, p = .435, W = .12); there was therefore no evidence that participants reported greater than chance likelihood of noticing difference in word qualities in the study list. Of those 23 who did report noticing word differences, 12 (52.17%) noticed the manipulation of valence. This suggests that the manipulation of valence in this list design was salient enough for at least some participants to recognize it. Four (17.39%) participants suggested that the words were somehow related to each other.

Question 5

This question was intended to directly ask about participants' cue-interaction beliefs regarding the difficulty of words. Five (12.20%) participants believed that words do not vary in difficulty, 36 (87.80%) believed that they do (χ^2 (1, N = 41) = 23.44, $p <$ .001, W = .76). The two most common reasons offered were the relatability of words 14 (38.89%) and the frequency of words 11 (30.56%). This is a similar response pattern as to the first question.

Question 6

This question was a manipulation check. Five (12.20%) participants did not notice the presence of emotional words and 36 (87.80%) did. Of those 36 who did, 22 (61.11%) noticed both emotional and exciting words, 13 (36.11%) noticed only emotional words (χ^2 (1. N = 41) = 23.44, $p < .001$, W = .77).

Question 7

This question assessed the presence of explicit beliefs related to the memorability of emotional or exciting words compared to neutral words. Four (9.76%) did not believe that emotional or exciting words were remembered differently, 37 (90.24%) did (χ^2 (1, N = 41) = 26.56, $p < .001$, W = .80).

Question 8

This question asked whether participants believed that emotional words were remembered differently. Five (12.20%) did not believe so, 36 (87.80%) believed that they were (χ^2 (1, N = 41) = 23.44, $p < .001$, W = .76). Of those 36, 13 (36.11%) said that the difference was due to some words being more relatable. Seven (19.44%) said that it was because the words provided an immediate emotional response or reaction when read. Five (13.89%) said that this was due to words resonating with their current emotional state, or otherwise relating to their current mood. It appears participants were emotionally engaging with the words, both through their memories and through their current and changing affect.

Question 9

This question asked whether participants believed that exciting words were remembered differently. Eleven (26.83%) did not believe exciting words were remembered any differently, 30 (73.17%) did (χ^2 (1, *N* = 41) = 8.81, *p* = .003, *W* = .46). Of those 30, 8 (26.67%) said that the difference in memorability was due to how relatable the words were. Six (20.00%) said the difference was due to how engaging the words were.

Question 10

This question asked whether participants believed that exciting words take more or less study time to remember. Seven (17.07%) said that they did not take more or less time. Four (9.76%) said that they took more time, 30 (73.17%) said that they took less time to remember (χ^2 (2, *N* = 41) = 29.61, $p < .001$, *W* = .85). Of those who said that they took more time 3 (75.00%) attributed this to the associations formed between the studied word and other words and thoughts. Of those who said exciting words take less time to study, 9 (30.00%) said that this was due to how relatable the words were, 4 (13.33%) said that this was because of how familiar the words felt.

Question 11

This question asked whether participants believed that emotional words take more or less study time to remember. Eleven (26.83%) said that emotional words did not take more or less time to study. Four (9.76%) said that they took more, 26 (63.41%) said that they took less (χ^2 (2, N = 41) = 18.49, p < .001, W = .67). Of those who believed that emotional words took more time to study, two (50.00%) said that it was due to how relatable the words were, 1 (25.00%) said it was because of how familiar the words were. Of those who said that emotional words took less time to study 11 (42.31%) said that it was due to how relatable the words were. Five (19.23%) said that it was due to an emotional response caused by the words.

Discussion

This experiment was intended to examine whether JOLs or self-paced study time varied based on valence and if the relationships between JOLs, study time and recall would change due to emotional valence. In the JOL analysis, the valence main effect was supported by the Bayes Factor analysis and the post-hoc testing that revealed that positive words were given higher JOLs than either neutral or negative words, and negative words were also given significantly higher JOLs than neutral words. This finding is in line with previous studies of metamemory for emotional words (Adelman & Estes 2013; Hourihan et al. 2017; Zimmerman & Kelley 2010). For the analysis of recall there was both a valence main effect and a block main effect. The valence main effect suggests that positive and negative words are indeed more memorable than neutral words, as has been found in other work (Adelman & Estes 2013; Zimmerman & Kelley 2010), and as predicted by participant JOLs. The recall block effect will not be discussed in detail, as the improvement corresponds to a testing effect resulting from the repeated-measures design of this study. However, this effect demonstrates that participants were in fact, learning the words and improving which also demonstrates that they were engaging with the study material. The lack of an increase in JOL ratings in the second block despite better recall may suggest that repetition, as an extrinsic cue, was not acknowledged consciously by participants, explaining the lack of influence on participants' JOL ratings. This is similar to the findings in Koriat (1997) where it was found that participants' JOLs were not as strongly affected by the extrinsic cues of list or item repetition as they were for intrinsic cues. However, in Koriat (1997) extrinsic cues did affect recall. The analysis of study time interestingly did not yield any significant findings and the Bayes Factor analysis tended to support the null hypotheses (moderate to extremely strong support for the null hypothesis) as well. Study time allocation was generally flat across valence levels and even across blocks. This suggests that valence does not significantly influence self-paced study control processes. Witherby (2019) found similar results: Study time was not affected by emotion.

Despite the clear valence effects for both JOLs and recall, there was no evidence that the relationship between JOLs and recall differed based on valence, suggesting that resolution was consistent. Resolution did slightly improve across blocks, but this was not supported by the Bayes Factor analysis. The relationship between recall and study time was also not different due to valence. Similarly, the gamma correlation analysis between JOL and study time yielded no significant results and only provided a conflicted interaction effect that was significant using frequentist statistics but not supported with the Bayes Factor analysis. Indeed, the only gamma correlation significantly different from zero in this analysis is for that of positive words in the first trial, it is possible that this caused the conflicting findings. Generally, self-pacing in this study design did not seem to affect metacognitive monitoring or memory performance. Valence however seemed to play an important role in both JOL formation and recall performance.

From the results of the metacognitive beliefs survey, the majority of participants were consciously aware of the presence of emotional words and were able to elaborate on their metacognitive process and strategy. Generally, participants held metacognitive beliefs that specifically concerned emotional words (meaning words with a non-neutral valence). Many of the participants suggested that they were applying the emotionality of the studied words to themselves either through an association with an existing memory or with their current emotional state (Questions 10 and 11, Question 8). Additionally, participants seem to have been utilizing their own intuition: that is, experiencing and recognizing emotional responses to words. (Question 8). From this, it is likely that metacognition for emotional words is driven both by individuals' explicit beliefs pertaining to emotional information as well as implicit feedback based on one's own reaction to the presentation of stimuli. In short, mnemonic cues were derived from memory associations and physiological associations (intuition). Despite the majority of participants holding beliefs about control responses for emotional words (exciting/emotional words should be studied for more/less time in questions 10-11) there was no quantitative evidence suggesting that these beliefs were being applied. The evidence supplied did not show a general increase or decrease in study time for emotional words. This could be an artifact of individuals' selective application of these beliefs regarding control.

Experiment Two

The second experiment is a direct replication of the first experiment except with a wordlist designed to manipulate arousal instead of valence. Participants studied a mixed list of high, moderate, and low arousal words, all of which had neutral valences. If memory for emotional items is driven by arousal as proposed by Mather and Sutherland (2009), we would expect an increase both in JOLs as well as memory for high arousal items compared to low arousal items. Adelman and Estes (2013) found that arousal was not a predictor of memory performance when individuals were studying words in a mixed-list design. This suggests that we should not expect a memory advantage for arousing items; however, this does not mean that JOLs will not be affected. Indeed, Hourihan et al. (2017) found that arousal influenced JOLs without a memory advantage in their first experiment. This suggests that we should expect an increase in JOL ratings for arousing words but not an increase in recall. We would also expect the distribution of self-paced study time to be the same as in the first experiment provided arousal does indeed lead to higher JOL ratings. We would not expect any changes in JOLs or recall accuracy between word arousal levels as there were no observed changes in Hourihan et al. (2017) and there were no improvements in accuracy (JOL-Recall resolution) across valence in Experiment 1. This second experiment is intended to compliment the first by targeting arousal independently of valence. We used three levels of arousal: high, moderate, and low arousal. The moderate arousal words are similar to the neutral valence words used in the first experiment. By using three distinct levels we are able to discuss both experiments with regard to that middle level that is similar across experiments.

Methods

Participants

For the second study 67 students (56 females, 11 males; mean age = 21.15, mean education 15.28 years) were recruited using a similar recruitment process as used in the first experiment. In this second experiment, recruitment was open for the duration of Memorial University's fall 2020 semester. None of the participants had participated in the first experiment. Some of those who participated in the study were excluded from the final sample. There were 40 blank or incomplete datafiles (some of which are likely artifacts of multiple study windows being opened mistakenly or refreshing the page, etc.). There were 14 instances of cheating (100% perfect recall, usually in the original presentation sequence). There were 11 cases excluded for describing being distracted during data quality questions (at the end of the survey) and 3 cases where participants stated that their data should not be used for data quality reasons. There were 9 cases where participants were excluded for inattention as characterized by giving every word the same JOL (for example, quickly pressing '8' to skip through the study phase). In total, there were 77 exclusions and 67 valid cases.

Materials and Procedure

The materials used in this experiment were the same as in the first apart from the word list. For this experiment, a 36-word mixed wordlist was created with arousal manipulated and separated into three levels- high arousal, moderate arousal, and low arousal. This time however, dominance did not significantly vary between levels, likely as a by-product of controlling for valence. All words across the three levels had valence scores centred around the middle of the valence scale: all neutral words. The full wordlist (Appendix Table B2) and the t-tests (Appendix Table B5) used in controlling for each variable can be found in Appendix B. The procedure in this experiment was the same as in Experiment 1, except the study stimuli were replaced with the wordlist that manipulated arousal instead of valence. For this experiment another 36-word list was constructed, using similar principles as the first. The high arousal category consisted of 12 words which had the highest arousal scores, and the low arousal category consisted of 12 words which had the lowest arousal scores. The moderate arousal category contained words that were clustered around the median of the scale. Words that were selected were chosen to be representative of their categories without differing significantly in any other important (and controllable measures) as with the list used in experiment 1.

