COMPARING REPETITION PRIMING IN WORDS AND ARITHMETIC EQUATIONS: ROBUST AND COMPARABLE PRIMING IS RESISTANT TO FEATURE CHANGES AND RELIES ON EMPIRICALLY SOUND DESIGN AND ANALYSIS

A thesis submitted in partial fulfilment of the requirements for the Degree

of Master of Science in Psychology

in the University of Canterbury

A. M. Humphries

University of Canterbury

2017

Table of Contents

Acknowledgments

The author wishes to express sincere appreciation to Dr Zhe Chen for her assistance in the preparation of this thesis and for her instrumental advice on experimental design, theoretical focus, and general writing style. In addition, special thanks to Jonathon Wiltshire for his patience and invaluable input during the programming stage. Thanks also to the members of the psychology department for their valued input over the past few years.

Notes

1. The five experiments reported here build on an experiment reported for my BSc. Honours' project. Hence, there is some overlap in the theoretical material covered.

2. Experiments 3 and 5 were written as a manuscript and submitted for publication to *Frontiers in Psychology* on 24 October 2017. Hence, there is overlap in the theoretical material as well as in the method and results sections.

Abstract

Previous studies have shown that stimulus repetition can lead to reliable behavioural improvements. Although this repetition priming (RP) effect has been reported in a number of paradigms using a variety of stimuli including words, objects, faces, and scenes only a few studies have investigated mathematical cognition involving arithmetic computation, and no prior research has directly compared RP effects in a linguistic task with an arithmetic task. In five experiments, between- and within-subjects designs were used to draw comparisons between a word and an arithmetic categorisation task. In a study-test block paradigm, stimuli were repeated identically to compare the magnitude of RP, and colour and response hand were manipulated to compare the effects of feature changes for repeated, otherwise identical, stimuli. In the same experiments, the merits of using the relative-difference or absolute-difference method of analysis was also investigated. The results show that the magnitude of RP was comparable between the two tasks, and that changing the colour or the response hand had a negligible effect on priming in either task. These results extended previous findings in mathematical cognition. They also indicate that priming does not vary with stimulus domain. The implications of the results were discussed with reference to both facilitation of component processes and episodic memory retrieval of stimulus-response binding. The results also indicate that, unless a fully randomised experimental design is used, the absolute-difference method of analysis may produce biased results that provide only a weak basis for drawing empirical conclusions.

Comparing repetition priming in words and arithmetic equations: Robust and comparable priming is resistant to feature changes and relies on empirically sound design and data analysis

Overview

Repetition priming (RP) refers to improvements in the speed and /or accuracy of processing a stimulus when that stimulus, or a similar one, is repeated. In a typical behavioural experiment investigating RP using visual stimuli, participants are presented with a series of displays containing stimuli such as numbers (e.g., Sciama, Semenza, & Butterworth, 1999), objects (e.g., Dobbins, Schnyer, Verfaellie, & Schacter, 2004), words (e.g., Graf & Ryan, 1990), faces (e.g., Valt, Klein, & Boehm, 2015), or scenes (e.g., Stevens, Kahn, Wig, & Schacter, 2011). With each display, participants are required to respond as quickly and as accurately as possible according to the task, and this often involves making a perceptual (i.e., red or green) or semantic (i.e., odd or even) decision that requires either a verbal or a manual (i.e., a key press on a computer keyboard) response. Reaction time (RT) and accuracy improvements from repetition are used as indices of RP. Although it has been shown that the magnitude of RP can vary depending on the number of repeats (e.g., Dobbins et al., 2004; Horner & Henson, 2009; Maljkovic & Nakayama, 1994; 1996; Treisman, 1992; Valt et al., 2015), the temporal lag between repeats (e.g., Henson, Rylands, Ross, Vuilleumeir, & Rugg, 2004), and the number of intervening items between repeats (e.g., Bentin & Moscovitch, 1988; Henson et al., 2004; Thomson & Milliken, 2012; 2013), the presence of RP indicates that behaviour is not just driven by the current stimulus situation but is also affected by what has been processed in previous stimulus situations (Schacter, 1990).

Most previous studies on RP used stimuli such as words, objects, faces, and/or scenes (e.g., Dobbins et al., 2004; Graf & Ryan, 1990; Horner & Henson, 2009; 2011; 2012;

Roediger & Blaxton, 1987; Stevens et al., 2011; Valt et al., 2015). Not many studies used numbers as stimuli (e.g., Dehaene, Naccache, Le Clec'H, et al., 1998; Kunde, Kiesel, & Hoffman, 2003; Naccache & Dehaene, 2001; Reynvoet, Caessens, & Brysbaert, 2002), and even fewer required participants to perform arithmetic computation on the numbers (e.g., Salimpoor, Chang, & Menon, 2010; Scheepers & Sturt, 2014, Scheepers, Sturt, Martin, Myachykov, Teevan, & Viskupova, 2011; Sciama et al., 1999). While there are indications that similarities in RP exist across stimulus domains and task types that are suggestive of a common mechanism(s) (Ward, Chun, & Kuhl, 2013), a direct comparison between word and arithmetic RP has not been undertaken. It therefore remains to be seen whether RP effects are stable across these two stimulus domains.

The experiments in this thesis used both a between- and within-subjects design to compare the magnitude of RP in a word and an arithmetic categorisation task. The thesis is divided into 4 sections. Section 1 consists of an overview of two commonly used RP paradigms and two methods of data analyses. Section 2 provides a selective review of the main theoretical perspectives on RP. Section 3 reports 5 experiments, and finally, Section 4 consists of the general discussion, including limitations and future directions.

1. RP paradigms and the methods of analyses

1.1 Two commonly used RP paradigms

RP is typically investigated in one of two paradigms: the prime-probe couplets paradigm, and the study-test block paradigm. In the prime-probe couplets paradigm, a single trial consists of a prime display followed by a probe display (e.g., Dehaene et al., 1998; Grainger & Jacobs, 1993; Kunde et al., 2003; Naccache & Dehaene, 2001; Neill, 1997; Reynvoet et al., 2002; Wong, 2000). The prime is either congruent or incongruent with the probe. In some cases both the prime and probe displays require a behavioural response (e.g., Neill, 1997; Wong, 2000), and in other cases only the probe display does (e.g., Dehaene et al., 1998, Naccache & Dehaene, 2001). In designs that require responses to both the prime and probe displays, the prime display often remains visible until a response is made, meaning it is accessible to conscious processing. A number of studies using this design have shown that priming is impaired when some features of the probe do not overlap with the prime, leading to the conclusion that RP is, to some degree, influenced by episodic retrieval of the prime trial (Neill, 1997; Wong, 2000). In designs where the prime does not require a response, the prime is often presented very briefly and is usually followed by a mask, which further eliminates any residual iconic memory before the probe is displayed. A number of studies using this design have shown that presenting a prime subliminally, i.e., below the threshold of conscious awareness, can result in robust behavioural improvement in response to a subsequent probe, suggesting that RP is automatic (e.g., Abrams, & Greenwald, 2000; Abrams, Klinger, & Greenwald, 2002; Dehaene et al., 1998; Dehaene, Jobert, Naccache, et al., 2004; Henson, Mouchlianitis, Matthews, & Kouider, 2008; Naccache & Dehaene, 2001). Due to this automaticity, RP is frequently discussed in terms of implicit memory processes (Roediger & McDermott, 1993).

Whereas the prime-probe couplets paradigm is particularly suitable for exploring the effect of immediate repetition, the study-test block paradigm (e.g., Dobbins et al, 2004; Graf & Ryan, 1990; Horner & Henson, 2009; 2011; Masson, 1986; Roediger & Blaxton, 1987; Salimpoor et al., 2010; Stevens et al., 2011) enables additional investigations into the effect of varying the time and the number of intervening trials between repeats. In this paradigm, both the study and test blocks consist of a series of trials, and a response is required on every trial. Although early claims were made that RP was invariant to a temporal delay of up to one week in a word-stem completion task (e.g., Tulving, Schacter, & Stark, 1982), subsequent research showed an attenuation of RP in the same task and over the same temporal delay (e.g., Roediger & Blaxton, 1987). Using a lexical decision task and a face classification task, Bentin and Moscovitch (1988) also showed that the attenuation of RP, with lags of zero, four, and 15 intervening trials, depended on stimulus type, the number of repetitions, and the type of task being performed. Furthermore, Henson and colleagues (2004) manipulated both time and the number of intervening trials in an object classification task and showed that the attenuation of RP resulted from both temporal delay and the interference of intervening items.

In addition to measuring the effects of repeating a stimulus in all its feature dimensions, features of otherwise identical stimuli can be individually manipulated along various perceptual (i.e., colour, shape, font, notation) and response (i.e., classification, decision, motor-action) based dimensions in order to measure the impact such changes have on RP. The experimental manipulations can be blocked or mixed. In the blocked version, the stimuli in the test block are identical to those in the study block in the control condition, and they differ in a single feature dimension (i.e., notation or colour) in the experimental condition. For example, in an experiment that involves the manipulation of colour and notation of words, a study set of words is presented with each word being allocated a particular colour and notation. In one test block (the colour block), the notation of each word is retained, and half the words retain the same colour but half change colour. In the other test block (the notation block), the colour of each word is retained, and half the words retain the same notation but half change notation. In the mixed version, the manipulated dimensions (i.e., notation and colour) are randomly mixed within the same test block. Using the same example, the identity of the study words is repeated in the test block, and some words remain identical (the control condition), some differ in colour (the colour condition), some differ in notation (the notation condition), and some differ in both colour and notation (the colour and notation condition).

