STAGE-SPECIFIC MECHANISMS OF DESIRABLE DIFFICULTY

ISOLATING STAGE-SPECIFIC MECHANISMS OF DESIRABLE DIFFICULTY IN LEARNING

By MELISSA PTOK, B.SC. (hons.)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the
Requirements for the Degree Master of Science

MASTER OF SCIENCE (2015)
McMaster University
Psychology, Neuroscience, and Behaviour

TITLE: Isolating stage-specific mechanisms of desirable difficulty in learning

AUTHOR: Melissa Jeanine Ptok, B.Sc. (hons.) (University of Guelph)

SUPERVISOR: Dr. Scott Watter

NUMBER OF PAGES: vii, 33

Abstract

According to the theory of desirable difficulty, conflict during practice can elicit a greater degree of processing and result in a later memory benefit (Bjork & Bjork, 1994). The present paper extends the work by Thomson et al. (in prep) that there may be a stage-specific mechanism involved. They found a desirable difficulty effect when directing a conflict towards the categorization stage through classifying names by gender. However, no such effect was seen when classifying words by size. They provided evidence that gender classification is more semantically central than categorizing items by size, which tends to be relative and depend on context. We took the same stimuli used in the Thomson et al. (in prep) but had participants make animacy judgments (animal or "thing") on words. A subsequent memory test revealed a desirable difficulty effect for incongruent words compared to congruent. Interestingly, animal words were better remembered overall compared to inanimate object words. A second experiment directed a conflict towards the categorization and response processing stages through classifying names as male and female with semantic (male/female) and response selection (left/right) primes. A subsequent memory test revealed a desirable difficulty pattern of results (although non-significant) where incongruent compared to congruent words were better remembered for the sematic primes and congruent compared to incongruent words were better remembered for response primes. These results suggest to-be-remembered material needs to be the focus of attention and increasing difficulty to any stage of processing does not give you a guaranteed desirable-difficulty effect.

Acknowledgments

My first year of graduate school has been everything I imagined: challenging, rewarding and above all, inspiring. First, I would like to express my sincere appreciation to my supervisor Dr. Scott Watter for your constant guidance and encouragement throughout the past year. Your patience and enthusiasm have pushed me to accomplish more than I thought possible within this time. I am excited and grateful that I get to continue working with you.

I would also like to thank my committee members, Dr. Judy Shedden and Dr. Karin Humphreys for their support and insightful input throughout the project. Judy, working in your lab continues to challenge me and I look forward to learning from you in the coming years.

Finally, I would like to thank my loved ones, who have supported me over the years. To my parents, Dianne and Ed, you have been my inspiration throughout life. Thank you for teaching me how to believe in myself and for your unwavering belief in what I am capable of. Dad, your steady guidance, encouragement and positive outlook on life has enabled me to reach for the stars. To Stephane, thank you for keeping me harmonious throughout this process by filling my life with an immeasurable amount of joy and happiness.

Table of Contents

List of Figures

INTRODUCTION

The ability of the human cognitive system to adapt to task demands through recruitment, modulation and regulation of cognitive processes is critical for performance and learning (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Collectively, these processes are known as *cognitive control* and have gained much attention in the field of psychology. The theory of cognitive control has allowed us to understand that human behavior is flexible, in that it can adapt given changes in context, which is an ability beneficial to everyday psychological functioning (Morton, Ezekiel & Wilk, 2011). The Stroop task studies are a common example of cognitive control (see MacLeod, 1991). In this task, participants are presented with "colour words" that are printed in different colours (i.e., the word *blue* printed in red). Here, participants are asked to respond to the font colour of the word. Cognitive control is used to suppress the tendency to read the word rather than indicating the colour of the print. A participant with a lack of cognitive control would read aloud the word rather than report the colour of the print.

There have been important findings that characterize the functioning of cognitive control. These offer an account of how the system determines when control is required, how it is modulated, and when it is no longer needed (Botvinick et al., 2001). Through the Stroop task, there is evidence that the recruitment of cognitive control processes begins as we attempt to perform a difficult task. This is evident as there is greater interference during the first few trials in each block compared to later trials (Henik, Bibi, Yanai & Tzelgov, 1997). Cognitive control appears to adjust to task demands, which can be seen through a rise in reaction time and accuracy following errors (e.g. Laming, 1968). Additionally, there appears to be less interference on frequent incongruent trials (when the print colour and word colour do not match) compared to when they appear less often (Lindsay & Jacoby, 1994). We also see that the cognitive control system withdraws as the need for control diminishes. This becomes evident as performance on an initially difficult task becomes increasingly automatic.

Based on these findings Botvinick et al. (2001) proposed a regulative and evaluative dimension of control that must be present to monitor and make assessments of task demands. This is known as the conflict-monitoring hypothesis of cognitive control. This model proposes that when conflict is detected between the print colour and colour word during an incongruent Stroop trial, a top-down increase in attention is directed toward the relevant information in order to reduce interference from the task-irrelevant information (i.e. the colour and colour word respectively). This increased demand for control of selective attention with increased conflict creates a congruency effect, where reaction times are slower for incongruent (e.g. the word *blue* printed in red) than for congruent trials (e.g. the word *blue* printed in blue) trials (Stroop, 1935).

Recent research has investigated this increase in selective attention directly and has reported a later memory benefit for incongruent items compared to congruent items on a surprise recognition memory test (Krebs, Boehler, Belder, & Egner, 2015; Rosner, D'Angelo, MacLellan & Milliken, 2014). These results are at odds with the typical divided attention prediction that increased processing demands will lead to a general decline in performance.

This idea that introducing difficulty during practice can benefit later memory stems from Bjork & Bjork's (1992) new theory of disuse, which suggested that our memories might be predictable through storage strength (how well learned a memory item is) and retrieval strength (how accessible the item is). Based on this theory, the more storage strength a memory trace has, the bigger boost in retrieval strength it will have. The amount of newly acquired information declines over time without attempt to retain it. Bjork & Bjork (1992) believed that we learn better by spacing out and interleaving multiple topics over a period of time rather than learning one topic all at once (massed presentation). This leads to sustained retention over time rather than having good initial performance but poor retention over time. Therefore, the interference-producing distractors that make initial learning difficult can actually enhance encoding and facilitate memory retention (Battig, 1972).

These difficulties that can have an enhanced effect on long-term retention are called *desirable difficulties* (Bjork, 1994). Studies involving varied learning conditions have found that making initial learning difficult can actually enhance long-term retention (Bjork & Bjork, 2011). Diemand-Yauman et al. (2010) suggest that these memory benefits occur as difficulty during learning produces more cognitive engagement and deeper processing, leading to enhanced encoding and subsequent retrieval. There have been some well-documented examples of times where difficulty experienced during initial leaning appears to be desirable for learning. These include those involving interleaving rather than grouping study topics (Kornell & Bjork, 2008), studies with initial encoding difficulties in the form of perceptual interference (Hirshman & Mulligan, 1991) and hard to read fonts (Diemand-Yauman, Oppenheimer & Vaughan, 2011).