Results

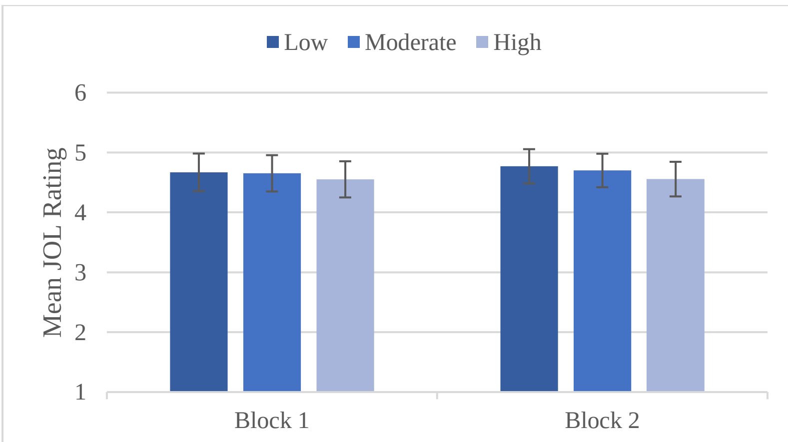
Experiment 2 was an extension of the first experiment except with arousal manipulated instead of valence. As such six more ANOVAs were conducted corresponding to those from Experiment 1. In Experiment 2 the response time trimming procedure removed 6.76% (326 of 4824) study blocks (67 participants x 36 words x 2 study blocks) from the analysis based on being an outlier on the study time variable.

JOL

As in the first experiment, we first conducted a 2 (study block) x 3 (arousal) ANOVA on JOL ratings. There were no main effects nor was the interaction significant. The Bayes factor analysis favoured the null hypothesis in all cases. For arousal $F(2,132) = 2.56$, $p = .081$, $BF_{Incl} = .11$,for block $F(1,66) = 0.14$, $p = .713$, $BF_{Incl} = .10$, and for the arousal x block interaction $F(2,132) = 0.50$, $p = .605$, $BF_{Incl} < .01$. The pattern of these results can be seen in Figure 4. These results suggest that JOLs did not change between the first study block and the second, nor were they influenced by arousal.

Figure 4

Experiment 2 JOL Ratings for Words Across Arousal Categories and Blocks



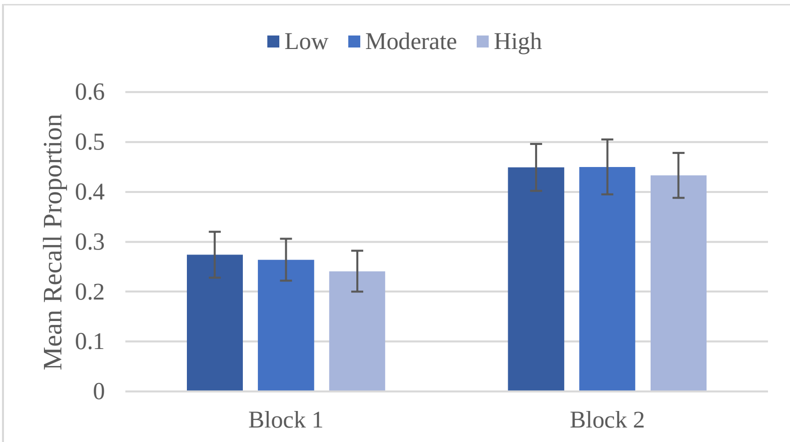
Note. Error bars represent 95% confidence interval around the mean

Recall

The analysis of recall yielded only a significant effect for block $F(1,66) = 167.54$, $p < .001$, $\eta^2p = .717$, $BF_{Incl} = inf^1$, with recall performance improving from the first to second block- as recall scores are expected to increase with repetition. The Bayes factor analysis supported the null hypothesis for arousal $F(2,132) = 0.98$, $p = .377$, $BF_{Incl} = .06$), and the arousal x block interaction $F(2,132) = 0.17$, $p = .848$, $BF_{Incl} = .02$. The pattern of these results can be seen in figure 5. This suggests that recall was wholly unaffected by arousal.

Figure 5

Experiment 2 Mean (correct) Recall Proportion for Words Across Arousal Categories and Blocks



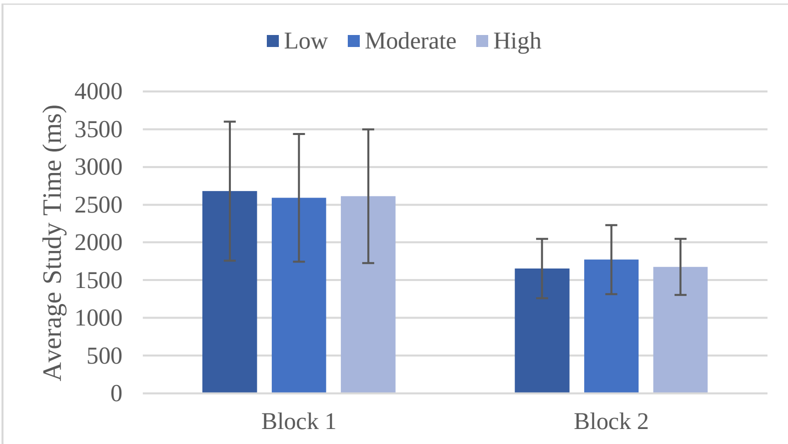
Note. Error bars represent 95% confidence interval around the mean

Study Time

The analysis of study time found a significant effect for block $F(1,66) = 7.49$, p =.008, $\eta^2p = .102$, BF_{Incl} = 398747.71. This result indicates that participants used less study time in the second study block than in the first. There were no other significant effects, and the Bayes factor analysis very strongly supported the null hypothesis for both arousal and the interaction: for arousal $F(2,132) = 0.11$, $p = .900$, BF_{Incl}= .02 and the arousal x block interaction $F(2,132) = 0.79$, $p = .458$, BF_{Incl}= .01, indicating that arousal had no influence on participants' study time allocation. The pattern of these results can be seen in Figure 6.

Figure 6

Experiment 2 Mean Study Times for Words Across Arousal Categories and Blocks



Note. Error bars represent 95% confidence interval around the mean

JOL x Recall Gamma

The analysis of the JOL x recall gamma correlations (Appendix Table E2) yielded a significant effect for block $F(1,55) = 15.38$, $p < .001$, $\eta^2p = .219$, $BF_{Incl} = 468.97$, suggesting that broadly, participants' JOL accuracy (resolution) improved from block 1 to block 2. There were no other significant effects, and the Bayesian analysis supported the null hypothesis for arousal and the interaction: for arousal $F(2,110) = .15$, $p = .863$, $BF_{Incl} = .03$ and the arousal x block interaction $F(2,110) = 1.33$, $p = .269$, $BF_{Incl} = .03$ suggesting that arousal did not influence the relation between JOLs and recall.

JOL x Study Time Gamma

There was a significant main effect of arousal in the JOL x study time gamma correlation (Appendix Table E2) analysis, $F(2,130) = 4.28$, $p = .016$, $\eta^2 p = .062$, $BF_{Incl} = 1.20$. Low arousal words had a higher gamma correlation than high arousal words ($t(65) = 2.55$, $p = .035$, $d = 0.30$, $BF_{10} = 4.24$), which could indicate that arousal influences how participants make JOLs after self-paced study. However, the small effect size combined with an inconclusive BF result suggests that we do not have enough evidence to conclude that there is truly the presence of an effect of arousal on the relation between study time and subsequent JOL. In addition, the analyses for the block $F(1,65) = 0.80$, $p = .375$, $BF_{Incl} = .16$ and for the interaction $F(2,130) = 1.87$, $p = .159$, $BF_{Incl} = .08$ were not significant and resulted in small inclusion scores, suggesting that the JOL by study time relationship did not change from the first study block to the second.

Recall x Study Time Gamma

The recall x study time gamma correlation (Appendix Table E2) analysis did not yield any significant effects. The Bayes factor analyses favoured the null hypothesis in all cases: for arousal $F(2,112) = 0.98$, $p = .380$, $BF_{Incl} = .05$, for Block $F(1,56) = 0.01$, $p = .919$, $BF_{Incl} = .08$ and for the arousal x block interaction $F(2,112) = 1.10$ $p = .337$, $BF_{Incl} = .00$. These results suggest that the relationship between study time and recall performance did not vary based on arousal or block and was consistent in this experiment.

A summary of all descriptive statistics pertaining to gamma correlations can be found in Appendix E.

Beliefs

Participants' responses to the metacognitive beliefs survey were coded according to the same response categories as in the first experiment. For a summary of codes and responses for each question refer to Appendix C. Responses were coded by two independent raters, one of whom was the author. There was a $\kappa = .511$ level of agreement averaged across questions (Kappas for individual questions are presented in Appendix table F2). Discrepancies were resolved by a third independent rater. Due to the non-standardized and open-ended nature of the questions, it is possible that the responses provided did not fall distinctly into the response categories, as such Kappa values should be interpreted with this context in mind. It should be noted that there were 67 valid surveys, and that questions where the participant is asked to explain their reasoning, it was possible for some responses to fall under two or more categories, thus some questions may have a total number of responses exceeding 67. All forced-choice responses were evaluated using a goodness of fit chi-square test. All of these response patterns except for question 6 (a manipulation check) were different from chance.

<u>Question 1</u>

The first question was asked to generally assess if participants found some words easier to remember throughout the study and if so, why. Three (4.48%) participants said they did not find a difference, 64 (95.52%) did (χ^2 (1, N = 67) = 55.54, p < .001, W = .91). Of those 64 who did report a difference, 32 (46.38%) said that this was because words were relatable, followed by 11 (15.94%) who said that this was due to word familiarity.

Question 2

The second question was asked to assess participants' control beliefs and whether they believe that studying words for longer made them more memorable. Twenty-Five (37.31%) participants did not believe that study time increased their memory, 42 (62.69%) did, however (χ^2 (1, N = 67) = 4.31, p = .038, W = .25). Of these 42, 12 (28.57%) said that they used a rehearsal strategy followed by 7 (16.67%) who used frequency to decide how to allocate study time. The majority of participants (34 individuals, 50.75%) used a set strategy to decide when to move on to the next item, 16 (24.88%) moved on when they made an explicit judgement and felt that a word was remembered.