Results of these feature manipulations provide evidence that the magnitude of RP can be sensitive to perceptual changes (e.g., Graf & Ryan, 1990; Masson, 1986; Neill, 1997; Roediger & Blaxton, 1987; Sciama et al., 1999; Stevens et al., 2011) and/or response changes (e.g., Horner & Henson, 2009; Kunde et al., 2003; Race, Shanker, & Wagner, 2009; Reynvoet et al., 2002; Valt et al., 2015), particularly where tasks involve atypical presentation formats (e.g., Masson, 1986; Sciama et al., 1999; Graf & Ryan, 1990), a high degree of task difficulty (e.g., Lamy, Zivony, & Yashar, 2011; Soldan, Clarke, Colleran, & Kuras, 2012), application of attention (e.g., Henson, Eckstein, Waszak, Frings, & Horner, 2014; Logan, 1988; 1990; Treisman, 1992), and/or multiple repetitions (e.g., Horner & Henson, 2009; Valt et al., 2015). However, a number of studies have also revealed that RP can be remarkably robust to feature manipulations (e.g., Biederman & Cooper, 1991; 1992; Dehaene et al., 1998; Fiser & Biederman, 2001; Naccache & Dehaene, 2001), indicating that RP can occur at multiple levels of abstraction depending on task demand (see Henson et al., 2014, for a review).

1.2 Methods of analyses in RP studies

An important consideration in investigating RP is the method of data analysis. To define priming, we use the notation

$$S_X = \frac{1}{n}\sum_{i=1}^{n} S_X^i \,, for\ X \in \{N, R\},$$

$$T_X = \frac{1}{n}\sum_{i=1}^{n} T_X^i \,, for\ X \ \in \ \{N, R\},$$

where S_X^i and T_X^i refer to the median prime/study and probe/test RT (or accuracy) for participant *i*, respectively, *N* and *R* refer to novel and repeated stimuli, respectively, and *n* is the number of participants.

Most previous studies on RP compute RP by measuring the differences in performance between the experimental and control conditions while only using the data from the probe trials or the test blocks (e.g., Biederman & Cooper; 1991; 1992; Fiser & Biederman, 2001; Horner & Henson, 2009; Masson, 1986; Race et al., 2009; Sciama et al., 1999; Stevens et al., 2011; Valt et al., 2015). In notation, $P = T_N – T_R$, where P is the priming effect reported. This method of analysis will be termed the *absolute-difference* method of analysis as RP is based on comparing the absolute difference in RT and accuracy between the probe trials/test blocks for the experimental (i.e., the primed or studied) condition and the control (i.e., the non-primed or non-studied) condition. In this method of analysis the data from the prime trials or study blocks are not included in the analysis. This method fits with the conception of RP as a phenomenon where studied (primed) items are more efficiently and effectively responded to than non-studied (non-primed) items (Schacter, 1990).

However, given that theories of RP attempt to explain current processing as a result of prior processing, it would seem important that all characteristics of the initial state be taken into consideration and, where possible, analysis be based on the change in outcome between prime/study and probe/test trials. In notation, $\Delta_N = S_N – T_N$, where Δ_N is the average change in RT (or accuracy) in the novel condition; and $\Delta_R = S_R – T_R$, where Δ_R is the average change in outcome in the repeat condition. The differential priming effect resulting from an experimental manipulation is then $P_\Delta = \Delta_N – \Delta_R$. This will be termed the *relative-difference* method of analysis as RP is based on the relative difference in RT and accuracy between study and test trials. For example, in a study-test block design, instead of directly comparing the performance in the studied test block with that in the unstudied test block (i.e., the absolute-difference method), one first calculates two difference scores: one between the study and test blocks for the studied condition, and the other between the study and test blocks for the unstudied condition. Thus, whereas the absolute-difference method is built on the assumption that performance does not differ across conditions in the study blocks, the relative-difference method takes into account (empirically) performance in these blocks.

To investigate whether these methods would lead to different patterns of results, in the experiments reported here, I used both methods to analyse the experimental data. Whereas the results based on the relative-difference method are reported in the main text, the results based on the absolute-difference method are shown in Appendix B.

2. Theoretical interpretations of RP

There are two main theoretical perspectives regarding the mechanism(s) that gives rise to RP. The first perspective explains behavioural advantages in terms of facilitation in component processes. This perspective focuses on how changes in neural activation patterns enable more efficient processing when the same or a similar stimulus is subsequently presented. The second perspective explains behavioural advantages in terms of memory encoding and retrieval processes. This perspective emphasises the perception of a stimulus as a retrieval cue for the previous encounter with that same, or a similar, stimulus. It is worth noting that although the two theoretical perspectives differ in their emphases on the mechanism(s) of RP, they are by no means mutually exclusive (Kahneman, Treisman, & Gibbs, 1992). In fact, recent evidence suggests that they are inter-dependent and co-operative (e.g., Lamy et al., 2011; Thomson & Milliken, 2012; 2013; Soldan et al., 2012; Valt et al., 2015).

In this section, I will first present a brief outline of the major theories falling within each perspective, with Section 2.1 covering those under the facilitation of component processes, and Section 2.2 those under the episodic retrieval perspective. Evidence supporting the integration of the two perspectives is discussed in Section 2.3.

2.1 Facilitation in component processes

This account focuses on how repeated processing of a stimulus produces changes in activated neurons and how these adaptations then facilitate subsequent processing when the same, or a similar stimulus, is presented. A number of physiological models have been proposed to explain the mechanisms that underlie these neural adaptations (see Grill-Spector, Henson, & Martin, 2006, for a review), with some models focused on factors related to changes in individual neurons and others to changes in the response properties of networks of neurons.

In terms of changes in individual neurons, there are three key RP models: *fatigue*, *sharpening*, and *facilitation*. The *fatigue* model emphasises the role of reductions in the amplitude of neurons' response curves (e.g., Kohn & Movshon, 2003); the *sharpening* model explains efficiency gains as the result of a reduction in the number of neurons that are activated when stimuli are repeated (e.g., Wiggs & Martin, 1998), and the *facilitation* model proposes that processing efficiencies are gained from faster onset of neural activation (e.g., James & Gauthier, 2006). With regard to changes in the networks of neurons, one of the key RP models suggests that processing efficiencies are gained as a result *of short-term feature weighting*, in which sustained excitability in activated neuronal populations leads to a behavioural advantage for features just experienced (e.g., Maljkovic & Martini, 2005; Maljkovic & Nakayama, 1994). Another model conceptualises RP as the result of increased synchrony, proposing that sensitivity to both firing rates and the timing of inputs from downstream neurons leads to synchronised activation and facilitates processing through enhanced inter-region communication (e.g., Ghuman, Bar, Dobbins, & Schnyer, 2008; Gilbert, Gotts, Carver, & Martin, 2010; Gotts, Chow, & Martin, 2012). Although these models all focus on the effects of stimulus-driven bottom-up processing, there is also evidence that top-down processing has an effect on the magnitude of RP and on the response of neurons in early sensory processing regions (e.g., Summerfield, Egner, Greene, Koechlin, Mangels, & Hirsch, 2006; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008; Summerfield, Wyart, Johnen, & De Gardelle, 2011).

Facilitation models that incorporate the influence of top-down feedback in modulating the responses of neurons in early processing regions are based on predictive coding theory (e.g., Friston, 2005; James & Gauthier, 2006) and have been implemented using models based on Bayes Theorem (e.g., Rao & Ballard, 1999). Under such models, improvements in behavioural responses are said to reflect the interplay between incoming stimuli and internally generated predictions that are driven by recent experience (Friston, 2005; Mesulam, 1998; 2008). The effect this interplay has on the physiological response of individual neurons and neural networks is said to be due to the influence of fulfilled expectations driving neural adaptation (Summerfield et al., 2006; 2008; 2011). Furthermore, this top-down effect seems to be moderated by attention. When task manipulations divert attention away from the stimuli being presented, expectancy related adaptation is reduced (e.g., Larsson & Smith, 2012). Hence, it is important to consider these facilitation models as neither mutually exclusive nor collectively exhaustive. Rather, it is more likely that they all play a role in providing neural processing efficiencies that are linked with the repetition of information, and with behavioural improvements in the form of RP.

An important aspect of the facilitation perspective is that RP can occur due to efficiency gains at any level of the neural processing hierarchy. As such, RP should not be adversely affected, at least to a significant degree, by changes in irrelevant features of a stimulus-response trial. Indeed, RP has been shown to be relatively robust to changes in visual perspective, size, and position in an object naming task (e.g., Biederman & Cooper, 1991; 1992; Fiser & Biederman, 2001). However, attenuation of priming was observed when different exemplars of objects were presented in the test phase of the task, and this was said to reflect the fact that RP for exact repetitions (even when presented in a novel orientation, size, or position) involves advantages at both a perceptual and conceptual level, whereas RP for different exemplars only involves advantages gained at a conceptual level (Biederman & Cooper, 1991; 1992).