One particularly interesting example of this was presented through a series of experiments conducted by Rosner, D'Angelo, MacLellan and Milliken (2014). They provide evidence for how spatially interleaved words appear to increase selective attention demands and produce a desirable difficulty effect on memory. In these experiments, the stimuli of two interleaving words were presented to the participant. One of the words was presented in red and the other in green; they were instructed to read the red word aloud while ignoring the green word. The green word was identical to the red word on half of the trials (congruent trials) and different on the other half of the trials (incongruent). A congruency effect was found for word naming where reading the word for incongruent trials was slower than for congruent trials. However, better memory was found for incongruent stimuli. More recent studies have suggested that this improved later memory for incongruent trials is due to an increase in selective attention demands for incongruent items rather than longer processing time for these items (Rosner & Milliken, 2015).

Other studies of divided attention suggest that a transient increase in attention to one task can enhance the performance on a second task. This phenomenon is known as the attentional boost effect (ABE) (e.g. Jiang & Swallow, 2010; Mulligan, Spataro & Picklesimer, 2014). This theory proposes that in order for difficulty manipulations to produce a later memory benefit, the difficulty must increase selective attention to the to-be-remembered information. Therefore, when to-be-encoded information is the recipient of the enhanced attention, an improvement of later memory will be seen. This was demonstrated in a study conducted by Jiang & Swallow (2010). Participants were presented with a sequence of pictures that each had a small square at the center of the image. In the divided attention condition, participants were instructed to monitor the colour of the square in the center of the screen while memorizing the images for a later memory test. Participants were instructed to press the space bar whenever they detected an infrequent white square (target) and to withhold a response to the frequent black squares (distractor). In the full attention condition, participants were told to memorize the images while ignoring the squares that appeared in the center of the image. Memory was later tested with a four choice recognition task. In the divided attention condition, the images that were encoded with the infrequent white square (target) were better remembered compared to the images that were encoded with the frequent black square (attentional boost effect) and seemed to eliminate the divided attention effect. In the full attention condition, where the participants made no response to the squares, no attentional boost effect was found.

All of these previous studies involving congruency and an enhancement of memory have involved a top-down model investigating task-wide changes. Thomson, Ptok, Humphreys & Watter (in preparation) have considered high-level attentional demands at different stages of processing to investigate whether desirable difficulty effects are task-general or stage-dependent. They presented a series of experiments that looked at difficulty manipulations during different stages of processing. Their prediction was for a stage-specific account that should promote desirable difficulty effects when a manipulation enhances selective attention toward semantic representations of target stimuli.

They conducted three experiments. The first was a psychological refractory period paradigm (PRP) study where responses were made for two tasks and then followed by a later memory test. Their PRP paradigm involved two simple tasks, each with the same left/right response alternatives. Thus, each trial could be either response-incongruent (the response to the second task was opposite the response of the first task) or response-congruent (the response to the second task was the same as the response of the first task). Prior work has demonstrated response priming from unattended Task 2 response activation to attended Task 1 performance (typically known as the Backward Compatibility

Effect), allowing manipulation of response compatibility priming on Task1 from a separate and semantically unrelated second task. In the present task, incongruent trials lead to slower Task1 performance, as in Rosner et al. (2014), but in contrast produced worse memory performance at subsequent test. This is inconsistent with the desirable difficulty effect. They argue that rather than assuming manipulations in attention will facilitate encoding wherever extra processing is required, only difficulties that increase selective attention at key stages of processing (representation of the meaning of the to-be-recalled stimulus) within a single task will lead to memory benefits (Thomson et al., in prep.).

The second and third experiments were designed to determine whether congruency differences on tasks with category or semantic primes could be manipulated to create desirable difficulties. During the second experiment, two words were presented separated by an SOA (stimulus onset asynchrony) of either 150 ms or 700 ms. Participants were required to make a speeded response to the first of two words by categorizing the first stimulus (T1) as either “big” or “small” relative to the computer screen, while ignoring the second stimulus (T1 prime) which were the words either *big* or *small.* Incongruent trials occurred when the judgment of the first word did not match the second stimulus word. A congruent trial occurred when the judgment of the first word matched the second stimulus word. A later memory test was presented to test participants’ memory of the T1 stimuli. Overall, typical congruency effects were observed where the reaction times for incongruent items were slower than for congruent items. Results for the memory phase of the experiment were split into fast T1 and slow T1 reaction times as it was expected that during fast T1 reactions times, participants were more focused and thus less susceptible to prime information. With this division they observed a memory benefit in congruent trials for the slow T1 reaction times. This is evidence for a divided attention cost rather than an example of a desirable difficulty. It was predicted that a desirable difficulty effect would be observed with this semantic priming but they argued that perhaps the big/small task was not automatic enough and too relative to this kind of desirable difficulty priming.

Experiment 3 was conducted as a second test of semantic category priming and desirable difficulty, and asked participants to make gender classifications on typical male and female names. Thomson et al. (in prep.) argued that gender was a strong semantic feature of names, and should not require additional evaluative processing in order to make this category judgment. This experiment was set up in the same way as Experiment 2, with T1 involving classifying names as either male or female, and the words *male* or *female* presented as T1 prime stimuli. For stimuli with slow T1 reaction times at study, better memory was observed for words with incongruent primes, reflecting a desirable difficulty effect. By eliciting more difficult initial processing in T1 categorization performance, an enhancement in long-term memory was present.

In general, difficulties usually do not lead to better long-term retention, but rather a memory cost due to enhanced processing demands. Finding a desirable difficulty effect is dependent on applying attentional resources to a particular stage of processing. The results from these studies offer evidence for how allocating attention to a particular stage of processing is beneficial to the learning process. In Thomson et al. (in prep.), categorizing by gender compared to relative object size may have allowed participants to focus more on the meaning of the task stimulus to resolve categorization conflict, versus engaging in a subsequent cognitive size comparison task at the expense of the meaning of the stimulus. The increase in selective attention to the central processing stage or representation experiencing high conflict here predicts enhanced memory encoding for gendered names, but a decrease in encoding for big/small objects when attention is focused on the subsequent size comparison task.

The goal of the present study is to further investigate the effect of desirable difficulties in order to elucidate its mechanism of action. In particular, two experiments were conducted in order to determine the type of tasks that could be manipulated in order to generate a process-specific increase in attention to produce the desirable difficulty memory effect.

Experiment 1 was designed to test whether judgments of animacy on the same concrete nouns from Experiment 2 of Thomson et al. (in prep.) (asking participants to classify a word as an animal or inanimate object) would produce the same congruency effect on enhanced long-term memory during a congruency priming manipulation. It was hypothesized that incongruent trial words would be more difficult to process and thus subject to worse immediate performance. A later memory test should reveal that incongruent words are better remembered if the conflict promotes an increase in attentional allocation to the core meaning of the word. This is compared to giving participants an equally cognitive demanding task (e.g. categorizing objects as relatively “big” or “small” compared to a reference object) when the attentional focus may not be directed towards an essential semantic feature of the word.