Question 3

This question was asked to get a broad understanding of how participants' outcome confidence was decided. Forty-four (65.167%) participants did not believe that their predicted recall was accurate, 23 (34.33%) did (χ^2 (1, N = 67) = 6.58, p = .010, W = .31). Of those who did not believe their recall predictions were accurate, 38 (88.36%) said it was because of their skill or confidence level. Of those who did believe their predictions were accurate, 17 (73.91%) said it was due to their skill or confidence level.

Question 4

This question was asked assess participants' sensitivity to word qualities, and if they would spontaneously offer explanations as to what manipulations may have occurred within the wordlist. Fifty (74.63%) participants did not notice anything about the presented words, 17 (25.37%) did. Unlike in the first experiment, the likelihood of participants noticing anything about the word qualities in the study list was significantly lower than would be expected by chance (χ^2 (1, N = 67) = 16.25, p < .001, W = .49). Of those 17 participants who did report noticing something about the words, 7 (41.18%) said that the words could be grouped together.

Question 5

This question was intended to directly ask about participants' cue-interaction beliefs regarding the difficulty of words. Three (4.48%) participants did not believe that words varied in their difficulty, the majority 64 (95.52%) believed that they did (χ^2 (1, N = 67) = 55.54, p < .001, W = .91). Of those 64, 22 (33.33%) said it was because some words were relatable, and 20 (30.30%) said that it was because some words had higher frequency.

Question 6

This question was a manipulation check. Thirty (44.78%) participants did not notice the presence of emotional words, and 37 (55.22%) did (χ^2 (1, N = 67) = 0.73, p = .392, W = .10). Of those 37 that said they did, 17 (45.95%) noticed both emotional and exciting words, 12 (32.43%) noticed emotional words only and 8 (21.62%) noticed exciting words only. The chi-square results suggest that individuals were not better at noticing emotionality than from chance.

Question 7

This question assessed the presence of explicit beliefs related to the memorability of emotional or exciting words compared to neutral words. Six participants (9.09%) did not believe that emotional or exciting words were remembered differently, 61 (91.04%) did (χ^2 (1, N = 67) = 45.15, p < .001, W = .82).

Question 8

This question asked whether participants believed that emotional words were remembered differently. Thirteen (19.40%) participants said that emotional words were not remembered differently. Fifty-four (80.60%) believed they were (χ^2 (1, N = 67) = 25.09, p < .001, W = .61). Of those 54, 20 (37.04%) said that this was because some words were more relatable. Interestingly, 13 (24.07%) did not provide an answer.

Question 9

This question asked whether participants believed that exciting words were remembered differently. Sixteen (23.88%) participants did not believe that exciting words were remembered differently and 51 (76.12%) believed that they were (χ^2 (1, N = 67) = 18.28, p < .001, W = .52). Of those 51, 13 (25.00%) believed that this was due to how engaging the words were. Eleven (21.15%) believed that this was due to how relatable the words were.

Question 10

This question asked whether participants believed that exciting words take more or less study time to remember. Fifteen (22.39%) participants said that exciting words did not take a different amount of study time, two (2.99%) said that they take more, and 50

(74.63%) said that they take less (χ^2 (2, N = 67) = 55.19, p < .001, W = .91). Of the two who said that exciting words take more time, one response fell into the category of 'other' and the other did not provide a reason. Of those who said that exciting words take less time to study, the most consistent response (11 or 22.00%) was that the study time difference was due to how relatable exciting words were. Sixteen (32.00%) had responses categorized as "other" and 11 (22.00%) did not provide a response.

Question 11

This question asked whether participants believed that emotional words take more or less study time to remember. Sixteen (24.24%) participants believed that emotional words took the same amount of time to study, two (3.03%) believed that they took more, and 48 (72.73%) believed that they took less, and one participant did not respond (χ^2 (2,N = 66) = 50.55, p < .001, W = .87). Of those two who believed that they took more, they were split between relatability and "other". Of those who said that emotional words take less time to study, 15 (31.25%) said that this was due to the relatability of the words, seven (14.58%) said it was due to an immediate emotional response. An equal number said that it was due to "other" reasons and 11 (22.92%) did not provide a response.

Discussion

This experiment was intended to replicate and extend the first by examining whether arousal influenced JOLs and self-paced study time allocation, and if the relationships between JOL, recall, and study time changed due to emotional arousal. The analysis of JOLs yielded surprising results in that arousal did not have any effect. This finding was supported by the Bayes Factor analysis. This finding stands in contrast to the results of Hourihan et al. (2017) and will be discussed below. The analysis of recall also lacked an arousal main effect, though it had the expected testing effect associated with repeated testing. The presence of a repetition effect in recall without an associated change in JOL ratings between blocks may again suggest that repetition was not recognized as a useful extrinsic cue as was found in the first experiment. The analysis of study time very interestingly showed a main effect of block, which was supported by the Bayes-factor analysis. Participants spent less time studying in the second block than they did in the first block. This finding will be discussed later in the general discussion.

The gamma correlation analysis between JOL and recall yielded a block main effect which was supported by the Bayes Factor analysis. This suggests that resolution robustly improved between blocks. The analysis of the gamma correlation between JOL and study time yielded an arousal main effect that was not supported by the Bayes Factor analysis and was of a low degree and will not be interpreted. There were no significant findings in the analysis of the gamma correlation between recall and study time. It appears the JOL and recall relationships with study time were consistent though study time itself was lowered in the second block. Generally, these findings are interesting, but unexpected. Our findings conflict with the findings of past research that suggest that arousal is a driving factor in memory and metamemory for emotional words (Mather & Sutherland 2009). These results and the possible role of distinctiveness will be discussed further in the general discussion.

Many of the survey findings in the second experiment were quite similar to the first; participants' beliefs were similar in proportion for both emotional (meaning valenced) as well as exciting (arousing) words, that is, many participants provided similar responses on questions explicitly targeting both valence and arousal. This is quite interesting as it may suggest that generally the two independent samples of participants held similar beliefs that were not radically changed by the different manipulation in the word lists separating them. Other questionnaire responses differed from those observed in Experiment 1. In this experiment, proportionally far fewer participants spontaneously noticed any manipulation of the words being presented. Of those who did, a single individual noted valence, and none noted arousal (Question 4). When prompted with a forced-choice question (Question 6), more participants noticed a wordlist manipulation but still far fewer than in the first experiment. Of those who did notice exciting words, in this experiment the majority responded that they noticed “both emotional and exciting words”. In addition, fewer stated that they noticed exciting words alone than those who stated that they noticed emotional words alone. As with the first experiment, the presence of control beliefs in questions 10 and 11 were not reflected in the quantitative analysis of study time.

General Discussion

This study aimed to investigate the role of emotional information in metamemory including both monitoring and control processes. Participants studied emotional words at their own pace and provided immediate JOLs before performing a free-recall task and repeated the study process again with the same words. The first experiment manipulated valence and the second manipulated arousal. In the first experiment, we predicted that words that were either positive or negative in valence would both be given higher JOLs and be more likely to be recalled than neutral words. In the second experiment, we predicted that words with high levels of arousal would be judged to be more memorable than words with low levels of arousal, but not show any recall differences. In addition to our predictions of metacognitive monitoring, we predicted that participants would inform their decision to allocate study time for each word to maximize efficiency: studying less memorable words for longer than more memorable words with word memorability determined by the JOLs given to the studied words. Thus, we predicted a negative relationship between JOLs and study time in-line with a metacognition-informs control model of metacognition as described in Nelson and Narens (1990). We found the expected pattern of results for monitoring in the first experiment but not the second.

Generally, we found some of the results we expected in the first experiment, namely that valence affected both JOLs as well as recall. Positive and negative words were given higher JOLs and were more likely to be recalled, compared to neutral words. This stands in contrast to the second experiment where we found that increased arousal did not lead to increased JOLs which defied our predictions.

The survey of beliefs presented after the study portion of each experiment support this: Many participants were explicitly aware of the valence of words, but none mentioned arousal explicitly. Many participants did not spontaneously notice any patterns or manipulation in the second experiment; proportionately, this number was almost double those in the first experiment and of those who did, valence was the most recognized (perceived) manipulation (see Question 4 in Appendix C). When the question was leading to the two types of manipulation, a great number recognized valence ("emotional") compared to arousal ("exciting"; see Question 6 in Appendix C). This could be a by-product of our experimental design. Using three levels of valence may still provide a distinct contrast between the conceptual categories of what is a positive word and what is a negative word, compared to what is a relatively neutral word whereas the distinction between high arousal, low arousal and moderate arousal words may break down, specifically with the presence of that middle category. The moderate arousal category is unique to this experiment, as are our findings regarding arousal. Indeed, a similar result showing that arousal did not influence JOLs, or recall was found in Hourihan et al. (2017) in their third experiment where a mixed-list design was used with a continuum of words ranging across different valence and arousal scores, thus reducing the relative distinctiveness of words based on a given emotional factor.

Taken together, this absence of an effect on JOLs in such mixed lists may suggest that metacognition for emotional information is driven primarily by a conscious recognition of those emotional items. Indeed, though few participants spontaneously recognized manipulations of arousal, a high number of participants in both studies believed that arousing words (as well as valenced words in question 8) were remembered differently (see question 9 in Appendix C) which suggests that conscious recognition may be the driving force behind the difference in JOLs. Theories which posit that emotion engages with metacognition subconsciously through processes such as autonomic arousal or an otherwise physiological reaction to viewing emotional information (Mather & Sutherland 2009) do not fully account for these findings. Though some participants did take note of their emotional responses to the material, far more suggested that the emotional words being studied were more relatable and, in the case of arousal, participants were split between exciting words being relatable and being engaging (Questions 10-11).