The effect on RP of manipulating perceptual features in word identification tasks also shows that RP can be robust to changes in some features but not to changes in other features. In one such task, Graf and Ryan (1990, Experiment 1) manipulated words so they were either presented normally or upside-down (a 180 degree rotation of whole words). The magnitude of RP in test blocks was unaffected by whether the words were presented in the same format or in the opposite format as they had first been presented in the study block. As the identification of words can be regarded as an automatic process (Stroop, 1935), this robust RP regardless of the format change was said to reflect facilitation at the level of word identification with the format of the word not interfering with this aspect of processing. However, when the manipulation involved an unusual format such as reversing the order of letters so the word was spelled backwards (i.e., fridge becomes egdirf), attenuated priming was observed (Graf & Ryan, 1990, Experiment 2). The reason why the manipulation had an effect when words were presented backwards rather than upside-down was explained with regards to differences in the processes engaged to decode the words. In the case of upside-down words, the manipulation involves a single global transformation that maintains the same overall shape of the word. However, in the case of the words presented backwards, there is a transformation in the order of individual letters. This transformation perhaps nullifies any advantage of having previously encountered the word in the study phase as the overall shape of the word is now irrelevant and it must, instead, be decoded in a right-to-left, letter-by-letter, manner (Graf & Ryan, 1990). This explanation bears similarities to the idea of transfer-appropriate processing (e.g., Morris, Bransford, & Franks, 1977; Roediger &

Blaxton, 1987), with facilitation occurring as a result of overlaps in the mental processes required for each presentation and attenuated RP occurring as a result of the engagement of non-overlapping mental processes.

An attenuation of RP with unusual presentation format has also been reported in an arithmetic task. Sciama and colleagues (1999) investigated the effects of changing the notation of repeated numbers (word form, dot array, or Arabic digit) in a study-test block design. The task required participants to perform addition on a visual display containing two numbers, and report the sum verbally. The results show that the repetition of notation from study to test led to greater priming than a change in notation but only when the test presentation was in either word form or dot array. A test display presented in Arabic digit notation was primed by all other notations. It was argued that this difference was due to word form and dot array notation being ‘atypical’ for arithmetic. These atypical notations led to an increased need for attention and stronger encoding of surface features that then interfered with the operation of a priming mechanism. Based on their findings, Sciama et al. (1999) concluded that arithmetic processing is supported by both a common- and form-specific code.

It is also possible that when the notation was repeated efficiency gains were due to facilitation at both a perceptual and conceptual level, but when notation changed the processing advantage was limited to conceptual level (e.g., Biederman & Cooper, 1991; 1992). However, this line of reasoning cannot account for the fact that RP was not affected by a change in notation when the test display was in Arabic digit notation, whereas RP was affected by a change in notation when the test display was in word form or dot array notation. To account for this, it is possible that for ‘typical’ notation the information is effortlessly processed from the perceptual symbolic form through to computational stages, whereas for ‘atypical’ notation the pathways linking the perceptual symbols with the required mental operation were not as effortlessly traversed, slowing processing and attenuating the RP effect.

This is also consistent with the idea of transfer-appropriate processing, an account that suggests that as well as perceptual feature overlaps the correspondence between mental processes engaged during study and test displays can enhance or impair facilitation and affect RP (Morris, Bransford, & Franks, 1977; Roediger & Blaxton, 1987).

The robust RP in the face of perceptual changes in these tasks can be explained with reference to facilitation through the component processes that dominate the task set (e.g., Schacter, 1990). In the case of object naming, RP results from a priming mechanism(s) operating predominantly at a conceptual level, with additional advantages for perceptual overlaps (e.g., Biederman & Cooper, 1991; 1992; Fiser & Biederman, 2001). In word identification, RP can again be assumed to be dominated by conceptual facilitation as RP was robust to perceptual changes that did not tax perceptual processing. However, when identification required a difficult and atypical analysis of the word display the repetition advantage was attenuated (Graf & Ryan, 1990). This may be due to a lack of overlap, hence a lack of facilitation, in the key neural pathways required for performing the task. Similarly, in the arithmetic task, RP was attenuated when processing difficulty was increased due to 'atypical' presentation format (Sciama et al., 1999), again perhaps due to a lack of overlap in the processes required to compute the response when the same arithmetic equation was presented. Hence, RP at least partially reflects the facilitation of processing through component processes that overlap during repeated presentations of stimuli, with the greatest advantages perhaps being gained at the level of processing at which the task, and therefore attention, is directed (Jacoby, 1983; Kirsner, Milech, & Stumpfel, 1986; Schacter, 1990).

2.2 Episodic retrieval

From the perspective of episodic retrieval models, RP results from the automatic encoding and subsequent retrieval of individual stimulus-response trials. Key models under this perspective include the *Instance Theory of Automatisation* (Logan, 1988; 1990), the *Event File* theory (Hommel, 1998; 2004; 2009), and the *Stimulus-Response Binding and Retrieval* (S-R) theory (Dobbins et al., 2004; Henson et al., 2014; Horner & Henson, 2009; 2011; 2012).

The instance theory of automatisation was initially developed to explain improved performance when learning had occurred. According to Logan (1988; 1990), when a stimulus is first encountered, algorithmic processing is required to compute the correct response. With additional trials, an accumulation of previous 'instances' of the same stimulus and response pairings builds up and this accumulation allows responses to be automatically retrieved from memory, rather than requiring re-computation. There are three key assumptions in this theory. First, memory encoding is an automatic and unavoidable result of attention. Second, memory retrieval is also an automatic and unavoidable result of attention. Third, every stimulus-response 'instance' is encoded, stored, and retrieved as a separate trace. Accordingly, novice performance is assumed to be slower and more error prone than expert performance due to deficiencies in knowledge (instances) rather than being due to a lack of resources (processing capacity), and improved performance, from additional repetitions of stimulus-response trials, is assumed to be the result of an accumulation of traces rather than the result of strengthening in a single trace.

Importantly, automaticity is said to be characterised by a transition from performance being dominated by algorithmic processing (transmission through component processes) to performance being dominated by the retrieval of previous associations, and RP is considered to be an early manifestation of this process (Logan, 1988; 1990). In contrast with models in the facilitation account, Logan assumes no improvement in algorithmic processing itself. Instead, all the advantage of repetition comes from the accumulation of instances. As the distribution for this accumulation of 'instances' can be described mathematically, it is concluded that memory retrieval will eventually outperform the algorithm. Furthermore, according to Logan (1988), even if the algorithm was to speed up (as per the facilitation in component processes perspective) memory retrieval would still eventually win and would dominate performance.

The event file theory was proposed by Hommel (1998; 2004), who investigated the role of response-related information in the episodic representation of a stimulus. The theory was built on the notion of an ‘object file’, a temporary representation of an object where perceptual features experienced together are bound together in a form that enables subsequent reviewing and updating (Kahneman & Treisman, 1984; Kahneman et al., 1992). In keeping with Logan (1988; 1990), Hommel argued that the integration of perceptual and response features is an automatic process with some selectivity in what is bound into the file, that the bindings take place at a local level between various stimulus and response features, and that these local bindings can themselves be integrated to form global bindings. He also claimed that retrieval is automatic and where the retrieved file is incongruent with the current display responses would show an interference effect. Moreover, task irrelevant features are not excluded from the binding process, instead, the strength of the binding between various features of the stimulus and the response (and the speed at which the binding decays) is modified by relevance and, importantly, response always interacts with stimulus features.

In line with the instance theory of automatisation (Logan, 1988; 1990) and the event file theory (Hommel, 1998; 2004), the S-R theory posits that RP reflects the retrieval of direct bindings between the initial stimulus and the resultant response, bypassing the intervening layers of computation (Dobbins et al., 2004). Such bindings encapsulate the complete and context-dependent features of the stimulus event, simultaneously encoding multiple levels of response including action, decision and classification (Horner & Henson, 2009; 2011; Logan, 1988; 1990; Race et al., 2009). However, and in contrast with Logan (1988; 1990), encoding is not an obligatory consequence of attention. In fact, attention is not considered to be required for encoding although it can serve to increase the strength of bindings (Henson et al., 2014), and such bindings can form from only one exposure to the S-R pairing (Dobbins et al., 2004; Henson et al., 2014). These bindings have been explained as an “action-trigger” with a repeated stimulus triggering the previous response through perceptual or conceptual associations with the original stimulus (Kunde et al., 2003).

Evidence for S-R bindings has been found in a number of studies. Using a study-test block design, Horner & Henson (2009) investigated S-R bindings in an object classification task in which participants had to answer one of two questions; “Is the object bigger than a shoebox?” or “Is the object smaller than a shoebox?” The blocked reversal of the question when the same stimulus was repeated served as a response component manipulation, changing the decision from ‘yes’ to ‘no’. RP was found when the same decision was made in the study and test blocks, and the magnitude of RP increased when more repetitions were presented at study. When the question was reversed at test (e.g., from a “bigger than” question to a “smaller than” question), the magnitude of RP was comparatively smaller and remained stable regardless of the number of study trials.

In subsequent experiments, Horner & Henson (2009) manipulated the motor component of the response (verbal or finger press) and/or the task itself (classify or name) between study and test. A graded RP effect was observed. RP was strongest when all response components (classification, decision, and motor-action) were repeated and became gradually weaker as additional response components were changed, leaving minimal RP when all were changed. The authors therefore concluded that once RP associated with response manipulations was accounted for there was little, if any, RP left. In other words, the RP found in the study could all be attributed to S-R bindings.