Experiment 2 was designed to introduce a situation in which two stages of processing are targeted. This study will also build upon results of Experiment 3 by Thomson et al. (in prep.), where male and female names were better remembered in *incongruent* trials compared to congruent trials when given a semantic categorization prime. Additionally, in their Experiment 1, better memory was found for *congruent* trials when a response selection stage of processing was primed. In this experiment, a combination of both semantic priming and response priming for male and female names was used to investigate the desirable difficulty mechanism. It is hypothesized that incongruent trials will have longer reaction times compared to congruent trials for both judgment tasks, as these items will present increased difficulty at initial performance. It is also hypothesized that incongruent trials primed with the semantic representation of the word will be better remembered compared to congruent trials; a desirable difficult effect will be present during these trials. Additionally, items primed with manual response information for the word will direct attentional focus to the response conflict during incongruent trials, rather than the meaning of the word, eliciting a typical divided attention memory cost.

Experiment 1: Animacy Judgment of Words

Methods

Participants. Twenty-five undergraduate students from McMaster University participated in a one-hour experimental session, in exchange for partial course credit. All participants spoke English fluently based on results of a questionnaire administered before the commencement of the experiment. One participant was excluded for demonstrating low accuracy during the first phase of the experiment, leaving 24 participants for reported data analyses.

Apparatus and Stimuli. The experiment was conducted on a 19-inch Samsung CRT monitor and presented in the program Presentation on Windows XP. Stimuli were presented as white text on black background.

Priming phase. A standard PC keyboard was used where the “Z” and “/” keys were used for response mapping. During each trial, two stimuli were presented. The top stimulus (T1) always appeared first. The second stimulus (T1 prime) appeared below the top stimulus at a delay of either 5ms stimulus onset asynchrony (SOA) or 600ms SOA. A total of 160 nouns were used for this phase of the experiment.

Memory phase. One stimulus was presented per trial where the keys “Z” and “/” were used for response mapping. A total of 160 nouns were used for this phase of the study where 80 were old words from phase 1 and 80 were new words.

Procedure

Priming phase. Participants sat comfortably in front of the computer where T1 (e.g. the word “dog”) was presented. At 5ms SOA or 600ms SOA, T1 prime (either the word “animal” or “thing”) appeared below T1. Participants were required to use the keys “Z” or “/” to make a speeded response regarding whether T1 was an “animal” or a “thing”. Response mapping was counterbalanced across participants. They were told to ignore the T1 prime stimuli and respond to the T1 stimuli as fast and accurately as possible. Response times (RTs) were recorded from the onset of the T1 word to the onset of the response, as detected by the keyboard. The priming phase included a practice block (which was consistent with the test blocks) of 40 trials, where participants could get comfortable with the task. An instruction page was presented at the beginning of both the practice and test blocks. There were four experimental blocks, each with 40 trials, for a total of 160 trials. Within the 160 trials, half were accompanied with a congruent prime and half with an incongruent prime, in random order. After each block, the participant was given a rest period. This phase of the experiment took participants between 15-20 minutes to complete.

Memory Phase. A surprise memory test was administered after the phase 1 blocks to test the learning for the T1 words. Participants were not aware of this test during the priming phase. A word was presented in the middle of the screen and the participant needed to indicate whether the word was "new" or "old" in relation to the first phase of the experiment. Old words came from T1 animal/thing stimuli of the priming phase and new words were not presented at any time during phase 1. The stimuli remained on screen until participants responded by pressing "Z" for old words and "/" for new words. The memory phase consisted of 4 blocks of 40 trials in each for a total of 160 trials. Of the 160 trials, 80 were words from the priming phase of the experiment and 80 were new words. Breaks were offered in between each block. This phase of the experiment took participants between 5-10 minutes to complete.

Results

Results for Experiment 1 mean reaction time (RT) and mean accuracy for the initial priming phase, and mean accuracy for the subsequent memory phase, are presented below.

Priming Phase.

Reaction Time. Mean reaction times for T1 stimuli were computed for each of the four conditions and are presented in Figure 1. A two-tailed paired sample t-test revealed that responses were significantly slower for incongruent items (615ms) than for congruent items (598 ms) at the 5 ms SOA (i.e. an RT priming effect was observed), $t(23)= 2.07$, $p= 0.05$. There was no significant difference in RT values between incongruent (614 ms) and congruent (618 ms) items in RT values at the 600 ms SOA, $t(23)= 0.65$, $p=0.52$.

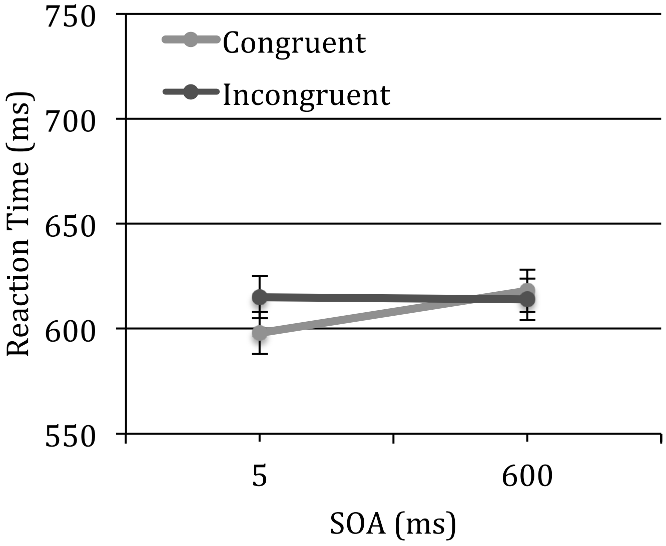


Figure 1. Experiment 1: Mean reaction times for congruent and incongruent T1s in the priming phase.

Accuracy. The mean accuracy scores within the priming phase were calculated for each of the 4 conditions to ensure a high accuracy. A two-tailed paired sample t-test revealed that there were no significant differences between the incongruent (0.95) and congruent (0.96) conditions at the 5 ms SOA, *t*(23)= .81, *p*= 0.42. The same is also observed between congruent (0.95) and incongruent (0.95) conditions at the 600 ms SOA *t*(23)= .53, *p*= 0.60.

Memory Phase

Accuracy.

The proportion of items remembered from the first phase of the experiment served as the dependent variable in a 2 x 2 repeated measures ANOVA that treated congruency (congruent, incongruent) and SOA (5ms, 600ms) as factors. These data are presented in Figure 2. The analysis revealed a significant main effect of congruency, *F*(1,23) = 6. 20, *p* =

0.02, $\eta_p{}^2 = .22$, indicating that incongruent words (0.75 at 5ms SOA, 0.78 at 600ms SOA) were remembered better overall than congruent words (0.71 at 5ms SOA, 0.72 at 600ms SOA). However, no significant main effect of SOA or interaction, *Fs* < 1 was found.