What these results suggest instead is that metacognition for emotional words can be informed both from mnemonic cues and experiences as well as an individual's awareness of their physiological state. That is, there is no immediately apparent holistic difference or bias in the way that participants were treating valence and arousal. The beliefs survey results suggest that the metacognitive beliefs and strategies used for both types of emotional information were similar. If monitoring were completely beliefs-driven, participants would more strongly favour strategies that prioritize their own experiences; if it were driven by physiological reactions then participants would have favoured strategies relying on bodily and emotional awareness, but neither was the case. The observed similarity in beliefs for valence and arousal, despite different patterns of performance for each, may instead be attributable to the encoding processes involved for the two different types of information (Mickley Steinmetz & Kensinger 2009). It is possible that participants adopted an ease-of-processing heuristic while studying. That is, emotional words which stood out may have been recognized by participants and may have been interpreted as more easily processed than neutral words. This could account for the JOL ratings in both experiments. Valenced words could have been interpreted as easier to remember than neutral words, however arousing words did not share this property. The application of the ease-of-processing heuristic could be the result of both physiological processes as well as individuals' beliefs about what type of information is, or should be, more memorable. In a study of word-font and study repetition, Kornell et al. (2011) found that ease-of-processing accounted for participants over-valuing the font size of words while simultaneously ignoring the memory benefit of repeated study. It is possible that our findings are similar in that participants in this study rated word memorability based on ease-of-processing during encoding while on the second study block, ignoring word repetition as a valuable predictor of memory performance. This is despite repetition being a common method for memorization and study. Despite recall performance improving between blocks, JOLs did not increase between blocks.

Changing to a discussion of metacognitive control, we found no systematic effects of emotion on study time. Most participants in both experiments tended to state that emotional and arousing words should take less time to study (Questions 10 and 11 in Appendix C) though we did not find evidence corroborating implementation of these beliefs, suggesting that participants were not actually applying them. This may imply a disconnect between participants' explicit metamemory beliefs and their performed study processes. This could be linked to the effects found by Koriat and Ackerman (2010) who investigated the differences in metacognitive inferences between self (performing the task) and mindreading for others (observing another performing the task and inferring their performance). In their first experiment they had participants perform a paired-associates task and then they were shown videos of a confederate performing the same study task. What they found was that participants who performed the task themselves used a different strategy than they suggested should be done for others. When participants were actually performing the task, they demonstrated the use of a study time heuristic called the memorization-effort heuristic: the less study time needed for a pair, the higher JOL it received. This was not the case however, as when participants watched the confederate perform, they did not consistently predict the confederate's performance with the same pattern as was used their own performance or JOL-study time relationship.

The relevance of Koriat and Ackerman's (2010) study is in their observed difference between one's actions and one's expectations of others. The disconnect that appears in the current study between beliefs and practice may be reflective of this metacognition versus mindreading discrepancy. Alternatively, perhaps when describing their explicit metacognitive beliefs, individuals view themselves in a third-person referential manner and describe their idealized beliefs rather than the principles underlying their own study process. This might suggest that we are unable to tap into the mnemonic cues used by individuals without inherent subjective bias like asking other introspective questions. Another possibility that may underlie our observed discrepancy between beliefs and control is simply that our design cannot account for subjective relationships between beliefs and control or the application of control methods other than study time, such as strategy choice. The ANOVA method of hypothesis testing used here relies on the existence of consistent systematic relationships between variables. If individuals apply their beliefs intermittently or inconsistently especially in relation to other participants, it is very possible that the sum of the effects will lead to a zero-sum or undetectable difference. The next step with this and similar data would then be to perform a moderation analysis that tests the relationship between the participants' stated metacognitive beliefs and the associated measures. Witherby et al. (2021) have an excellent and highly relevant review on this subject discussing the influences of metacognitive relationships for emotional information and a potential pathway model that could be incorporated within such an extension.

There is one other result worth discussing. In Experiment 2, study time decreased significantly from the first block to the second block, and this was not evident in the first experiment. This suggests a difference in the way individuals' metacognitive control was being used between experiments. We can only speculate what may underlie this difference without objective comparison, but it is possible that the study strategy used by participants was fundamentally different between experiments. In the first experiment, there was an obvious and salient distinction between types of words, and in the second it is evidenced that there was no distinction to guide study habits. In short, when participants were studying valenced words, they may have used an entirely different strategy than those who were studying the second experiment (arousal) wordlist. If this were the case, the first strategy seems to have implied a consistent study time allocation approach to words across not only valences but blocks, whereas the second strategy involved practice of words in the first block and less demand for restudy in the second block.

This study has several noteworthy shortcomings. The first is that, due to global circumstances, it was conducted entirely remotely, which severely limits the extent of control over the study environment and the quality of data requires the assumption that participants engaged with the study material mindfully. More cases than was desirable have been excluded due to the quality of the responses, ranging from the exclusion of entire participants who did not respond during the study, cheated, or provided reasons for disqualification in the data quality portion of the survey.

Another shortcoming that is prevalent and persistent across virtually all memory studies involving words and lexical characteristics is inadequate control. At the time of writing, it is known to be difficult to distinguish word valence and dominance. It would be possible to generate small lists of controlled words but difficult to produce lists of a suitable length for a memory experiment. Consequently, this study does not fully account for the dominance dimension, and it necessarily confounds with valence in the first experiment as the two strongly co-vary. The wordlists used here controlled for character length and frequency. Character length was controlled in order to allow words to be matched in both character length as well as the amount of screen space allocated during presentation. Frequency was controlled as it has been shown in Hourihan et al. (2017) to affect JOL ratings for emotional words. Notable factors that were not controlled include: number of syllables, concreteness, age of acquisition and word relatedness. These factors are not often controlled or acknowledged in similar free-recall memory studies, however Bireta et al. (2021) assert that valence does not affect *serial recall* when these dimensions are controlled. It may be possible that the findings of this study are thus attributable to a lexical dimension that was not controlled. This shortcoming may someday have a better solution with the continued efforts of those who study lexical data (e.g., Ensor et al. 2021; Macmillan et al. 2021), and for the sake of transparency, the full wordlists have been provided in Appendix B. Future work could benefit from examining the role of word distinctiveness as it applies to emotion, and how the properties of emotion interact with individuals' perception of word distinctiveness.

As an extension of the distinctiveness comparison, a mixed list consisting of a range of valence and arousal scores is a logical next step from this research, similar to Hourihan et al.'s (2017) third experiment. In such a mixed-list design, the JOL and recall effects for valence would likely disappear if distinctiveness were driving them. One other avenue for future research would be to investigate the role of metacognitive control in a controlled lab setting using measures such as pupillometry and gaze tracking (e. g., Bradley et al. 2008; Lempert et al. 2015; Henderson et al. 2018) to understand not only how participants allocate the study time they have available but also their attention. It is possible that differences exist which are as of yet impossible to capture in an online study.

An additional direction for future research directly stemming from the data gathered in this study would be to commence work on modelling the role of metacognitive beliefs in the metacognitive monitoring-control system with a moderator analysis design. This type of work which objectively relates beliefs survey data to quantitative data would be invaluable for further progressing metacognitive research.

Summary

We found that valence and arousal influence metamemory differently and thus should be considered independently in future work. Related to this point, our survey results showed that despite their independence, valence and arousal share similar subjective metacognitive experiences and likely share similar mnemonic influence and monitoring. Thus, the difference between the two cannot be reduced to a systematic distinction (beliefs vs. fluency, conscious vs. unconscious). On the side of metacognitive control, our results show that emotional factors did not broadly influence how people choose to control their study time for words. Any observed memory benefits such as in the first experiment for positive and negatively valenced words thus are unlikely to stem from study-time based strategies or otherwise differential allocation of study time. The key finding of this study is that valence and arousal produced seemingly similar subjective experiences yet provided different patterns of effect without manifesting a direct and observed change in one's control. We conclude with the finding that metamemory for emotional information is likely driven by valence, and that emotion is not likely to be considered when control decisions regarding self-paced study are made.

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, *129*(3), 530–535. https://doi.org/10.1016/j.cognition.2013.08.014

Anaszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, *41*(2), 521–543. https://doi.org/10.1177/0149206314560412

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81*(1), 126–131. https://doi.org/10.1037/h0027455

Bireta, T. J., Guitard, D., Neath, I., & Surprenant, A. M. (2021). Valence does not affect serial recall. *Canadian Journal of Experimental Psychology, 75*, 35-47. https://doi.org/10.1037/cep0000239

Borgatta, E. F. (1961). Mood, Personality, and Interaction. *The Journal of General Psychology*, *64*(1), 105–137. https://doi.org/10.1080/00221309.1961.9920431

Bradley, M. M., & Lang, P. J. (2010). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings (Technical Report C-2)*. Gainesville: University of Florida, Center for Research in Psychophysiology.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering Pictures: Pleasure and Arousal in Memory. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition. 18(2). 379-390. https://doi.org/10.1037/0278-7393.18.2.379

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Charbonnier, J., & Wartena, C. (2019). Predicting Word Concreteness and Imagery. *Proceedings of the 13th International Conference on Computational Semantics - Long Papers*, 176–187. https://doi.org/10.18653/v1/W19-0415

Dunlosky, J., & Metcalfe, J. (2009). Chapter 3: Methods and Analyses. In J. Dunlosky & J. Metcalfe, *Metacognition* (pp. 37-59). Thousand Oaks, CA: Sage Publications, Inc.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, *98*(1). https://doi.org/10.1016/S0001-6918(97)00051-6

Fiacconi, C. M., & Dollois, M. A. (2020). Does word frequency influence judgments of learning (JOLs)? A meta-analytic review. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *74*(4), 346–353. https://doi.org/10.1037/cep0000206

Fiacconi, C. M., Mitton, E. E., Laursen, S. J., & Skinner, J. (2020). Isolating the contribution of perceptual fluency to judgments of learning (JOLs): Evidence for reactivity in measuring the influence of fluency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 926–944. https://doi.org/10.1037/xlm0000766.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Georghiades, P. (2004). From the general to the situated: Three decades of metacognition. *International Journal of Science Education*, *26*(3), 365–383. https://doi.org/10.1080/0950069032000119401

Henderson, R. R., Bradley, M. M., & Lang, P. J. (2018). Emotional imagery and pupil diameter. *Psychophysiology*, *55*(6), e13050. https://doi.org/10.1111/psyp.13050

Hoffmann-Biencourt, A., Lockl, K., Schneider, W., Ackerman, R., & Koriat, A. (2010). Self-paced study time as a cue for recall predictions across school age. *British Journal of Developmental Psychology*, *28*(4), 767–784. https://doi.org/10.1348/026151009X479042

Hourihan, K. L., & Bursey, E. (2017). A misleading feeling of happiness: Metamemory for positive emotional and neutral pictures. *Memory*, *25*(1), 35–43. https://doi.org/10.1080/09658211.2015.1122809

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2017). The influences of valence and arousal on judgments of learning and on recall. *Memory & Cognition*, *45*(1), 121–136. https://doi.org/10.3758/s13421-016-0646-3

IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp

Jarosz, A. F., & Wiley, J. (2014) What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving*. *7*(1), 2-9. https://doi.org/10.7771/1932-6246.1167

JASP Team (2021). *JASP*. [Version 0.16]. Retrieved from https://jasp-stats.org.