However, a different conclusion was reached by Valt and colleagues (2015), who manipulated multiple levels of response independently and/or additively in a face classification task, and found evidence of both facilitation and S-R bindings. RP was robust to a change in decision (i.e., “Is the celebrity an actor/actress?” or “Is the celebrity a non-actor/non-actress?”) after one study presentation, and this result was taken as evidence of facilitation at the level of semantic classification. However, after two study presentations and, critically, when the decision at test was changed compared with the decision in the second study block, RP was reduced but still present. This indicates a degree of interference that is predicted by S-R binding at the level of stimulus-decision, alongside robust RP as expected under the facilitation perspective. When additional levels of the response were also manipulated, there was a RP gradient in line with Horner & Henson (2009). However, strong RP was still present even when all the response components were changed. This suggests that facilitation in component processes at both a perceptual and a conceptual level was an important contributor to the RP effect. Interestingly, although this facilitation effect was retained in the RT data regardless of changes to response components, the accuracy data showed clear interference effects when the decisions were reversed at test. This difference may indicate that measures of RT reflect the influence of facilitation and measures of accuracy reflect the influence of S-R bindings (Soldan et al., 2012). Furthermore, it may indicate that facilitation and retrieval rely on separate underlying processes (Soldan et al., 2012), that both facilitation and S-R binding contribute to RP (Horner & Henson, 2011; Lamy et al., 2011), and that they do so in an interactive and co-operative manner (Soldan et al., 2012; Valt et al., 2015).

Results that can be interpreted in favour of facilitation and/or retrieval accounts have also been reported using the prime-probe paradigm with tasks utilising numerical stimuli. Dehaene and colleagues (1998) used a masked prime-probe design in a task that required participants to categorise a single-digit probe as greater than or less than five. The manipulation involved changing the notation of the digits (word form or Arabic digit) so that the prime and probe could be presented in the same or different notation, as well as being the same or a different number. The results showed RT advantages for probes that had been preceded by an identical prime and by primes belonging to the same category, regardless of notation. As the primes were subliminal, the RT advantage from primes belonging to the same category, but not being the same number, was argued to show that unconscious processing could proceed through semantic associations to the corresponding motor response (Dehaene et al., 1998).

In a follow-up study, Naccache & Dehaene (2001) grouped the stimuli so that some numbers were never presented in the probe displays and could not, therefore, have been directly associated with a completed motor response. In this case, RP was again evident regardless of notation and, importantly, it was evident for all stimuli, even when the prime had never been consciously processed as a probe. It was therefore claimed that subliminal priming was, at least partially, due to facilitation in semantic processing (Naccache & Dehaene, 2001). Additionally, Reynvoet and colleagues (2002) used the same masked prime-probe design in a task requiring the categorisation of Arabic digits as odd or even, and changed the mapping between category and response hand between blocks of trials. The results showed that RTs were faster to probes when they were preceded by a prime that was close in value, and fastest when the two were identical, evidence of a semantic feature gradient in mathematical priming. Furthermore, the RT gradient did not depend on previous conscious processing of the primes. This was interpreted as an indication that binding had occurred between the semantic category and the motor response (Reynvoet et al., 2002).

The results from several studies reviewed above provide evidence that is in line with both facilitation and retrieval (Dehaene et al., 1998; Horner & Henson, 2009; Naccache &

Dehaene, 2001; Reynvoet et al., 2002; Valt et al., 2015). As such, it seems sensible to investigate the overlaps between the two perspectives with a view to their integration. Such an integration seems to allow for a priming mechanism(s) to operate anywhere within the processing domain, and to be dominated by facilitation of component processes or the retrieval of specific traces, depending on task demands (e.g., Lamy et al., 2011; Soldan et al., 2012). Furthermore, it is conceivable that *S-R bindings/event files/instances* are an index of task relevant *synchronisation* culminating in improved functional connectivity between local and global networks involved in stimulus processing and response production, manifesting as RP.

As the instance theory of automatisation (Logan, 1988; 1990) explicitly includes both algorithmic (information flow through component processes) and retrieval processes, this seems a logical place to start a discussion on such an integration.

2.3 Integration of facilitation and retrieval perspectives

In the instance theory of automatisation, Logan (1988; 1990) envisaged each stimulus event as a race between algorithmic and retrieval processes, with more instances leading to faster and more accurate retrieval that then tends to 'win' the race. This has similarities with the S-R theory, where component processes are bypassed and RP results from the direct retrieval of the previous S-R episode (Dobbins et al., 2004; Henson et al., 2014). However, there are also fundamental differences, and these differences highlight important questions.

According to Logan (1988; 1990), the development of bindings is a gradual process, during which the algorithm continues but is eventually beaten by the improved speed of retrieval processes. As there is no mention of what might transpire once performance is dominated by retrieval, it is unclear what happens to the algorithmic processing once a stable 'solution' has been reached and retrieval has 'won'. It seems unlikely that a highly evolved and energy conserved system would continue expending resources on algorithmic processing once it becomes superfluous.

Instead of considering binding as a gradual process, the S-R theory regards it as the result of a single exposure to a stimulus event. By this account, the component processes (algorithm) that were required to compute the response for the first stimulus event are bypassed on repetition rather than beaten over time (Dobbins et al., 2004). Evidence of such bindings is obtained in the form of a response congruency effect. When a stimulus is presented it acts as a cue, prompting the retrieval of the previous S-R episode/event file/instance involving that stimulus, and so retrieving the previous response. If the current display is incongruent with the retrieved display (i.e., the word is presented in a different colour or the classification question is different), this incongruence creates conflict and causes interference resulting in slower and more error prone behaviour (Hommel, 1998). An obvious question to ask is: what elicits the interference if the component processes are bypassed and the previous response is retrieved directly? It appears that without component processes (algorithms) operating at the same time as retrieval processes, the perceptual representation and/or response representation would be based on the retrieved stimulus event, not that which would be computed if component processes continued. As such there would be no conflict to resolve and so no interference when the response differed from that elicited by the same stimulus at the previous encounter.

In integrating the two perspectives, three key ideas need to be established. First, the algorithm and retrieval can operate simultaneously, as per the instance theory of automatisation (Logan, 1988; 1990). Second, the algorithm can speed up with repetition, as per models of RP under the facilitation in component processes perspective (e.g., Gotts et al., 2012; James & Gauthier, 2006; Kohn & Movshon, 2003; Maljkovic & Martini, 2005; Wiggs

& Martin, 1998). Third, performance can be dominated by either facilitation or retrieval depending on features of the task (e.g., Lamy et al., 2011; Soldan et al., 2012).

Under the assumption that these ideas are correct, the integration of the two perspectives then raises additional questions. If algorithmic and retrieval processes both continue and feed into a system that compares the outputs to maximise the adaptiveness of behaviour, how and where would such a comparison and decision process occur? Once retrieval has 'won' the race and the behaviour has become automatic, what happens to the algorithm? Does it transition to operating at a higher order (more abstract) level of processing? If so, what would constitute evidence of such a transition? And can evidence of this be obtained under current experimental paradigms? Or else, is the algorithm terminated? In which case, how and where is this signalled? Are partial traces then discarded or encoded and retained? And if they are retained, what would constitute evidence of partial traces? And can evidence of these partial traces be detected?

For now, these questions remain unanswered, but recent studies combining behavioural and neural imaging measures provide support for the integration of algorithmic and retrieval processes. Using a visual object classification task in a blocked study-test design, Horner & Henson (2012) measured neural responses using functional magnetic resonance imaging (fMRI) and electroencephalogram (EEG). They observed repetition suppression (i.e., a reduction in the blood oxygen level dependent (BOLD) response) in the fMRI analysis when recording over the prefrontal regions that are associated with response processing, and over the occipital and temporal regions that are associated with perceptual and conceptual processing of visual information. However, a reversal of the question in test blocks elicited repetition enhancement (i.e., an increase in the BOLD response) in prefrontal regions and this repetition enhancement was accompanied by a reversal of the response-locked event related potential (ERP) signal. These fMRI and EEG findings were interpreted as the neural markers of interference due to the retrieval of an incongruent S-R binding (Horner & Henson, 2012). Furthermore, at the same time the repetition suppression in occipital and temporal regions remained stable. This suggests that facilitation in early processing regions is unaffected by whether a response is congruent or incongruent and that algorithmic processing is not wholly bypassed.

In fact, it has been shown that improvements in RT reflect facilitation while interference effects from retrieving incongruent S-R episodes are reflected in the accuracy data (e.g., Soldan et al., 2012; Valt et al., 2015) and that interactions between facilitation and binding emerge when S-R pairings have been experienced at least three times prior to the test phase (e.g., Valt et al., 2015). Furthermore, the magnitude of RP is greatest for immediate repetition and only increases in small increments for additional repetitions (Lander et al., 2009; Logan, 1998; 1990; Maljkovic & Nakayama, 1996; Treisman, 1992) while response congruency effects are smallest after one repetition and increase in magnitude with additional repetitions (Dobbins et al., 2004; Horner & Henson, 2009; Valt et al., 2015). Hence, it seems likely that both facilitation and retrieval models are required to explain RP. For example, it is possible that, given the constrained nature of cognitive processing resources, a degree of facilitation in early sensory and/or late motor processing regions is necessary in order to ‘free-up’ resources that allow a ‘binding’ mechanism to be engaged.