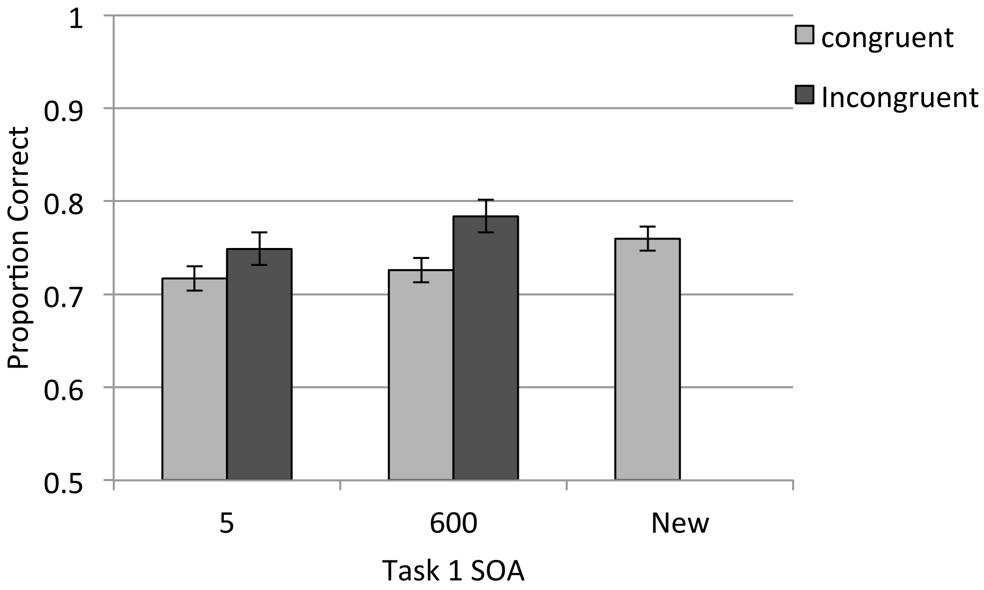


Figure 2. Experiment 1. Mean proportion correct for task 1 trials on recognition memory test separated by congruency and SOA at initial performance.

As congruency effects were found but no interaction was reported, we further analyzed the data in order to see if there were differences in congruency between the two types of words (animal/thing). These are displayed in Figure 3. A 2 x 2 repeated measures ANOVA treated congruency (congruent/ incongruent) and animacy (animal/thing words) as factors. A significant main effect was observed for congruency, $F(1,23)= 5.64$, $p= 0.03$, $\eta_p{}^2 = .20$, showing the general desirable difficulty effect with higher accuracy for incongruent trials. A significant main effect was also seen for animacy, $F(1,23)= 6.60$, $p= 0.02$, $\eta_p{}^2 = .22$, demonstrating that animal words were remembered significantly better overall compared to inanimate "thing" words. Additionally, the interaction approached significance $F(1,23)= 2.9$, $p= 0.09$, $\eta_p{}^2 = .12$. To examine this interaction more closely, two-tailed paired sample t-tests were examined separately for animal and thing words. For "thing" words, incongruent words (0.75) were remembered better than congruent words (0.68), $t(23)= 2.51$, $p = 0.02$. There was no difference found for animal words between the incongruent (0.80) and congruent trials (0.77), $t(23)= 0.69$, $p = 0.50$.

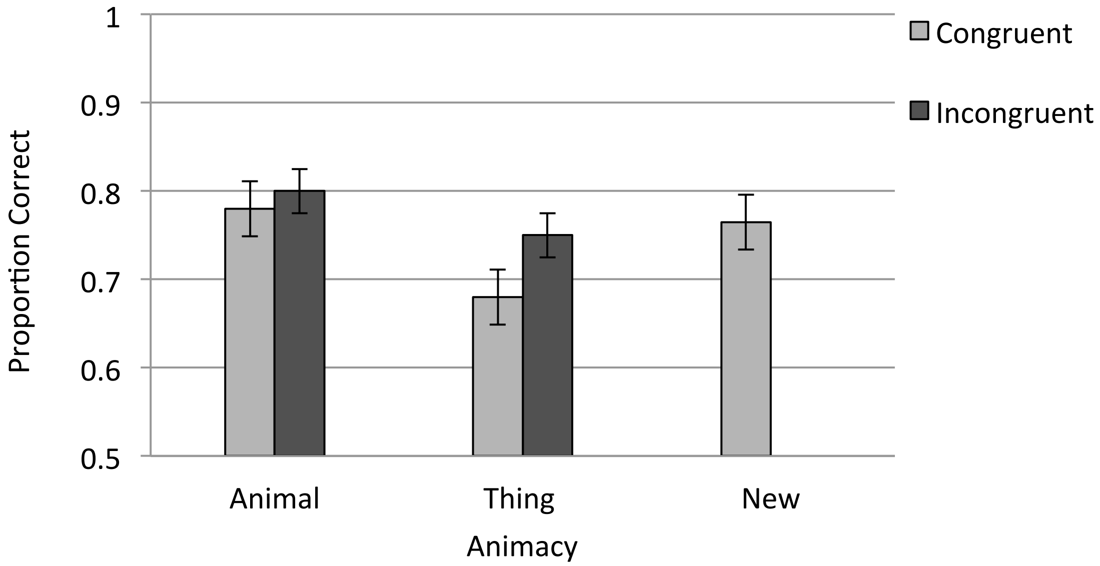


Figure 3. Experiment 1. Mean proportion correct for task 1 trials on recognition memory test separated by congruency and animacy.

These results were investigated further by calculating 2 x 2 repeated measures ANOVAs separately for animal words and then "thing" words. In both of these analyses, congruency (congruent/ incongruent) and SOA (5ms/ 600ms) were treated as factors. The analysis for animal words revealed no significant main effect of congruency $F(1,23) = 1.64$, $p = .21$, with no main effect of SOA and no interaction, *F*s < 1. However, "thing" words revealed a main effect for congruency $F(1,23) = 5.94$, $p = .02$, $\eta_p{}^2 = .21$, with no significant main effect for SOA $F(1,23) = 2.4$, $p = .13$, and no interaction, $F < 1$, with incongruent words at both the 5 ms (0.72) and 600ms (0.78) SOA remembered better compared to congruent words at both the 5ms (0.67) and 600ms (0.70) SOA.

Discussion

This experiment was designed to further investigate the stage-specific processes of desirable difficulty. This was completed through the use of a priming task where varying the congruency between tasks created a source of difficulty directed at the categorization stage of processing. Long-term memory and learning were measured by the proportion of T1 words that were remembered from phase 1 (designed to elicit priming effects) of the experiment.

During the priming phase of the experiment, it was predicted that a larger congruency effect at short SOAs and a smaller congruency effect at long SOAs would be present if priming effects were produced. It was predicted that the T1 prime stimuli would have more of an effect on T1 RT at the short SOAs as the processing of the T1 judgment would still be underway. At the long SOA it was predicted that the processing of T1 stimuli may be close to finishing as the T1 prime was presented and thus less of a priming effect would be present. Consequently, when there was an interference with the T1 word and the

T1 prime at the short SOA, the reaction times were longer for incongruent trials compared to congruent trials. This indicated that our selective attention manipulation was effective at interfering with and facilitating the T1 judgment task. Congruent items required less cognitive control over selective attention, demonstrated through faster reaction times (MacLeod, 1991; Stroop, 1935).