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), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A., & Ackerman, R. (2010). Metacognition and mindreading: Judgments of learning for Self and Other during self-paced study. *Consciousness and Cognition*, *19*(1), 251–264. https://doi.org/10.1016/j.concog.2009.12.010

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The Ease-of-Processing Heuristic and the Stability Bias: Dissociating Memory, Memory Beliefs, and Memory Judgments. Psychological Science, 22(6), 787–794. https://doi.org/10.1177/0956797611407929

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*(4), 917–940. https://doi.org/10.1037/a0030811

Laursen, S. J., & Fiacconi, C. M. (2021). Examining the effect of list composition on monitoring and control processes in metamemory. *Memory & Cognition*, *49*(3), 498–517. https://doi.org/10.3758/s13421-020-01107-4

Lempert, K. M., Chen, Y. L., & Fleming, S. M. (2015). Relating Pupil Dilation and Metacognitive Confidence during Auditory Decision-Making. *PLOS ONE*, *10*(5), e0126588. https://doi.org/10.1371/journal.pone.0126588

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. [R package]. Retrieved from https://cran.r-project.org/package=emmeans.

Mather, M., & Sutherland, M. (2009). Disentangling the Effects of Arousal and Valence on Memory for Intrinsic Details. *Emotion Review*, *1*(2), 118–119. https://doi.org/10.1177/1754073908100435

MacMillan, M. B., Neath, I., & Surprenant, A. M. (2021). Re-assessing age of acquisition effects in recognition, free recall, and serial recall. *Memory & Cognition*, *49*, 939-954. https://doi.org/10.3758/s13421-021-01137-6

Mickley Steinmetz, K. R., & Kensinger, E. A. (2009). The effects of valence and arousal on the neural activity leading to subsequent memory. *Psychophysiology*, *46*(6), 1190–1199. https://doi.org/10.1111/j.1469-8986.2009.00868.x

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs*. [R package]. Retrieved from https://cran.r-project.org/package=BayesFactor.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O. (1990). Metamemory: A Theoretical Framework and New Findings. In *Psychology of Learning and Motivation*, *26*, 125–173. https://doi.org/10.1016/S0079-7421(08)60053-5

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 676–686. https://doi.org/10.1037/0278-7393.14.4.676

Qualtrics (2020). *Qualtrics* (May 2020) [Computer software]. Available at https://www.qualtrics.com

R Core Team (2021). *R: A Language and environment for statistical computing*. (Version 4.0) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2021-04-01).

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356-374.

Ruckmick, C. A. (1921). A preliminary study of the emotions. *Psychological Monographs*, *30*(3), 30–35. https://doi.org/10.1037/h0093142

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Singmann, H. (2018). *afex: Analysis of Factorial Experiments*. [R package]. Retrieved from https://cran.r-project.org/package=afex.

Tauber, S. K., & Dunlosky, J. (2012). Can older adults accurately judge their learning of emotional information? *Psychology and Aging*, *27*(4), 924–933. https://doi.org/10.1037/a0028447

The jamovi project (2021). *jamovi*. (Version 1.8) [Computer Software]. Retrieved from https://www.jamovi.org.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1024–1037. https://doi.org/10.1037/0278-7393.25.4.1024

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language*, *64*(2), 109–118. https://doi.org/10.1016/j.jml.2010.11.002

van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E. J., Derks, K., Dablander, F., Gronau, Q. F., Kucharsky, S., Gupta, A. R., Sarafoglou, A., Voelkel, J. G., Stefan, A., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E. J. (2020). A Tutorial on Conducting and Interpreting a Bayesian ANOVA in JASP. *L'Année Psychologique*, 120, *73-96*. https://doi.org/10.3917/anpsy1.201.0073

Van Selst, M., & Jolicoeur, P. (1994). A Solution to the Effect of Sample Size on Outlier Elimination. *The Quarterly Journal of Experimental Psychology Section A*, *47*(3), 631–650. https://doi.org/10.1080/14640749408401131

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *The Psychological Record*, *58*(3), 475–482. https://doi.org/10.1007/BF03395630

Witherby, A. E. (2019). *How Do Students Regulate Their Learning of Emotional and Neutral Information?* (13898892) [Master's thesis, Texas Christian University]. ProQuest

Witherby, A. E., Tauber, S. K., & Dunlosky, J. (2021). Why Do Emotional Stimuli Influence Judgments of Learning? Theory, Evidence, and Future Directions. In D. Moraitou & P. Metallidou (Eds.), *Trends and Prospects in Metacognition Research across the Life Span* (pp. 3–24). Springer International Publishing. https://doi.org/10.1007/978-3-030-51673-4_1

Zimmerman, C. A., & Kelley, C. M. (2010). "I'll remember this!" Effects of emotionality on memory predictions versus memory performance. *Journal of Memory and Language*, *62*(3), 240–253. https://doi.org/10.1016/j.jml.2009.11.004

Appendices

Ethics Approval Letter (Appendix A)

MEMORIAL UNIVERSITY

Interdisciplinary Committee on Ethics in Human Research (ICEHR)

St. John's, NL Canada A1C 5S7 Tel: 709 864-2561 icehr@mun.ca
www.mun.ca/research/ethics/humans/icehr

ICEHR Number:	**20201914-SC**
Approval Period:	April 30, 2020 – April 30, 2021
Funding Source:	
Responsible Faculty:	Dr. Kathleen Hourihan Department of Psychology
Title of Project:	*Self-paced Study and Word Dimensionality in Metacognition*
Title of Parent Project:	*Metacognition in Individuals*
ICEHR Number:	2012-239-SC

April 30, 2020

Mr. Ryan Lee

Department of Psychology, Faculty of Science

Memorial University of Newfoundland

Dear Mr. Lee:

Thank you for your submission to the Interdisciplinary Committee on Ethics in Human Research (ICEHR) seeking ethical clearance for the above-named research project. The Committee has reviewed the proposal and agrees that the proposed project is consistent with the guidelines of the *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans* (TCPS2). *Full ethics clearance* is *granted to* April 30, 2021. ICEHR approval applies to the ethical acceptability of the research, as per Article 6.3 of the *TCPS2*. Researchers are responsible for adherence to any other relevant University policies and/or funded or non-funded agreements that may be associated with the project.

The *TCPS2* **requires** that you submit an Annual Update to ICEHR before April 30, 2021. If you plan to continue the project, you need to request renewal of your ethics clearance and include a brief summary on the progress of your research. When the project no longer involves contact with human participants, is completed and/or terminated, you are **required** to provide an annual update with a brief final summary and your file will be closed. If you need to make changes during the project which may raise ethical concerns, you must submit an Amendment Request with a description of these changes for the Committee's consideration prior to implementation. If funding is obtained subsequent to approval, you must submit a Funding and/or Partner Change Request to ICEHR before this clearance can be linked to your award.

All post-approval event forms noted above can be submitted from your Researcher Portal account by clicking the ***Applications: Post-Review*** link on your Portal homepage. We wish you success with your research.

Yours sincerely,

Russell J. Adams, Ph.D.

Chair, Interdisciplinary Committee on

Ethics in Human Research Professor of Psychology and Pediatrics Faculties of Science and Medicine

RA/bc cc: Supervisor – Dr. Kathleen Hourihan, Department of Psychology, Faculty of Science

Wordlists (Appendix B)

B1 Experiment 1 Wordlist (Arousal Controlled Words)

Word	ValMn	ValSD	AroMn	AroSD	DomMn	DomSD	ValLvl	Lg10WF
alone	2.41	1.77	4.83	2.66	3.7	2.42	Negative	4.1969
blame	2.77	1.41	4.93	2.02	3.47	1.59	Negative	3.477
decay	2.68	1.66	4.44	2.28	3.63	1.84	Negative	2.0253
filth	2.47	1.68	5.12	2.32	3.81	2.06	Negative	2.3655
frown	1.87	1.31	4.27	2.24	3.7	1.58	Negative	2.0212
grave	2.18	1.54	4.78	2.85	2.52	1.89	Negative	3.1274
guilt	2.14	1.33	5.36	2.95	2.96	1.83	Negative	2.8814
loser	2.25	1.48	4.95	2.57	3.02	2.17	Negative	3.1405
louse	2.81	1.92	4.98	2.03	3.57	2.26	Negative	1.9395
scorn	2.84	2.07	5.48	2.52	3.93	2.64	Negative	1.7782
trash	2.67	1.45	4.16	2.16	5.24	1.85	Negative	3.0596
chore	2.74	1.38	3.52	2.47	3.26	1.95	Negative	1.716
await	4.64	1.87	4.54	2.78	3.75	2.25	Neutral	2.2601
break	4.59	2.09	5.23	2.06	5.26	1.73	Neutral	4.0522
defer	4.1	1.54	4.48	2.37	4.76	2.44	Neutral	1.4914
brick	4.79	1.97	4.46	2.13	5.14	1.63	Neutral	2.716
glass	4.75	1.38	4.27	2.07	5	1.46	Neutral	3.4909
gloat	4.3	2.11	4.52	2.33	5.41	1.78	Neutral	1.9085

haste	4.15	1.7	5.04	2.49	4.19	1.86	Neutral	2.0294
issue	4.13	1.57	5.4	2.24	4.63	1.87	Neutral	3.241
rough	4.74	2	5.33	2.04	4.81	1.7	Neutral	3.2806
stiff	4.68	1.97	4.02	2.41	4.93	2.04	Neutral	2.7193
stool	4.56	1.72	4	2.14	4.98	1.85	Neutral	2.2553
wires	4.63	1.5	4.7	2.02	4.73	1.57	Neutral	2.5514
alive	7.25	2.22	5.5	2.74	6.39	2.15	Positive	3.8965
bathe	7.37	1.52	4.63	2.48	5.93	1.96	Positive	2.1106
bless	7.19	1.69	4.05	2.59	5.52	2.22	Positive	3.2526
bloom	7.21	1.42	4.96	2.03	5.36	1.77	Positive	2.4502
excel	7.45	1.55	5.45	2.16	7	1.75	Positive	1.5798
fairy	6.24	2.01	4.25	1.9	5.43	1.79	Positive	2.9304
hobby	7.24	1.48	5.36	2.59	7.24	1.83	Positive	2.5502
house	7.26	1.72	4.56	2.41	6.08	2.12	Positive	4.4185
loyal	7.55	1.9	5.16	2.42	6.91	2.23	Positive	2.7875
oasis	7.79	1.42	5.04	2.87	6.11	1.91	Positive	1.9956
truth	7.8	1.29	5	2.77	6.47	2.11	Positive	3.9913
unity	7.29	1.44	4.64	2.68	6.43	1.73	Positive	2.143