3. The Present Experiments

Early research suggested that numerical information is represented in networks that are structured in a similar way to that of semantic information (e.g., Ashcraft & Battaglia, 1978; Ashcraft & Stazyk, 1981), in which nodes for associated concepts are closer together than nodes for un-associated concepts (e.g., Collins & Loftus, 1975; Rosch, 1975). This suggestion gained support from subsequent studies where a gradient in RTs was observed with responses being slowest when the numerical distance between the prime and probe was greatest (e.g., Naccache & Dehaene, 2001; Reynvoet et al., 2002). Furthermore, abstract numerical information and the more conceptually elaborative linguistic information appear to rely on a common underlying syntactic structure (e.g., Scheepers et al., 2011; Scheepers & Sturt, 2014). However, to date no direct comparison has been undertaken in terms of whether RP effects are also equivalent in tasks that require semantic categorisation of words and arithmetic categorisation of equations (but see Logan, 1988; 1990, for a similar comparison). As the study of RP can reveal the nature of an underlying learning mechanism(s) that gives rise to processing efficiencies gained from stimulus repetition, it is important to know whether the mechanism(s) that underlies RP with conceptually rich stimuli such as words and objects is the same as that which underlies RP in mathematical cognition. This question is particularly important for the contemporary society, because of the growing reliance on technology and the introduction of computer coding as a core curriculum area within primary school education in many countries.

The experiments reported in this thesis were designed to compare the magnitude of RP in a linguistic and arithmetic task in a study-test block design. The linguistic task involves classifying English words as belonging to the category of 'animal' or 'object'. The arithmetic task involves classifying 2-operand equations of the form '3+26' or '26+3' according to whether the answer is odd or even. Perceptual manipulations have been shown to vary in their effect on RP depending on computational difficulty and the length of time it takes for participants to respond (Soldan et al., 2012). Furthermore, under the episodic retrieval perspective, task irrelevant perceptual features are hypothesised to bind more strongly when a high degree of attention is required for the task (Henson et al., 2014; Logan, 1988; 1990). As such, a task irrelevant perceptual feature, colour, is manipulated in Experiments 1-3 to compare the effect that colour has on RP in the two tasks. The linguistic task is relatively effortless and likely involves a high degree of parallel processing so can be regarded as highly automatic (Stroop, 1935). However, the arithmetic task is computationally effortful, requires a high degree of attention and likely involves a degree of serial processing (Sackur & Dehaene, 2009). As such, the colour manipulation tests for a difference in perceptual congruency effect (in which the magnitude of RP changes as a function of the perceptual congruency between the study and test presentations) based on computational differences between the two tasks.

The present study also seeks to extend current research on the role of S-R bindings in RP that underlies mathematical cognition. Most studies that investigated the role of S-R bindings in RP used conceptually rich stimuli such as visual images of familiar objects (e.g., Dobbins et al., 2004; Horner & Henson, 2009; Race et al., 2009) or photos of famous faces (e.g., Valt et al., 2015) in study-test block designs, or numerical stimuli (e.g., Dehaene et al., 1998; Kunde et al., 2003; Naccache & Dehaene, 2001; Reynvoet et al., 2002) in prime-probe designs. To date, there has only been one study that investigated S-R bindings in a study-test design with numerical stimuli (e.g., Salimpoor et al., 2010). However, because response-related components were not directly manipulated in that study, it is unclear whether the behavioural improvement is more closely aligned with facilitation or retrieval.

Salimpoor and colleagues (2010) compared RP and neural repetition effects, using fMRI in both a visual search task (requiring responses according to whether or not the digit ‘5’ was present in a visual array), and a verification task (requiring the categorisation of mathematical equations as correct or incorrect). In the visual search task, there was evidence of RT improvements and repetition suppression (i.e., decreased BOLD response) for repeated displays. As these measures were not correlated, the RT improvements were interpreted as being due to speeded stimulus identification and the repetition suppression was said to reflect ‘general cognitive processes’ (Salimpoor et al., 2010). In the verification task, both repetitions suppression and repetition enhancement (i.e., increased BOLD response) were observed when stimuli were repeated. Linear regression showed that RP, in the form of RT improvements, was positively correlated with repetition enhancement in medial temporal lobes, but not with repetition suppression in any region. RT improvements were also associated with increased connectivity between the hippocampus, the dorsal mid-cingulate cortex, and the supplementary motor area (Salimpoor et al., 2010). Due to these correlations, the enhancement of neural responses in the math task was interpreted as being due to the formation and retrieval of S-R memory traces, in line with episodic retrieval theories of RP.

This interpretation does not explain the lack of neural enhancement found in the visual search task, an important consideration given that episodic retrieval theories seek to explain priming across the board rather than just in a subset of cognitively challenging tasks. As the verification task is computationally challenging and explicitly utilises working memory (while the visual search task is not so demanding), it is possible that the repetition enhancement effects found by Salimpoor and colleagues (2010) were due to the engagement of working memory, previously shown to increase neural activation (Desimone, 1996), rather than being a neural signature of the formation of S-R bindings. Furthermore, the conclusion that observed behavioural improvements were due to S-R bindings was based on a correlational analysis between RT improvements and neural repetition effects, with no manipulation of a response component. Such a manipulation is regarded as critical when investigating S-R bindings (Horner & Henson, 2011; Henson et al., 2014). Hence, it remains to be seen whether behavioural evidence of a response congruency effect can be elicited in a mathematical task using a study-test block design. As such, a response feature, the hand used to respond, is manipulated in Experiments 4 and 5 to compare the effect this has on RP in the two tasks. Due to hypothesised differences in the strength of bindings from attention (Henson et al., 2014; Logan, 1988; 1990; Treisman, 1992) and task difficulty (Lamy et al., 2011;

Soldan et al., 2012) it is possible that evidence of a response congruency effect is more likely to manifest in the arithmetic task than in the word task.

As variations across conditions during study are more likely to be captured in the relative-difference method rather than in the absolute-difference method, the relative-difference method provides a more complete picture of RP. In the experiments reported here, the primary data analyses are conducted using the relative-difference method. Secondary analyses are conducted using the absolute-difference method and reported in Appendix B. This enables a comparison to be drawn between the two methods of analysis in order to ascertain to what degree data analysis strategies may influence the findings in RP.

In summary then, the experiments reported here were designed with the following objectives in mind: (1) to compare the magnitude of RP between a word and a math task; (2) to determine whether RP would be affected to a similar degree in the two tasks by a change in a task irrelevant perceptual feature dimension (Experiments 1-3) or by a change in motor action (Experiments 4 and 5); and (3) to compare the RP results obtained under the relative-difference and absolute-difference methods of analysis.

3.1 Experiment 1

Experiment 1 had three objectives. First, to compare the magnitude of RP in a linguistic and an arithmetic categorisation task. As numerical information is thought to be represented in semantic networks that are analogous to those representing linguistic information (e.g., Ashcraft & Battaglia, 1978; Ashcraft & Stazyk, 1981) and RT gradients are linked to semantic relationships between repeated stimuli in numerical tasks (e.g., Naccache & Dehaene, 2001; Reynvoet et al., 2002), RP (in the form of the relative improvement in performance from study to test) is expected to be comparable in the word and the math task.

Second, to determine whether changes in a task irrelevant feature, in this case colour, would affect the magnitude of priming, and if so, whether the effect on priming would depend on the cognitive domain investigated. While it was unknown how the manipulations at test might affect priming in a linguistic task relative to an arithmetic task, there is evidence for structural priming between the two domains (e.g., Scheepers et al., 2011; Scheepers & Sturt, 2014). Scheepers and Sturt (2014, Experiment 1) used a prime-probe couplet paradigm, with the prime being an arithmetic equation that had a left- or right-branching structure (e.g., 3 x 4 + 6 vs. 3 + 4 x 6), and the probe a linguistic expression that also had a left- or right-branching structure (e.g., with alien monster movie vs. lengthy monster movie). The task was to solve the equation and then to rate the sensicality, on a 1-5 scale, of the linguistic expression. A robust priming effect was found. Participants rated the linguistic expressions as being more sensical when the prime and probe had the same structure compared with when they had different structures. In a subsequent experiment, structural priming was again found when the prime was linguistic expressions and the probe was arithmetic equations. These results suggest that arithmetic and language share syntactic representations. They also raise the possibility that if an aspect of a stimulus is manipulated between study and test in both a word task and a math task, the effect on priming might be comparable between the two tasks. However, under the episodic retrieval perspective, the relatively high degree of attention that is required for the math task, compared with the word task, may render that task more sensitive to feature binding (e.g., Henson et al., 2014; Logan 1988; 1990). If that is the case, a perceptual congruency effect would be expected in the math task but not the word task.

The third objective of the experiment was to compare results using the relative-difference method of analysis with those using the absolute-difference method of analysis, and to determine whether the former is a more accurate measure of RP than the latter.

Method

Participants. Forty eight participants between the ages of 18 and 46 years (M = 20.5 years, SD = 4.7 years) were recruited from the University of Canterbury (11 males and 37 females) in return for course credit or a $10 voucher.

Apparatus and stimuli. The experiments were presented on a PC with a 50-cm x 30-cm monitor in width and height. E-Prime 2.0 with a refresh rate of 60 ms was used to generate the stimuli and to collect responses. Participants were tested individually in a dimly lit room. The viewing distance was approximately 60 cm.