During the memory phase of the experiment, analyses revealed that incongruent words were remembered more on average compared to congruent words across SOA conditions. This indicates that we are seeing a desirable difficulty effect overall. This may indicate that directing attention to the core of the stimulus meaning (by making an animacy judgment) leads to better encoding of that word and producing a boost in long-term memory retention. What is interesting with this effect is that the same set of stimulus words used in Thompson et al. (in prep.) were used here. However, the particular judgment task used here ("is the word an animal or thing?" versus the previous "is the word bigger or smaller than the computer monitor?") elicited this desirable difficulty effect.

As there were differences of congruency, these results were further explored to see if any differences between the types of words were present. When investigating this relationship more closely we found that "thing" words demonstrated a significant desirable difficulty effect, whereas no significant desirable difficulty effect for animal words was found. More interestingly, animal words were better remembered on average in both congruency conditions compared to "thing" words. Specifically, about 80% of alive words were remembered for both congruency conditions. This may indicate that animacy information for animal words is activated more automatically or more strongly, and thus when introduced with a difficulty (incongruent trials), less attentional focus (cognitive control) needs to be directed to the word in order to overcome the conflict. Thus, no difference in encoding is made between congruency conditions and the desirable difficulty effect is more difficult to see. Alternatively, some feature of animal words may make them generally more memorable, obscuring our ability to observe a potential desirable difficulty effect due to semantic category conflict.

Experiment 2: Name Gender Categorization

Experiment 2 combined both semantic category priming and response selection priming in a single experiment, presenting participants with one or the other kinds of priming stimuli randomly through a series of name gender classification trials. The goal of this experiment was to attempt to observe both a desirable difficulty memory effect from incongruent semantic priming, and a divided attention memory cost effect from incongruent response priming, within a single experiment using a completely within-participants design.

Methods

Participants. Twenty-five undergraduate students from McMaster University participated in a one-hour experimental session in exchange for partial course credit. A language proficiency questionnaire was administered before commencement of the experiment to ensure English fluency. A total of two participants were excluded for low accuracy and excessively slow performance, leaving 23 participants for the reported data analysis.

Apparatus and Stimuli. The apparatus was the same as in Experiment 1. Stimuli consisted of 120 male and female names (half male, half female) in both the priming phase and memory phase. During the memory phase, 60 names were "old" (from the priming phase) and 60 names were "new" names. These names were common within the western/English language and only obvious male and female names (no ambiguous names such as Alex) were included. During both the priming and memory phase of the experiment, participants made their responses by hitting the "Z" and "/" keys.

Procedure

Priming phase. The priming phase was similar to Experiment 2 except participants were classifying names as male or female. A second priming stimulus (T1 prime) was presented with an SOA of 0; participants were told to ignore this stimulus. These T1 prime words were either a semantic prime (the words "male" or "female") or a response selection prime (the words "left" or "right"). Both these priming sets were randomized and were presented equally throughout the experiment. Participants were randomly assigned to one of two counterbalanced conditions where responses were made by pressing "Z" for female and "/" for male, or "/" for male and "Z" for female. A practice block, which consisted of 20 trials was completed in order to ensure the participants became comfortable with the task. The experimental trials consisted of three blocks with 40 trials each for a total of 120 trials. This portion of the experiment took participants about 15-20 minutes to complete.

Memory Phase. The memory phase was conducted the same as for Experiment 1, with four blocks consisting of 30 trials in each for a total of 120 trials. Participants were offered breaks between each block. This portion of the experiment took participants about 5-10 minutes to complete.

Results

Results for Experiment 2 mean RT and mean accuracy for the phase 1 name gender classification task, and mean memory accuracy for the phase 2 memory test are presented below.

Priming Phase.

Reaction Time. Mean RTs for T1 names were computed for each of the four conditions. These are presented in Figure 4. A 2 x 2 repeated measures ANOVA that treated congruency (congruent, incongruent) and prime type (male/female, left/right) as factors revealed no significant effects for congruency, prime type, or the interaction, all *F*s < 1. This suggests that no significant RT differences were seen between male/female primes during congruent (629 ms) and incongruent (634 ms) trials as well as left/right primes during congruent (635 ms) and incongruent trials (645 ms).

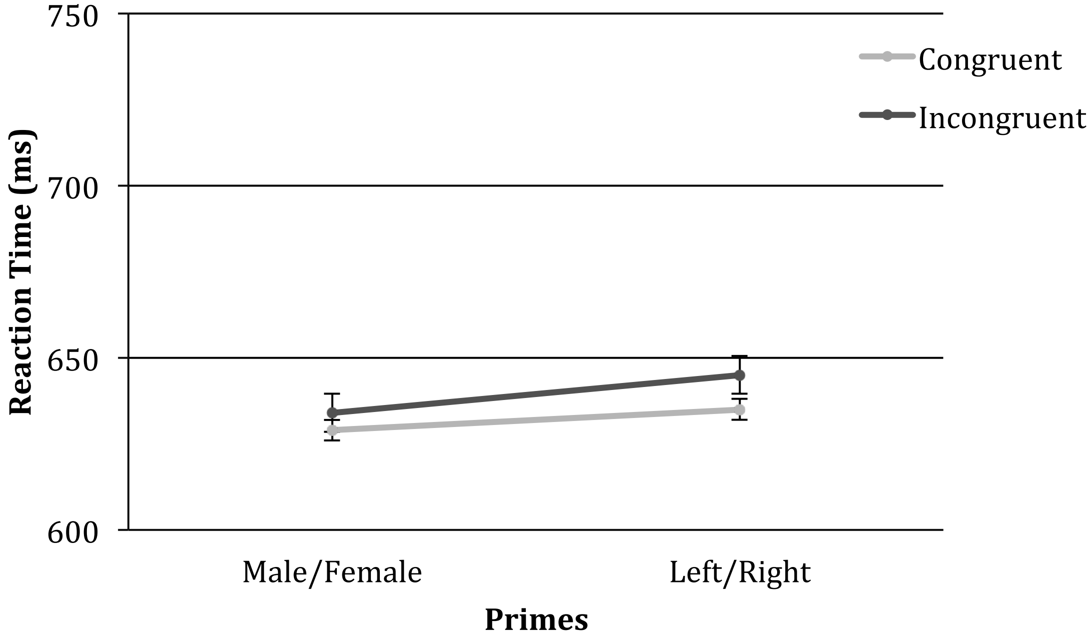


Figure 4. Experiment 2: Mean reaction times for congruent and incongruent T1s in the priming phase.