B2 Experiment 2 Wordlist (Valence Controlled Words)

Word	ValMn	ValSD	AroMn	AroSD	DomMn	DomSD	AroLvl	Lg10WF
bench	4.61	1.4	3.59	2.07	4.68	1.38	Low	2.6937
bland	4.1	1.08	3.29	1.89	4.88	1.27	Low	1.7482
blues	4.11	2.18	3.46	2.01	4.75	2.41	Low	2.7193
board	4.82	1.23	3.36	2.12	4.98	1.77	Low	3.5149
chalk	4.89	1.69	3.48	2.24	4.7	1.59	Low	2.2648
cloth	5.27	1.53	3.55	2.05	5.14	1.66	Low	2.4942
habit	4.11	1.77	3.95	2.11	4.3	1.79	Low	2.8686
paper	5.2	1.21	2.5	1.85	4.47	1.67	Low	3.722
penny	5.06	1.7	3.52	1.88	5.06	1.53	Low	3.0934
plain	4.39	1.46	3.52	2.05	4.71	1.68	Low	3.0469
slush	4.66	1.88	3.73	2.23	4.91	1.48	Low	1.4624
snail	4.31	1.67	3.86	2.27	5.68	2.04	Low	1.959
await	4.64	1.87	4.54	2.78	3.75	2.25	Medium	2.2601
brick	4.79	1.97	4.46	2.13	5.14	1.63	Medium	2.716
flock	5.5	1.22	4.5	1.63	4.8	1.27	Medium	2.3766
gloat	4.3	2.11	4.52	2.33	5.41	1.78	Medium	1.9085
heavy	3.69	1.38	4.58	1.93	4.1	1.62	Medium	3.3826
lower	4	1.58	4.36	1.91	4.18	1.63	Medium	3.1517
onion	4.43	1.91	4.6	2.27	4.97	1.16	Medium	2.3365

organ	5.24	1.57	4.54	1.86	4.96	1.75	Medium	2.5694
pinch	3.83	1.7	4.59	2.1	4.76	1.73	Medium	2.4955
plate	5.3	1.68	4	1.89	5.55	1.96	Medium	3.1169
spine	5.12	1.13	4.48	2.06	5.32	1.46	Medium	2.4683
stove	4.98	1.69	4.51	2.14	5.36	1.87	Medium	2.5888
chase	5.07	2.03	6.5	1.95	4.64	1.97	High	3.2238
clown	5.39	2.15	5.43	2.23	4.86	2.03	High	2.9074
crave	4.88	1.58	6.13	1.43	3.97	1.78	High	2.0294
drill	4.4	1.57	5.4	1.69	4.8	1.73	High	2.8463
erupt	5.43	1.57	6.57	2.14	3.8	1.97	High	1.3222
evade	4.47	1.83	5.6	2.24	5.53	2.27	High	1.6628
noisy	5.02	2.02	6.38	1.78	4.93	1.76	High	2.4116
rigor	4.96	1.82	5.57	2.5	5.18	2.13	High	1.6532
storm	4.95	2.22	5.71	2.34	4.54	2.04	High	3.1973
swear	4.43	1.45	5.46	2.05	4.79	1.79	High	3.6529
tense	3.56	1.36	6.53	2.1	5.22	2.02	High	2.7185
bribe	4.14	1.88	6	1.62	4.48	1.95	High	3.3918

B3 Practice Words								
Word	**ValMn**	**ValSD**	**AroMn**	**AroSD**	**DomMn**	**DomSD**	**Level**	**Lg10WF**
ton	4.68	1.47	4.33	1.98	5.04	2.17	Practice	2.5832
cab	4.53	1.07	3.97	1.88	4.27	1.93	Practice	3.2617
bus	4.51	1.57	3.55	1.8	4.84	1.75	Practice	3.578
fur	4.51	1.88	4.18	2.44	4.32	1.97	Practice	2.6263
ash	4.04	1.76	4.41	2.06	4.89	1.48	Practice	-

Legend

Word – name of the word to be presented[1]
ValMn - mean valence rating[1]
ValSD – standard deviation of valence rating[1]
AroMn – mean arousal rating[1]
AroSD – standard deviation of arousal rating[1]
DomMn – mean dominance rating[1]
DomSD – standard deviation of dominance[1]
Lvl/Level – the level category of the word for valence/arousal
Log10WF – is the log base 10 operation applied to the overall word frequency[2]

1. Measure from Bradley & Lang 2010
2. Measure from Brysbaert & New 2009

Note: all practice words are only used in practice trials to allow participants to gain familiarity with the study procedure and are not present elsewhere in the study.

B4 Experiment 1 Wordlist t-test Summary

Test	Statistic	t-statistic	df	p	Cohen's *d*	Lower 95% CI	Upper 95% CI
Negative - Neutral	ValMn	-17.077	22	< .001	-6.972	-2.264	-1.774
	AroMn	0.326	22	0.747	0.133	-0.37	0.509
	DomMn	-5.275	22	< .001	-2.154	-1.716	-0.747
	Lg10WF	-0.072	22	0.944	-0.029	-0.668	0.623
Negative - Positive	ValMn	-32.884	22	< .001	-13.425	-5.121	-4.514
	AroMn	-0.712	22	0.484	-0.291	-0.58	0.284
	DomMn	-10.163	22	< .001	-4.149	-3.217	-2.126
	Lg10WF	-0.579	22	0.568	-0.237	-0.907	0.511
Neutral - Positive	ValMn	-20.435	22	< .001	-8.342	-3.082	-2.514
	AroMn	-1.124	22	0.273	-0.459	-0.619	0.184
	DomMn	-6.477	22	< .001	-2.644	-1.901	-0.979
	Lg10WF	-0.527	22	0.603	-0.215	-0.868	0.516

B5 Experiment 2 Wordlist t-test summary

Test	Statistic	t-statistic	df	p	Cohen's *d*	Lower 95% CI	Upper 95% CI
Low - Moderate	ValMn	-0.113	22	0.911	-0.046	-0.468	0.419
	AroMn	-8.59	22	< .001	-3.507	-1.228	-0.75
	DomMn	-0.017	22	0.986	-0.007	-0.407	0.4
	Lg10WF	0.078	22	0.939	0.032	-0.463	0.499
Low - High	ValMn	-0.489	22	0.629	-0.2	-0.511	0.316
	AroMn	-14.401	22	< .001	-5.879	-2.809	-2.102
	DomMn	0.722	22	0.478	0.295	-0.237	0.491
	Lg10WF	0.16	22	0.874	0.065	-0.567	0.662
Moderate - Positive	ValMn	-0.313	22	0.757	-0.128	-0.56	0.413
	AroMn	-10.311	22	< .001	-4.21	-1.762	-1.172
	DomMn	0.594	22	0.559	0.242	-0.324	0.584
	Lg10WF	0.117	22	0.908	0.048	-0.492	0.551

Survey Questions and Responses (Appendix C)

Note: Experiment 1 frequencies are placed in square [box] brackets on the left, experiment 2 frequencies are placed in curly brackets {braces} on the right. All percentage frequencies are based on the parent row- thus the frequencies for each code are the percentage of responses within the response category itself rather than overall for the question. For example, in question 1, the frequencies for ***familiar*** *are the percentage of the* ***yes*** *responses that were categorized under the* ***familiar*** *code. In addition, if a participant gave a response that fit within two categories, then they were coded under both. This means that the sum of codes may be greater than the total number of responses in the parent category.*

Q1. In the experiment, do you think you recalled some words more easily than others? (Yes, No)

No [1|2.44%] {3|4.48%%} ; **Yes** [40|97.56%] {64|95.52%}

- If "yes", please explain why you think you recalled some words more easily than others.
    - ***Familiar*** – Words are interesting, arousing, or appealing to individual

      [6|14.63%] {11|15.94%%}
    - ***Frequency*** – Words that are more common or repeated in daily lives

      [4|9.76%] {5|7.25%%}
    - ***Interconnected*** – Words that are related to other presented words

      [7|17.07%] {9|13.04%}

- ***Relatable*** – Words that are related to an individual through memories or events

  [13|31.71%] {32|46.38%}

- ***Other*** – Any other reason words may be memorable

  [6|14.64%] {10|14.49%}

- ***Blank***

  [5|12.20%] {2|2.90%}

Q2. Do you believe that words were remembered better when you spent more time studying them? (Yes, No)

No [13|31.71%] {25|37.31%} ; **Yes** [28|68.29%] {42|62.69%}

- If "yes", please describe what made you decide to study some words for longer than others.