Each trial consisted of a central fixation followed by a word or a mathematical equation presented at the center of the screen. The fixation was a black cross that extended 0.06 degrees of visual angle in both width and height. Both the words and the mathematical equations were written in Courier New, font size 40. Depending on the experiment condition, the stimuli were either black or coloured. In the latter condition, the colours used were black, blue, cyan, green, lime, magenta, maroon, navy, olive, orange, purple, and red.

In the word task, participants determined whether the word referred to an animal or an object. The stimulus set consisted of 184 words that varied in length from three to nine letters. Half of them referred to an animal and half to a household object. Half were long words that had more than five letters, and the other half were short words that consisted of five or fewer letters. These 184 words were then randomly assigned to condition ensuring each mini-block of eight was fully balanced. They were then imported into E-prime. In the math task, participants determined whether the answer to the equation was an odd number or an even number. The stimulus set consisted of 216 equations. All equations were in the format of a single-digit number plus a double-digit number, with half of them starting with the single-digit number (e.g., 5 + 15) and the other half starting with the double-digit number (e.g., 16 + 4). Half of the equations had an odd answer and half had an even answer. Half of the equations were categorised as difficult and half as easy. Equations were categorised as difficult when they involved addition with a 'unit' digit that was greater than 6 or when they involved addition that crossed the boundary between 'tens' (i.e., 16 + 5 crosses from the 'tens' to the 'twenties'). These 216 equations were then randomly assigned to condition ensuring each mini-block of eight was fully balanced. They were then imported into E-prime.

Design and procedure. The experiment used a 2 (task: word vs. math) x 2 (feature: identity (ID) vs. colour) x 2 (condition: same vs. change) mixed design, with task being a between-subjects factor. Participants were randomly assigned to either task and completed that task in one sitting. Both tasks included two sessions: an ID session where the identity of the stimulus was manipulated (same vs. change) and a colour session where the colour of the stimulus was manipulated (same vs. change). Only one stimulus dimension varied between the study and test blocks in each session. In other words, in the ID session, all stimuli had the same colour in both the study and test blocks. Likewise, in the colour session, all stimuli had the same identity in both the study and test blocks. In both sessions, only stimulus identity was task relevant while colour was task irrelevant. The order of the two sessions was counterbalanced across participants.

Both tasks followed the same procedure and within each task the two sessions followed the same blocked design with four rounds of study-test cycles. Each of these four cycles consisted of two mini-blocks of eight study words/equations followed by two mini-blocks of eight test words/equations. This gave a total of 32 trials per study-test cycle and 128 trials per session, for a total of 256 trials in each task.

In the ID session, all stimuli were black. In the ID-Same condition, the words or equations used in the study phase were re-used in the test phase. In the ID-Change condition, different words or equations were used in the study and test phases. No words or equations were used twice within a task except to fulfil the repetition condition. In addition, the set of words or equations used in the ID session was not re-used in the colour session, so there was no overlap of stimuli between the two sessions. In the colour session, all words or equations were presented in coloured font and the identity of the stimuli in the study block was always repeated in the test block. In the Colour-Same condition, the stimuli had the same colour in both the study phase and the test phase. In the Colour-Change condition, the stimuli had one colour in the study phase but a different colour in the test phase. No colours were used twice except to fulfil the repetition condition. Within each study-test cycle, the order of the two mini-blocks at test was randomised and the order of presentation of individual stimuli within each study and test mini-block was also randomised. Hence, the maximum possible lag between presentations of the same stimuli was 30 trials and the minimum was zero, with an average lag of 15 intervening trials. Figure 1A shows the four conditions as they relate to the study and test cycles in both the ID and colour sessions for the word task, the conditions were the same for the math task.

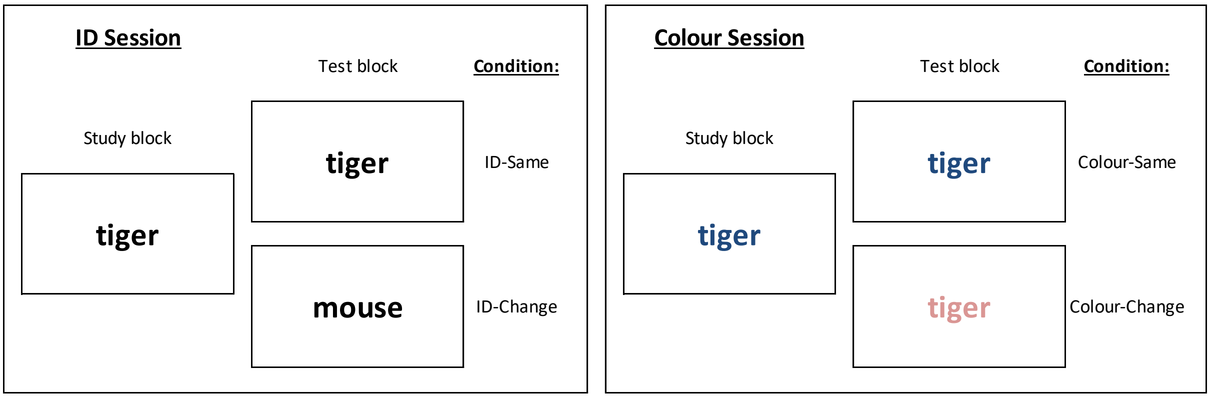


Figure 1A. A schematic representation for the ID and colour sessions in the word classification task in Experiment 1. The allocation of words to mini-block was the same for every participant but the order in which the words were presented within each mini-block and the order of the mini-blocks within each of the study and test cycles was randomised. Note that the identity of the stimuli in the colour session was always the same between the study and the test blocks regardless of whether there was a change in colour between the two blocks.

Each trial began with a fixation cross for 1000 ms, followed by a target display for 120 ms, and then a blank screen until a response had been made. The inter-trial stimulus interval was 1000 ms. The participants used the index and middle fingers of their right hand to press one of the two labelled keys on a computer keyboard. In the word task, the participants pressed the “o” key if the word referred to an animal, and the “p” key if the word referred to an object. In the math task, they pressed the “o” key if the answer to the equation was an odd number, and the “p” key if it was an even number.

Each participant first completed a brief practice block of 32 trials before proceeding to the experiment proper. All participants were encouraged to take a break between the two sessions. The total amount of time for the experiment was approximately 20 minutes.

Results and discussion

The median RT was chosen for all the analyses reported here after first assessing the distributional skew and investigating any potential advantages of using square root or log transformations. In all the experiments reported here, analyses of variance (ANOVAs) were first conducted on the difference scores between the study and test blocks across all conditions in RTs[1] and error rates. Full results from these ANOVAs are provided in Appendix A. Where appropriate (i.e., there was a priori prediction for RP), planned t-tests were also carried out to assess the statistical significance of RP in individual conditions. Tables showing the t-test results are also provided in Appendix A and attention will be drawn to them where appropriate.

[1] As the RTs in the math task were twice as long as the RTs in the word task, it was known in advance that there would be potential issues of heteroscedasticity and homogeneity of variance. As such, all experiments were analysed under square root and logarithm transformations to investigate these issues, as well as to ensure the assumption of normality was met. The results remained the same under these transformations and all assumptions were shown to be met. As it was decided to report results in the actual units of measure, for ease of understanding the magnitude of the effects, no further remarks will be made concerning these assumptions and they should be regarded as having been met for all experiments.

For each participant, the median difference score was first calculated between the study and test blocks (Study – Test) in each condition, and these difference scores were then used to compute RP. In other words, the statistical analyses and the subsequent interpretation of the results reported below, unless otherwise noted, are based on the data using the relative-difference method of analysis. To allow an exploratory comparison between these results and results using the absolute-difference method, statistical analyses were also conducted using the data from the latter method. These results are presented in Appendix B and comparisons are drawn where appropriate.

Data exceeding 3SD (both above and below) from each individual participant's mean RT were excluded, and this resulted in the exclusion of less than 2% of the data. In the word task, data from two participants was also excluded due to the average error rate exceeding 25% in one or more condition and data from one further participant was excluded due to their median RT being more than 3SD above the average of the median RTs for all participants in the word task. In the math task, data from four participants was also excluded due to the average error rate exceeding 25% in one or more condition. The means of median RTs and error rates for the remaining participants are shown in Table 1A for the word task (N=21) and Table 1B for the math task (N=20).