Accuracy. The mean accuracy within the priming phase was calculated for each of the four conditions. A 2 x 2 repeated measures ANOVA that treated congruency (congruent, incongruent) and prime (male/female, left/right) as factors revealed no significant effects for congruency, prime type, or the interaction, all Fs < 1. This suggests that no significant accuracy differences were seen between male/female primes during congruent (0.96) and incongruent (0.95) trials as well as left/right primes during congruent (0.95) and incongruent trials (0.95).

Memory Phase.

Accuracy. The proportion of items remembered from the first phase of the experiment served as the dependent variable in a 2 x 2 repeated measures ANOVA that treated congruency (congruent, incongruent) and prime (male/female, left/right) as factors. The mean proportion of the names remembered from phase 1 of the experiment separated by prime type and congruency is presented in Figure 5. The analysis revealed no significant main effect of congruency, $F < 1$, or prime, $F < 1$. While no significant interaction was seen, $F(1,22) = 2.78$, $p = 0.11$, the interaction was marginal, with the pattern of memory performance in the hypothesized direction (better memory for incongruent items with semantic priming, and worse memory for incongruent items with response priming). Values are presented in Figure 5 where the memory accuracy for each condition was: 0.68 (male/female, congruent); 0.70 (male/female, incongruent); 0.74 (left/right, congruent); 0.70 (left/right, incongruent); and 0.80 (new).

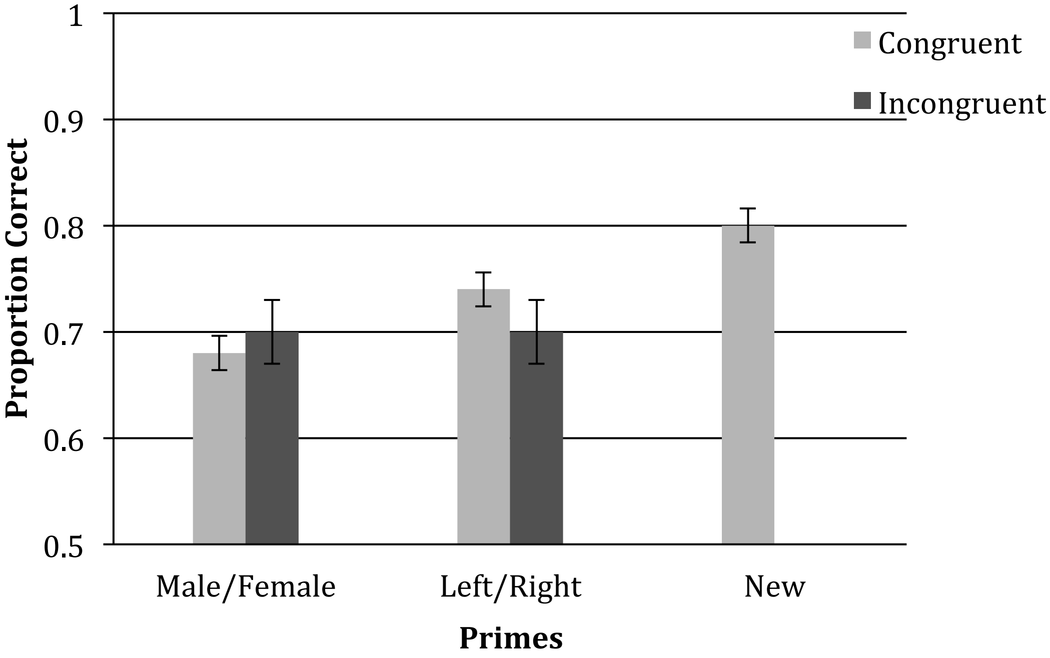
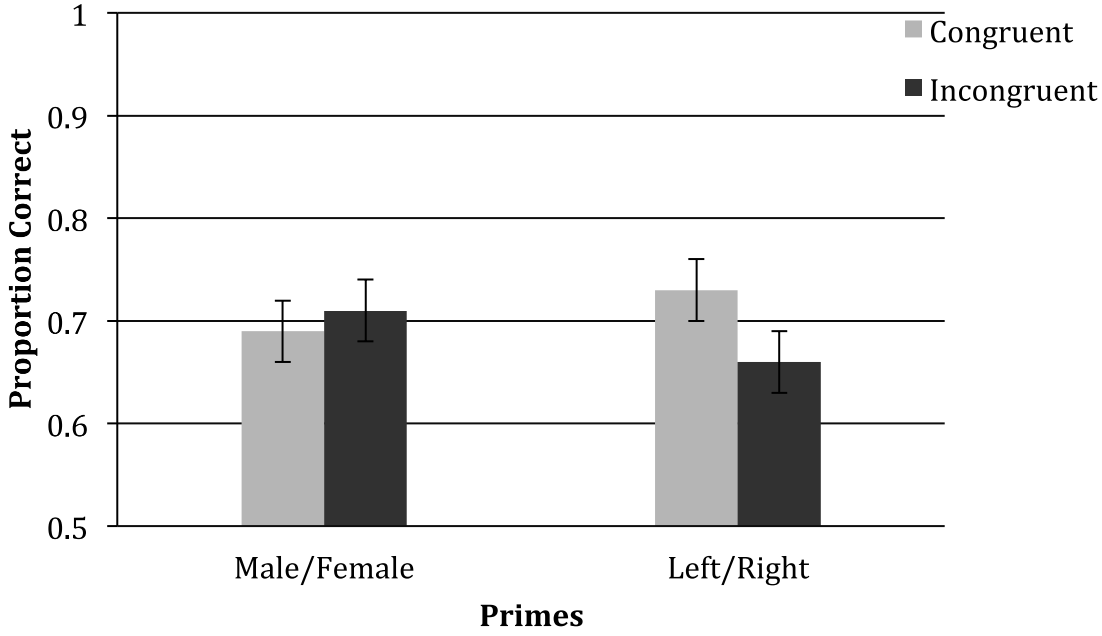


Figure 5. Experiment 2. Mean proportion correct for task 1 stimuli on later recognition memory test, separated by congruency and prime type.

Fast and Slow Accuracy. The accuracy scores for the memory phase were split into fast T1 RT scores and slow T1 RT scores using the mean RTs from the priming phase, following analysis procedures used by Thomson et al. (in prep.). These were computed based on the participant's average, where fast T1 RT scores are the values above the participant's average and slow T1 RT scores are the values below the participant's average. These results are presented in Figure 6a (fast T1 scores) and 6b (slow T1 scores).

Two 2 x 2 repeated measures ANOVAs were performed separately for fast T1 RTs and then for slow T1 RTs. In both these analyses, congruency (congruent, incongruent) and prime type (male/female, left/right) were treated as factors. The analysis for fast T1 RT stimuli revealed no significant main effect of congruency $F < 1$, or prime type $F(1,22) = 2.1$, *p*= 0.17. Similarly, no interaction was found, *F* < 1, indicating no reliable congruency effects in fast T1 RTs as previously shown by Thomson et al. (in prep.).

Analysis for slow T1 RT stimuli revealed no significant differences of congruency, *F* < 1 or prime *F* < 1. Even though the interaction was in the expected direction, it did not reach significance *F*(1,22) = 1.78, *p* = 0.19.