  - ***Familiarity*** – Words are studied based on being interesting, arousing, or appealing

    [3|10.71%] {6|14.29%}

  - ***Frequency*** – Study of words based on how common or repeated in daily lives words are

    [5|17.86%] {7|16.67%}

  - ***Rehearsal*** – Words are studied until individual feels they are remembered

    [10|35.71%] {12|28.57%}

  - ***Relatable*** – Study of words based on how related to an individual they are

    [5|17.86%] {5|11.90%}

  - ***Other*** – Any other control strategy

[5|17.86%] {11|26.19%}

- ***Blank***

  [0|0%] {1|2.38%}

- How did you decide when to move on to the next item?
  - ***Feeling*** – A feeling or sensation of confidence in memorability for a word (implicit feeling) determined when participant was finished studying an item

    [6|14.63%] {9|13.43%}

  - ***Remembered*** – Participant concluded study when they felt or judged that a word was learned (Explicit judgement)

    [14|34.15%] {16|23.88%}

  - ***Strategy*** – Participant moved on after employing a specific learning strategy or other routine (for example- reading the word three times and then moving on)

    [18|43.90%] {34|50.75%}

  - ***Other*** – Participant used a different strategy to determine when to conclude their studying

    [3|7.32%] {7|10.45%}

  - ***Blank***

    [0|0%] {1|1.49%}

Q3. Do you believe your predictions of future recall were accurate? (Yes, No)

No [30|73.17%] {44|65.67%} ; **Yes** [11|26.83%] {23|34.33%}

- Please explain your response to the above question

No

- **Skill or Confidence** – Participant judged that their JOLs were accurate or inaccurate due to a lack of confidence or skill, or previous poor performance

  [21|70.00%] {38|88.36%}

- ***Study Time*** – Participant judged the accuracy of their future recall based on their study time of the words

  [1|3.33%] {2|4.55%}

- ***Other*** – Participants judged their JOL-Recall accuracy based on some other reason not already covered

  [7|23.33%] {4|9.09%}

- ***Blank***

  [1|3.33%] {0|0%}

Yes

- **Skill or Confidence** – Participant judged that their JOLs were accurate or inaccurate due to a lack of confidence or skill, or previous poor performance

  [6|54.55%] {17|73.91%}

- ***Study Time*** – Participant judged the accuracy of their future recall based on their study time of the words

  [1|9.09%] {1|4.35%}

- ***Other*** – Participants judged their JOL-recall accuracy based on some other reason not already covered

[4|36.06%] {3|13.04%}

- *Blank*

  [0|0%] {2|8.70%}

Q4. Did you notice anything about the content of the words being presented? (Yes, No)

No [18|43.90%] {50|74.63%} ; **Yes** [23|56.10%] {17|25.37%}

- If "yes", please explain your observation
  - ***Arousal*** – Participant noticed that words varied in arousal or exactingness

    [0|0%] {0|0%}

  - ***Concreteness*** – Participant noticed the concreteness or abstractness of words

    [1|4.35%] {2|11.76%}

  - ***Frequency*** – Participants noticed that some words stood out due to their commonness or uniqueness/uncommonness

    [1|4.45%%] {2|11.76%}

  - ***Grouping*** – Participant noticed that the presented words were easily grouped with each other to form sentences or stories

    [4|17.39%] {7|41.18%}

  - ***Valence*** – Participant noticed that words varied in valence

    [12|52.17%] {1|5.88%}

  - ***Other*** – Participant noticed something not previously mentioned about the wordlist

    [4|17.39%] {4|23.53%}

- ***Blank***

  [1|4.35%] {1|5.88%}

Q5. In general, do you believe that certain words are easier (or harder) to remember than others? (Yes, No)

No [5|12.20%] {3|4.48%} ; **Yes** [36|87.80%] {64|95.52%}

- If "yes", please explain why you think certain words vary in their memorability.
  - ***Concreteness*** – Memorability has to do with how concrete/abstract a word is

    [4|11.11%] {6|9.09%}
  - ***Emotion*** – Word memorability varies with emotional content

    [5|13.89%] {6|9.09%}
  - ***Frequency*** – Words are more memorable based on how often they are used

    [11|30.56%] {20|30.30%}
  - ***Relatable*** – Relatable words are more memorable

    [14|38.89%] {22|33.33%}
  - ***Other*** – Words vary in their memorability for some other reason that is not listed

    [2|5.56%] {9|13.85%}
  - ***Blank***

    [0|0%] {3|4.55%}

Q6. Did you notice any emotional or exciting words in your list(s)? (Yes – Both, Yes - Emotional, Yes – Exciting, No – Neither emotional or exciting)

No [5|12.20%] {30|44.78%} ; **Yes** [36|87.80%] {37|55.22%}

"Yes" subcategories

- ***Both emotional and exciting***

  [22|61.11%] {17|45.95%}

- ***Emotional***

  [13|36.11%] {12|32.43%}

- ***Exciting***

  [1|2.78%] {8|21.62%}

- ***Blank***

  [0|0%] {0|0%}

Q7. Do you believe that emotional or exciting words are remembered differently than neutral words? (Yes/No)

No [4|9.76%] {6|8.96%} ; **Yes** [37|90.24%] {61|91.04%}

Q8. Do you believe that emotional words are remembered differently (e.g., "Happy" vs. "Sad")? (Yes, No)

No [5|12.20%] {13|19.40%} ; **Yes** [36|87.80%] {54|80.60%}

- If "yes", please explain how you think different emotional words vary in their memorability.

  - ***Current Emotional State*** – Emotional words resonate with current emotional state

    [5|13.89%] {6|11.11%}

- ***Immediate Emotional Response*** – Word affects or changes an individual's current emotional state (they feel the valence of the presented word)

  [7|19.44%] {5|9.26%}

- ***Relatable*** – Emotional words are relatable to life- memories, environment, past

  [13|36.11%] {20|37.04%}

- ***Other*** – Emotional words vary in some other way

  [8|22.22%] {10|18.52%}

- ***Blank***

  [3|8.33%] {13|24.07%}

Q9. Do you believe that exciting words are remembered differently? (Yes, No)

No [11|26.83%] {16|23.88%} ; **Yes** [30|73.17%] {51|76.12%}

- If "yes", please explain how you think exciting words vary in their memorability.
  - ***Engagement*** – Exciting/boring words are more engaging, stimulating, or interesting

    [6|20.00%] {13|25.00%}

  - ***Relatable*** – Exciting/boring words are relatable to life- memories, environment, past

    [8|26.67%] {11|21.15%}

  - ***Saliency*** – Exciting/boring words stand out

    [4|13.33%] {7|13.46%}

  - ***Valence*** – Exciting/boring words differ in valence

[3|10.00%] {3|5.77%}

- ***Other*** - Exciting/boring vary in some other way not listed

  [5|16.67%] {7|13.46%}

- ***Blank***

  [4|13.33%] {11|21.15%}

Q10. Do you believe that exciting words take more or less study time to remember? (More, Less, No)

No [7|17.07%] {15|22.39%} ; **More** [4|9.76%] {2|2.99%} ; **Less** [30|73.17%] {50|74.63%}

➢ If "more" or "less", please describe how you believe these words should be studied.

  More

  - ***Associations*** – Study time based on the associations formed between the studied word and other words or thoughts

    [3|75.00%] {0|0%}

  - ***Familiarity*** – Study time based on familiarity or feeling of memorability

    [0|0%] {0|0%}

  - ***Immediate Emotional Response*** – Study time based on the emotional reaction produced by engaging with the word

    [0|0%] {0|0%}

  - ***Relatable*** – Study time based on how relatable a word is

    [0|0%] {0|0%}

  - ***Same***

[0|0%] {0|0%}

- ***Other*** – Study time is based on something else

  [0|0%] {1|50.00%}

- ***Blank***

  [1|25.00%] {1|50.00%}

Less

- ***Associations*** – Study time based on the associations formed between the studied word and other words or thoughts

  [1|3.33%] {4|8.00%}

- ***Familiarity*** – Study time based on familiarity or feeling of memorability

  [4|13.33%] {3|6.00%}

- ***Immediate Emotional Response*** – Study time based on the emotional reaction produced by engaging with the word

  [2|6.67%] {4|8.00%}

- ***Relatable*** – Study time based on how relatable a word is

  [9|30.00%] {11|22.00%}

- ***Same***

  [2|6.67%] {1|2.00%}

- ***Other*** – Study time is based on something else

  [9|30.00%] {16|32.00%}

- ***Blank***

  [3|10.00%] {11|22.00%}

Q11. Do you believe that emotional words take more or less study time to remember? (More, Less, No)

No [11|26.83%] {16|24.24%} ; **More** [4|9.76%] {2|3.03%} ; **Less** [26|63.41%] {48|72.73%}

More

- ***Associations*** – Study time based on the associations formed between the studied word and other words or thoughts

  [0|0%] {0|0%}

- ***Familiarity*** – Study time based on familiarity or feeling of memorability

  [1|25.00%] {0|0%}

- ***Immediate Emotional Response*** – Study time based on the emotional reaction produced by engaging with the word

  [0|0%] {0|0%}

- ***Relatable*** – Study time based on how relatable a word is

  [2|50.00%] {1|50.00%}

- ***Same***

  [0|0%] {0|0%}

- ***Other*** – Study time is based on something else

  [1|25.00%] {1|50.00%}

- ***Blank***

  [0|0%] {0|0%}

Less

- ***Associations*** – Study time based on the associations formed between the studied word and other words or thoughts

  [1|3.85%] {5|10.42%}

- ***Familiarity*** – Study time based on familiarity or feeling of memorability

  [0|0%] {1|2.08%}

- ***Immediate Emotional Response*** – Study time based on the emotional reaction produced by engaging with the word

  [5|19.23%] {7|14.58%}

- ***Relatable*** – Study time based on how relatable a word is

  [11|42.31%] {15|31.25%}

- ***Same***

  [2|7.69%] {2|4.17%}

- ***Other*** – Study time is based on something else

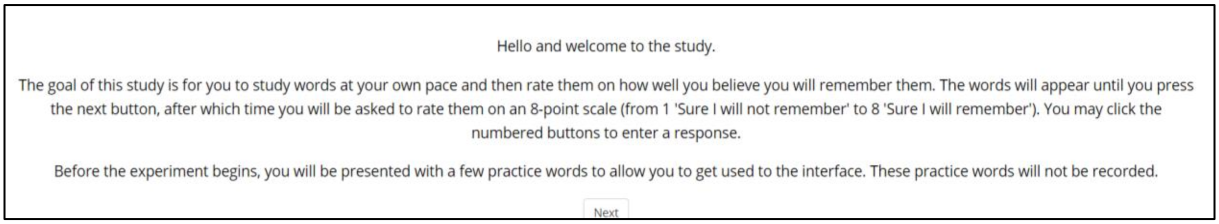
  [4|15.38%] {7|14.58%}

- ***Blank***

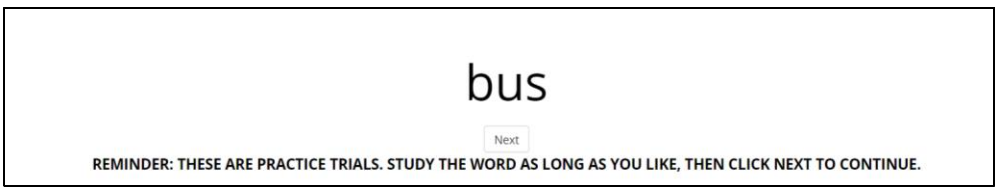
  [3|11.54%] {11|22.92%}

Survey Demonstration Screenshots (Appendix D)