Table 1A. *Means of median reaction times, expressed in milliseconds (ms), and percentage of errors (%) for the classification of words in Experiment 1 (N=21).*

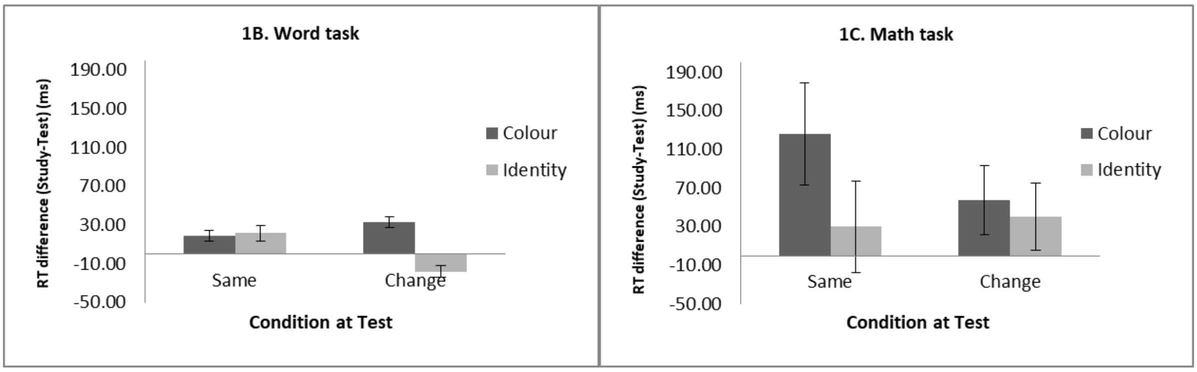
	Reaction Time (ms)				Percentage Error (%)			
	Study		Test		Study		Test	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ID-Same	586	70	565	60	5.0	5.3	3.6	3.9
ID-Change	599	79	617	74	4.8	5.6	7.8	5.3
Colour-Same	589	65	570	61	3.8	3.0	4.2	3.7
Colour-Change	617	70	584	63	6.3	5.6	5.6	4.9

Table 1B. *Means of median reaction times, expressed in milliseconds (ms), and percentage of errors (%) for the classification of equations in Experiment 1 (N=20).*

	Reaction Time (ms)				Percentage Error (%)			
	Study		Test		Study		Test	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ID-Same	1,277	497	1,247	509	6.6	5.1	6.7	4.1
ID-Change	1,398	527	1,358	552	9.9	6.3	8.5	6.4
Colour-Same	1,408	432	1,282	418	8.9	5.3	9.4	7.4
Colour-Change	1,408	490	1,350	450	9.4	5.9	11.1	7.6

Figures 1B and 1C show the mean of the RT difference score in each condition for the word and the math task, respectively. A 2 (task: word vs. math) x 2 (feature: ID vs. colour) x 2 (condition: same vs. change) mixed ANOVA was conducted on the RT data (i.e., the difference scores between the study and the test blocks). The results showed that the intercept was significantly different from zero, $F(1, 39) = 13.70$, $p = .001$, partial $\eta^2 = .26$. As difference scores were used in the analysis, this result indicated that RTs were faster at test than at study. There was also a main effect of task, $F(1, 39) = 5.74$, $p = .021$, partial $\eta^2 = .13$, indicating a larger RT improvement in the math mask (64 ms) than in the word task (14 ms). No other effects reached significance.

To test RP for each condition, t-tests for single samples on the study-test difference scores in RTs were conducted, and the results are provided in Appendix A, Tables 1E and 1F. In the word task, there were statistically significant RT improvements when identity was repeated, irrespective of whether the colour was the same or changed at test, indicating robust RP. Surprisingly, there was also a significant detrimental effect on RT from study to test (-18 ms) in the ID-Change condition, in which novel words were presented at test. (I will refer to this effect as a “novelty effect” in the text below.) In the math task there was a significant improvement in RT in the Colour-Same condition (126 ms). In the other three conditions, participants showed a numerical improvement in RT, but the improvement was not statistically significant.



Figures 1B & 1C. RT differences between the study and test blocks in the word and math classification tasks as a function of feature and condition in Experiment 1. Error bars show +/- 1 standard deviation of the mean. A positive number indicates faster RT in the test block than in the study block. A negative number indicates slower RT in the test block than in the study block.

A similar ANOVA was conducted on the error rates. The only significant result was a 3-way interaction of task, feature and condition, $F(1, 39) = 5.69$, $p = .022$, partial $\eta^2 = .13$. To clarify the interaction, two separate ANOVAs were conducted, one for the word task and the other for the math task. For the word task, the only significant result was a feature by condition interaction, $F(1, 20) = 6.32$, $p = .021$, partial $\eta^2 = .24$. Consistent with the results in RT, there was a statistically significant decrease in accuracy in the ID session when novel words were presented at test (-3.0%). No significant change in accuracy occurred when words were repeated (1.4%). In the colour session, there was no significant change in accuracy regardless of whether the colour changed (0.7%) or remained the same (-0.4%). In the math task there were no significant effects.

There are four key findings in Experiment 1. Firstly, in both tasks, a significant main effect of RP was found, indicating that RP can survive an average of 15 intervening trials. The finding of RP in the math task extended the results of a recent study by Salimpoor et al. (2010), who used a three-operand arithmetic verification task (e.g., 3 + 5 – 4 = 4) in a study-test block design with an average lag of 16, and found significantly faster RTs in the test block compared with the study block.

Secondly, colour change had no effect on the magnitude of priming in the word task, but there appeared to be a boost to processing efficiency when the same colour was retained at test in the colour session of the math task. This pattern of data is in line with previous studies (e.g., Horner & Henson, 2009), and with suggestions that task difficulty and the level of attention required to perform a task enhance the effects of feature binding (Henson et al., 2014; Logan, 1990; Soldan et al., 2012; Treisman, 1992). Perhaps the computational challenge in the math task promoted binding and/or promoted the longevity of binding between the colour and identity of equations.

There was no evidence that colour change reduced the magnitude of RP in either task. This was likely due to three factors: the tasks being semantic in nature, the repetition of stimulus identity between the study and test blocks in the colour session, and the large number of intervening trials between repeats. If RP arises primarily from task relevant features (Hillstrom, 2000; Maljkovic & Nakayama, 1996; but see Huang, Holcombe, & Pashler, 2004), repeating the identity of a stimulus between study and test should be sufficient to generate RP, with repetition in other task irrelevant features being immaterial. Because this experiment used a study-test block design with an average of 15 trials between repeats, the lack of a reduction in priming when changing colour at test is equivocal with regard to the issue of binding between colour and identity. It is possible that no binding occurred between colour and identity during encoding. It is also possible that there was binding, but it decayed during the intervening trials. This issue will be discussed more fully in the General Discussion (Section 4).

Thirdly, the overall improvement in RT from study to test was larger in the math task (64 ms, being a 5% improvement) than in the word task (14 ms, being a 2% improvement). This difference is likely to be partly due to the *novelty effect* in the word task, an effect not present in the math task. With regard to the error rates, the *novelty effect* was again present in the word task but not the math task and the overall change in accuracy for the two tasks was comparable, -0.35% in the word task and -0.20% in the math task.

The reason for the *novelty effect* in RT and accuracy for the word task is not clear, but there are a number of possible explanations. As the test block always follows the study block it is possible that fatigue played a role but as the same effect was not apparent in the math task fatigue is unlikely to be able to fully account for the impaired performance. The effect could also be due to some type of proactive interference where the activation of multiple related concepts interferes with processing through the accumulation of 'noise' within associative networks. A similar negative effect on subsequent processing has been reported in several studies focused on visual object priming and has been labelled as visual antipriming (e.g., Deason, 2008; Marsolek, Schnyer, Deason, Ritchey, & Verfaellie, 2006; Zhang, Fairchild, & Li, 2017). Under this explanation, there is a negative effect on current processing as a result of the previous activation of related information whose representations overlap and so compete with the current information (Marsolek et al., 2006; Marsolek, 2008; Zhang et al., 2017). Therefore, a potential reason for the difference between the word and the math task is that although words and numbers are thought to be represented in similar associative networks (e.g., Ashcraft & Battaglia, 1978; Ashcraft & Stazyk, 1981) there is something different about their cumulative processing. Namely, each word provides access to a unique and meaningful concrete concept that is embedded in a complex network of associated concepts while each equation results in a rather more abstract concept in the form of a numerical answer that can also be obtained through a large number of other equations. Hence, network activation from the processing of equations does not necessarily add additional 'noise' that could interfere with subsequent processing.

It is also possible that the effect is an artefact of the experimental design as the words were not randomly allocated to condition for each participant. Instead, all participants were presented with the same words in each condition (albeit in a different random order) and so there may have been something more difficult about the words presented in the novel test blocks. However, the math task was also subject to the same design features so whether this is a complete explanation remains to be seen. A further possibility is that, as this experiment used a between-subjects design, the lack of effect in the math task may indicate there was something different about the two groups of participants. These two possibilities are explored in Experiments 2 and 3, to follow.

Finally, it is notable that the *novelty effect* in the word task was found only when the data were analysed under the relative-difference method, which takes into account performance in both the study and test blocks. When the data were analysed under the absolute-difference method of analysis (see Appendix B), any evidence of this effect is masked by relying on measurements in the test blocks only. As a result, instead of an increase in RT and error rate from study to test in the ID-Change condition and a decrease in RT and error rate in the ID-Same condition, RP was found in the form of a RT and accuracy advantage in the test blocks for the ID-Same condition compared with the ID-Change condition.

The use of the two different methods of analysis led to differences in several other aspects of results, too. In the word task, whereas the relative-difference method showed no effect of a colour change, the absolute-difference method showed a significant RT impairment when changing the colour of words at test compared with retaining the same colour. In the math task, the relative-difference method indicated a numerical increase in RP in the Colour-Same condition compared to the Colour-Change condition, but the absolute-difference method showed no effect of manipulating the colour of repeated equations. The above differences in results using two different methods of analysis underscore the importance of taking into account the performance in the study blocks in addition to performance in the test blocks when calculating RP.