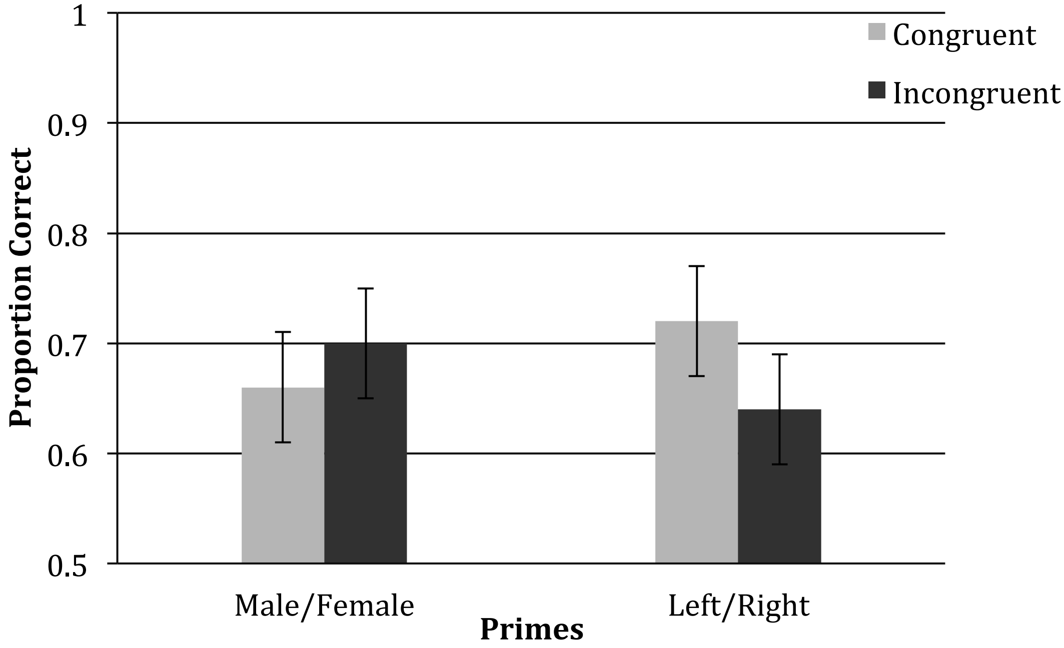


Figure 6. Experiment 2. Mean proportion correct for task 1 trials on recognition memory test separated by prime for both fast (top; 6a) and slow (bottom; 6b) T1 RTs.

Discussion

Similar to Experiment 1, Experiment 2 was also designed to investigate the stage-specific processes of desirable difficulty. In this experiment, T1 was primed with either the semantic representation of the word (male/female) or the left/right response cue. The potential influence of this priming on long-term memory and learning was measured by the proportion of T1 words that were remembered from phase 1 of the experiment.

In phase 1, while the congruency effect in reaction times was in the predicted direction, no significant priming effect was observed. This lack of an initial reliable priming effect on Task 1 RTs could help explain why later congruency effects on memory performance were also not reliably observed.

Although congruency effects in memory performance also did not reach significance, we did observe a marginal interaction effect in the direction of our predictions, with a contrasting pattern of congruency effects on semantic and response selection primes. For semantic male/female primes, a numerically higher proportion of incongruently primed names were remembered compared to the names in the congruent trials. The opposite pattern was numerically observed for the left/right primes, with a higher proportion of words for the congruent trials remembered compared to the incongruent trials.

When results were separated by fast and slow T1 RTs, again results did not reach significance but also followed the predicted direction. We were particularly interested in examining the slow T1 RTs results, as we believed these trials would demonstrate a stronger priming effect compared to when participants were focused and ready in the fast

T1 RTs. During these slow T1 RTs incongruent words with the male/female prime were better remembered compared to the incongruent trials with the left/right prime. The opposite pattern was seen for the left/right response primes.

If these non-significant patterns were shown to be reliable, they would suggest that incongruent information for the male/female primes may have caused an increase in selective attention to aid in better encoding of the T1 stimuli. Although these results were not significant, the pattern of results are similar to the attentional boost effect (Jiang & Swallow, 2010), as well as results seen in the work done by Thomson et al. (in prep.).

GENERAL DISCUSSION

The present paper extends the ideas of Thomson et al. (in prep.), which suggests a stage-specific mechanism involved in desirable difficulty. In their paper, they found that directing a difficulty towards the response selection stage of processing elicited typical divided attention effects. Additionally, targeting the categorization stage through classifying names by gender produced this desirable difficulty effect. They provided evidence that this gender classification is more semantically central than categorizing items by size, which tends to be relative and depend on context. Making a judgment on an object's size may divert attention away from the core of the stimulus to another processing stage, while classifying a name as male or female may increase attention to the core semantic representation or meaning of the stimulus which may enhance encoding. These results point to a stage-dependent model where desirable difficulties cause process-specific enhancements of attentional work rather than task-general increases. More specifically, making a *specific* aspect of a task harder is necessary to direct cognitive resources such as attention towards the to-be-learned information to engage a deeper processing thereby eliciting the desirable difficulty learning benefits.

Experiment 1 in the present paper was completed as a follow-up to these experiments. Here we tested the idea that central semantic priming should give a desirable difficulty effect (as shown by Thomson et al., in prep., with male/female names). We used their same concrete noun stimuli as their Experiment 2 (big/small judgment), however asked participants to make more central animacy judgments. If animacy is a central feature to these words, we predicted we would see a desirable difficulty effect as they did with their male/female name judgment. This is exactly what we found; memory for incongruent stimuli (conflict inducing trial) was better than for congruent trials.

An interesting finding is that the desirable difficulty benefit was only seen for "thing" stimuli. However, overall memory accuracy for animal words was better than for "thing" words for both congruent and incongruent trials. We wondered whether this was a frequency effect as studies investigating association value (Allen & Garton, 1968) have suggested that low frequency words are better remembered (as animal words in this experiment were). Additionally, we wondered whether animal words were more imagable than "thing" words. Past research has found that easily imagable words are recalled better than less imagable words (Marcel, 1983). However, after checking our stimulus sets against Kucera and Francis (1967) norms, we found no reliable differences.

Even though we are seeing this ceiling memory performance seen in animal words, "thing" words still demonstrated this desirable difficulty effect. We see an immediate decline in performance during the priming task that elicits a later memory boost. This replicates findings by Rosner et al. (2014), which showed that increasing attentional work produces immediate performance costs, but leads to better encoding of that word. With this in mind, having generally better memory for items (animal words) may decrease this selective attention boost we see when we present a difficulty at encoding.