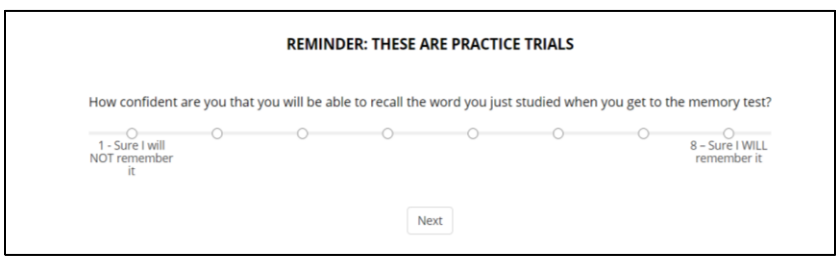
A. Introduction screen



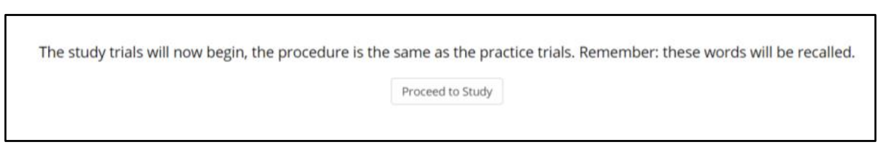
B. Presentation of a practice word



C. Practice JOL rating

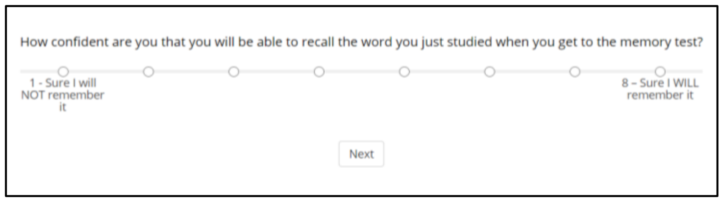


D. Transition from practice words to study words

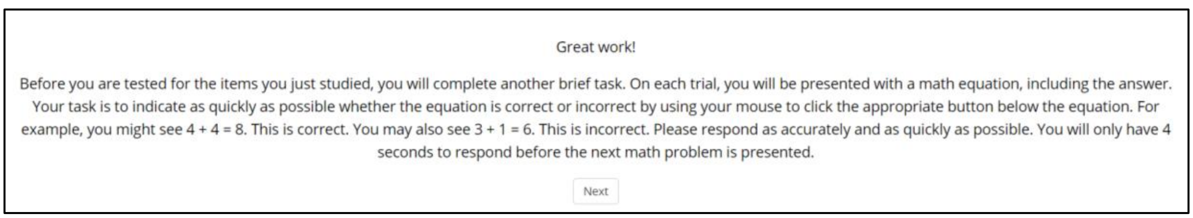


E. Example of the presentation of a study word

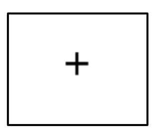
F. JOL rating screen



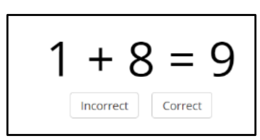
G. Transition to retention interval task



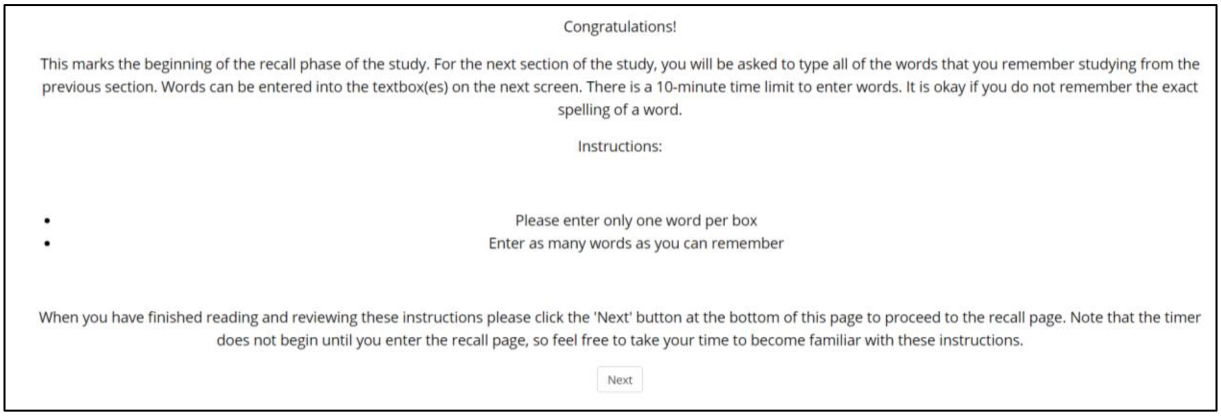
H. Brief presentation of fixation cross



I. Example math question used in fixation interval



J. Transition to free-recall task



Congratulations!

This marks the beginning of the recall phase of the study. For the next section of the study, you will be asked to type all of the words that you remember studying from the previous section. Words can be entered into the textbox(es) on the next screen. There is a 10-minute time limit to enter words. It is okay if you do not remember the exact spelling of a word.

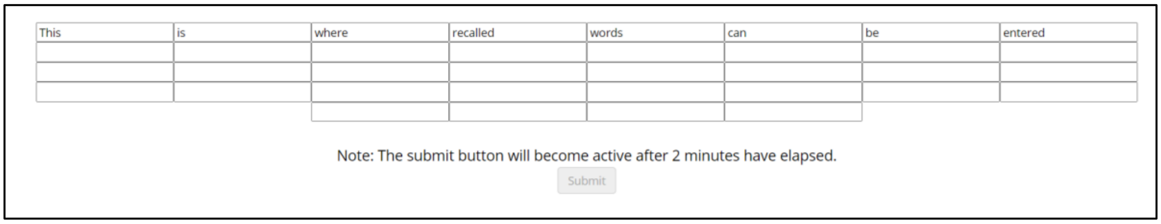
Instructions:

- Please enter only one word per box
- Enter as many words as you can remember

When you have finished reading and reviewing these instructions please click the 'Next' button at the bottom of this page to proceed to the recall page. Note that the timer does not begin until you enter the recall page, so feel free to take your time to become familiar with these instructions.

Next

K. Example of the recall grid



This	is	where	recalled	words	can	be	entered

Note: The submit button will become active after 2 minutes have elapsed.

Submit

Gamma Correlation Descriptive Statistics (Appendix E)

E1 Experiment 1 Gamma Correlations

		Block 1					Block 2				
					95% CI					95% CI	
Relationship	**Valence**	**N**	**Mean**	**SD**	**Lower**	**Upper**	**N**	**Mean**	**SD**	**Lower**	**Upper**
JOL x Recall	Negative	35	0.278	0.684	0.052	0.505	39	0.363	0.56	0.188	0.539
	Neutral	38	0.138	0.611	-0.056	0.332	39	0.349	0.532	0.182	0.516
	Positive	37	0.359	0.532	0.188	0.53	40	0.307	0.565	0.132	0.482
JOL x Study Time	Negative	39	-0.039	0.311	-0.137	0.058	40	0.047	0.309	-0.048	0.143
	Neutral	40	0.057	0.376	-0.06	0.173	40	-0.02	0.292	-0.11	0.071
	Positive	39	0.104	0.364	-0.01	0.218	40	-0.043	0.421	-0.174	0.087
Recall x Study Time	Negative	38	0.002	0.392	-0.123	0.127	41	0.048	0.475	-0.097	0.194
	Neutral	36	0.155	0.507	-0.01	0.321	40	0.008	0.423	-0.123	0.139
	Positive	39	0.098	0.476	-0.051	0.247	39	-0.017	0.435	-0.153	0.119

E2 Experiment 2 Gamma Correlations

		Block 1					Block 2				
					95% CI					95% CI	
Relationship	**Arousal**	**N**	**Mean**	**SD**	**Lower**	**Upper**	**N**	**Mean**	**SD**	**Lower**	**Upper**
JOL x Recall	Low	63	0.181	0.621	0.028	0.335	66	0.431	0.562	0.295	0.566
	Moderate	63	0.282	0.65	0.122	0.443	65	0.325	0.601	0.179	0.471
	High	65	0.2	0.661	0.04	0.361	67	0.437	0.486	0.32	0.553
JOL x Study Time	Low	67	0.063	0.303	-0.009	0.136	67	0.062	0.357	-0.023	0.147
	Moderate	67	0.082	0.315	0.007	0.158	66	-0.029	0.354	-0.115	0.056
	High	67	-0.049	0.313	-0.124	0.026	67	-0.026	0.274	-0.091	0.04
Recall x Study Time	Low	63	0.111	0.48	-0.008	0.23	66	-0.001	0.437	-0.107	0.104
	Moderate	63	-0.023	0.456	-0.136	0.09	66	-0.034	0.448	-0.142	0.074
	High	65	-0.003	0.526	-0.131	0.125	67	0.042	0.425	-0.06	0.144

Beliefs Interrater Reliability (Appendix F)

F1 Experiment 1 Beliefs Interrater Reliability

KAPPA	**Q1**	**Q2AB**	**Q2C**	**Q3**	**Q4**	**Q5**	**Q8**	**Q9**	**Q10**	**Q11**
Subjects	35	28	41	39	22	34	33	26	30	27
Raters	2	2	2	2	2	2	2	2	2	2
Agreement %	74.286	60.714	75.61	89.744	72.727	91.176	60.606	65.385	50	70.37
Kappa	0.663	0.511	0.632	0.802	0.609	0.877	0.479	0.566	0.361	0.57
z	7.942	6.207	6.252	7.653	5.532	8.711	4.967	5.97	4.155	5.236
p-value	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001

F2 Experiment 2 Beliefs Interrater Reliability

KAPPA	**Q1**	**Q2AB**	**Q2C**	**Q3**	**Q4**	**Q5**	**Q8**	**Q9**	**Q10**	**Q11**
Subjects	60	40	65	64	16	60	39	39	39	38
Raters	2	2	2	2	2	2	2	2	2	2
Agreement %	55	62.5	70.769	75	68.75	86.667	58.974	69.231	48.718	47.368
Kappa	0.385	0.536	0.53	0.573	0.585	0.819	0.399	0.585	0.355	0.347
z	6.555	7.956	7.039	7.066	4.523	11.351	4.761	6.912	4.905	4.721
p-value	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001