In Experiment 2 we manipulated the processing stage (categorization and response selection) to which processing conflict and hence selective attention and control was directed. In particular, we investigated a task where attention was focused at a response mapping stage of processing (left and right response primes) and where it was focused on the categorization stage (male and female primes). We predicted that we would be able to see typical divided attention costs and desirable difficulty within the same participant if a stage-specific mechanism was being activated. Although results did not reach significance, they were in the direction predicted. For example, typical divided attention effects were observed for left/right primes, as memory was better for congruent stimuli compared to incongruent stimuli. Additionally, results were approaching a desirable difficulty effect for the male/female primes, as there was better memory for incongruent items compared to congruent items. Although these differences did not reach significance, they follow the pattern of results by Thomson et al. (in prep.).

Recent research has demonstrated that better memory encoding for incongruent compared to congruent items may be a result of an increase in cognitive control (Botvinick et al., 2001) or a control mechanism aimed at reducing response ambiguity (Rosner et al., 2015). The present study offers evidence for a process-specific model of desirable difficulty instead of this top-down general increase of memory retention based on the conflict monitoring theory. Specifically, desirable difficulty is not evident for all difficult selective attention-encoding conditions but rather is a selective mechanism that varies in different processing stages.

As the results in Experiment 2 showed the predicted pattern of opposite effects of semantic and response congruency on later memory performance, yet did not reach significance, we should attempt to replicate this experiment with more participants and possibly a modified design. These weaker results may be due to the delivery of these different priming types. It is possible that having these priming types randomly interleaved may cause confusion for the participant, causing them to be more conservative and controlled with their responses. This could explain why a decrease in RT priming effects was seen in the first phase of the experiment, and would predict generally reduced congruency effects on later memory. Future research may consider blocking these priming types to induce more automatic processing so that priming effects are more observable.

Additionally, varying the nature of the primary attentional task will provide further insight into the attentional mechanisms involved in producing the desirable difficulty effect. For example, it would be interesting to investigate the process-specific components involved with memory costs and improvements via task switching. This research could investigate univalent (low control) and bivalent (high control) tasks with different degrees of preparation time, and test how degree and timecourse of switching demand is associated with later memory for task stimuli.

Overall, understanding the nature of mechanisms involved in desirable difficulty is of significance to many facets of learning. Educators have believed that initial performance is a good predictor of long-term retention. More specifically, reducing cognitive load during initial retention equates to better learning (Diemand-Yauman et al., 2010). Teaching styles also tend to be structured to support easy initial learning, leading to unfavorable long-term retention (Bjork & Bjork, 2011). Students have also often attributed fluent learning experiences to better comprehension of material, which then lead to inefficient learning and improper study strategies (Yue et al., 2013).

Research is continuing to provide evidence of how increasing difficulty during study phases improves long-term retention and learning (Richland et al., 2005; Bjork & Bjork, 2011; Thomson et al., in prep.). It is crucial for educators to implement methods that allow learners to retain flexible knowledge that can be applied in multiple contexts over long periods of disuse. Consequently, investigating the stages involved in the desirable difficulty effect will allow us to further understand the mechanisms involved in initial learning, which will lead to successful long-term retention for learners.

References

Allen, L. R., & Garton, R. (1968). The influence of word knowledge on the word frequency effect in recognition memory. Psychonomic Science, 10, 401 402.

Baddeley, A. D., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. Journal of Experimental Psychology: General, 13, 518-540.

Battig, W. F. (1972). Interference during learning as a source of facilitation in subsequent retention and transfer. Paper presented at the annual meeting of the American Educational Research Association.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J.Metcalfe and A. Shimamura (Eds.), Metacognition: Knowing about knowing. (pp.185-205). Cambridge, MA: MIT Press.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.

Bjork, E. L., & Bjork, R. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). New York: Worth Publishers.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108,* 624-652.

Craik, F.I.M.. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 2, 671-684.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition, 118,* 111– 115.

Henik, A., Bibi, U., Yanai, M., & Tzelgov, J. (1997). The Stroop effect is largest during first trials. *Abstracts of the Psychonomic Society, 2,* 57.

Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 507–513.

Jiang, Y. V., & Swallow, K. M. (2010). An overview of the attentional boost effect. *Journal of Vision, 10*, 7, 143.

Kucera, H., & Francis, W. (1967). *Computational analysis of present- day American English*. Providence, RI: Brown University Press.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585–592

Krebs, R. M., Boehler, C. N., De Belder, M., & Egner, T. (2013). Neural conflict-control mechansims improve memory for target stimuli. *Cerebral Cortex, 25*, 833-843.

Kuiper, N.A., & Rogers, T.B. (1979). Encoding of personal information: Self-other differences. *Journal of Personality and Social Psychology, 37*, 499-514.

Laming, D. R. J. (1968). Information theory of choice-reaction times*.* London:Academic Press.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 219-234.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. Psychological Bulletin, 109(2), 163–203.

Marcel, A.J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition*. Cognitive Psychology, 15*, 197-237.

Morton, J. B., Ezekiel, F., & Wilk, H. A. (2011). Cognitive control: easy to identify but hard to define. *Cognitive Science, 3*, 212-216.

Mulligan, N. W., Spataro, P., & Picklesimer, M. (2014). The attentional boost effect with verbal materials. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 40*, 1049-1063.

Naveh-Benjamin, M., Craik, F. I., Gavrilescu, D., Anderson, N. D. (2000). Asymmetry between encoding and retrieval processes: evidence from divided attention and a calibration analysis. *Memory & Cognition, 6,* 965-976.

Naveh-Benjamin, M., Guez, J., & Marom, M. (2003). The effects of divided attention at encoding on item and associative memory. *Memory & Cognition, 31*, 1021-1035.

Rogers, T.B., Kuiper, N.A., & Kirkner, W.S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*, 677 688.

Schneider, W. & Shiffrin, R.M. (1977).Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66.

Spataro, P., Mulligan, N. W., & Rossi-Arnaud, C. (2013). Divided attention can enhance memory encoding: The attentional boost effect in implicit memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 39*, 1223-1231.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Swallow, K. M., & Jiang, Y. V. (2010). The attentional boost effect: Transient increases in attention to one task enhance performance in a second task. *Cognition, 115,* 118-132.

Swallow, K.M., & Jiang, Y.V. (2011). The role of timing in the attentional boost effect. *Attention, Perception, & Psychophysics, 73,* 389-404.

Swallow, K.M., & Jiang, Y.V. (2012). Goal-relevant events need not be rare to boost memory for concurrent images. *Attention, Perception, & Psychophysics. 74(1), 70-82.*

Rosner, T .M., D'Angelo, M. C., MacLellan, E., & Milliken, B. (2014). Selective attention and recognition: Effects of congruency on episodic learning. *Psychological Research*. doi: 10.1007/s00426-014-0572-6

Rosner, T. M., & Milliken, B., (2015). Congruency effects on recognition memory: a context effect. *Journal of Experimental Psychology. 69(2),* 206-212.

Thomson, S. J., Ptok, M. J., Humphreys, K. R., Watter, S. (2015). Stage-Specific Attentional Mechanisms of Desirable Difficulty. (in preparation).

Uncapher M.R., Wagner A.D. (2009). Posterior parietal cortex and episodic encoding: insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiology, Learning and Memory. 91*:139–154.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin*, *133*, 273-293.