3.2 Experiment 2

Experiment 1 demonstrated that robust RP effects could be found in both a word and an arithmetic categorisation task. It also showed a *novelty effect* when the data were analysed using the relative-difference method. To further investigate these results, in particular whether the pattern of data had something to do with the population from which the participants came (i.e., mostly 1st year psychology students), a replication of Experiment 1 was undertaken using a within-subjects design with participants who were studying first year engineering mathematics. The goal was to examine whether the results of Experiment 1 were partly due to differences between participants.

Method

Participants. Twenty six new participants between the ages of 17 and 35 years (M = 19.5 years, SD = 3.7 years) were recruited from the University of Canterbury (17 males and 9 females). All participants were first year students studying engineering mathematics and participated in return for a $10 voucher.

Apparatus and stimuli. Both the apparatus and stimuli were the same as those used in Experiment 1.

Design and procedure. The experiment used a 2 (task: word vs. math) x 2 (feature: ID vs. colour) x 2 (condition: same vs. change) within-subjects design. The order of the tasks was counterbalanced, with half the participants completing the word task first, and the other half completing the math task first. Within those groups the order of the sessions were also counterbalanced, with half completing the ID session first and the other half completing the colour session first. All the other aspects of the procedure were the same as those in Experiment 1. The entire experiment took about 40 minutes to complete.

Results and discussion

The data were treated in the same way as that in Experiment 1, and this excluded less than 2% of the data. Data from nine participants was also excluded: one due to their average error rate exceeding 25% in one or more condition in the word task and four for the same reason in the math task, and three further participants due to their median RT being more than 3SD above the average of the median RTs for all participants, one in the word task, one in the math task, and one whose median RT exceeded the average in both tasks. The means of median RTs and error rates for the remaining participants are shown in Table 2A for the word task and in Table 2B for the math task.

Table 2A. *Means of median reaction times, expressed in milliseconds (ms), and percentage of errors (%) for the classification of words in Experiment 2 (N=18).*

	Reaction Time (ms)				Percentage Error (%)			
	Study		Test		Study		Test	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ID-Same	567	46	534	44	3.0	4.6	2.1	2.6
ID-Change	560	58	580	60	2.5	3.7	5.2	3.5
Colour-Same	552	57	526	50	2.3	3.4	1.6	2.7
Colour-Change	568	46	547	51	4.7	6.2	2.4	4.9

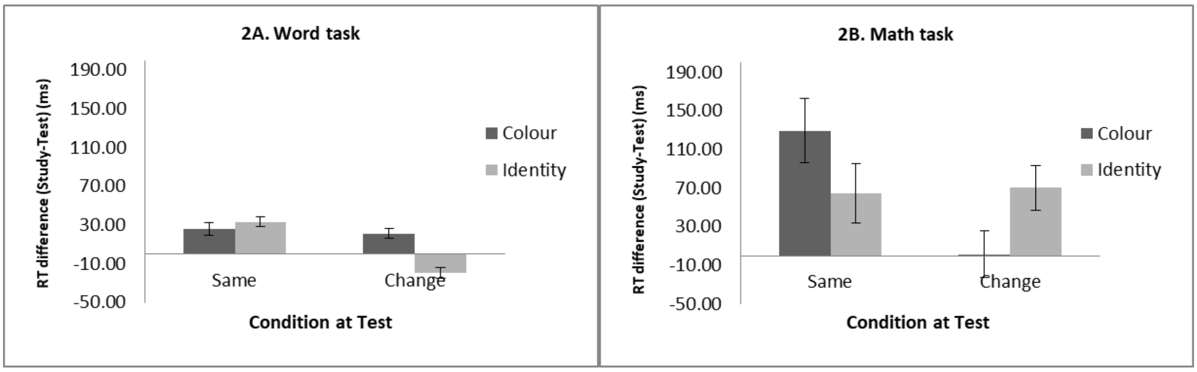
Table 2B. *Means of median reaction times, expressed in milliseconds (ms), and percentage of errors (%) for the classification of equations in Experiment 2 (N=18).*

	Reaction Time (ms)				Percentage Error (%)			
	Study		Test		Study		Test	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ID-Same	945	200	880	148	6.4	4.7	5.4	6.0
ID-Change	1,014	230	943	206	6.6	6.9	7.2	4.9
Colour-Same	1,045	231	915	182	6.7	6.1	7.2	4.1
Colour-Change	923	172	922	189	5.8	4.4	4.9	3.5

Figures 2A and 2B show the mean of the RT difference in each condition for the word and the math task, respectively. A 2 (task: word vs. math) x 2 (feature: ID vs. colour) x 2 (condition: same vs. change) repeated-measures ANOVA was conducted on the RT data, again using the difference scores between the study and the test blocks. The results showed that the intercept was significantly different from zero, $F(1, 17) = 16.70$, $p = .001$, partial $\eta^2 = .50$, indicating that RTs were faster at test than at study. There was also a main effect of task, $F(1, 17) = 6.89$, $p = .018$, partial $\eta^2 = .29$, with greater RT improvement in the math task (67 ms) than in the word task (15 ms), and a main effect of condition, $F(1, 17) = 9.30$, $p = .007$, partial $\eta^2 = .35$, indicating greater improvement in RT when features were repeated (63 ms) than when they changed (18 ms). In addition, task, feature, and condition interacted, $F(1, 17) = 13.36$, $p = .002$, partial $\eta^2 = .44$. To clarify the interaction, two separate ANOVAs were conducted, one for the word task and the other for the math task.

For the word task, the intercept was significantly different from zero, $F(1, 17) = 39.01$, $p < .001$, partial $\eta^2 = .70$, indicating that RTs were faster at test than at study. There was also a significant effect of feature, $F(1, 17) = 5.23$, $p = .035$, partial $\eta^2 = .24$, indicating RT improvements were greater in the colour session (24 ms) than the ID session (6 ms), and a significant effect of condition $F(1, 17) = 28.81$, $p < .001$, partial $\eta^2 = .63$, indicating RT improvements were greater when features were repeated (29 ms) than when they changed (1 ms). There was also a significant feature by condition interaction, $F(1, 17) = 25.37$, $p < .001$, partial $\eta^2 = .60$. In the ID session, the repetition of words at test resulted in an improvement in RT (33 ms) but when novel words were presented performance was impaired (-20 ms). In the colour session there were statistically significant RT improvements regardless of whether colour was repeated (26 ms) or changed (21 ms).

For the math task, the intercept was significantly different from zero, $F(1, 17) = 11.47$, $p = .004$, partial $\eta^2 = .40$, indicating that RTs were faster at test than at study. There was also a significant effect of condition $F(1, 17) = 5.35$, $p = .033$, partial $\eta^2 = .24$, indicating RT improvements were greater when features were repeated (97 ms) than when they changed (36 ms). There was also a significant feature by condition interaction, $F(1, 17) = 8.27$, $p = .011$, partial $\eta^2 = .33$. In the ID session, there was a statistically significant improvement in RT from study to test regardless of whether the equations were novel (70 ms) or repeated (65 ms). In the colour session, there was a RT improvement when the colour remained the same at test (130 ms), but no improvement when the colour of repeated equations was changed at test (2 ms). No other effects reached significance.



Figures 2A & 2B. RT differences between the study and test blocks in the word and math classification tasks as a function of feature and condition in Experiment 2. Error bars show +/- 1 standard deviation of the mean. A positive number indicates faster RT in the test block than in the study block. A negative number indicates slower RT in the test block than in the study block.

A similar ANOVA was conducted on the error rates. Task and feature interacted, $F(1, 17) = 5.23$, $p = .035$, partial $\eta^2 = .24$. In the word task, accuracy improved from study to test in the colour session (1.5%) but worsened in the ID session (-1%). In the math task, the improvement was the same in both sessions (0.2%). Feature and condition also interacted, $F(1, 17) = 4.99$, $p = .039$, partial $\eta^2 = .23$. For the ID session, accuracy improved between the study and test blocks in the ID-Same condition (1.0%) but not in the ID-Change condition (-1.6%), indicating a positive effect of repetition and a detrimental effect of novelty in test blocks. For the colour session, there was an improvement in the Colour-Change condition (1.6%) but not in the Colour-Same condition (0.1%). The latter result indicates no interference when colour changes from the study to the test block. No other effects reached significance.

The results of Experiment 2 generally cohere well with those from Experiment 1. There was again a greater RT improvement (67 ms, being a 7% improvement) in the math task than the word task (15 ms, being a 3% improvement) and a comparable effect on accuracy in both tasks, an improvement of 0.30% in the word task and 0.19% in the math task.

In the math task, once again, there was an absence of impairment when presenting novel equations at test, with comparable RT improvement for both the ID-Same (65 ms) and ID-Change (70 ms) conditions, perhaps suggesting that the activation of related equations adds to general processing efficiency rather than resulting in interference. This could be due to the many-to-one relationship between two-operand equations and a numerical sum as well as being due to the abstract nature of the stimulus domain. Furthermore, relative to the other conditions, a processing boost was again apparent in the Colour-Same condition when presenting equations in the same colour at test, and, in addition, there was no RP when the colour changed at test. As mentioned before, a boost for feature congruency and a cost for feature incongruency are in line with previous studies (e.g., Horner & Henson, 2009). These results suggest that task difficulty and attention may promote functional binding between identity and task irrelevant perceptual features (e.g., Henson et al., 2014; Logan, 1990; Soldan et al., 2012; Treisman, 1992).

In the word task there was again a *novelty effect* in both RT and accuracy. This adds weight to the idea that the activation of multiple related items may have a negative effect on information processing, as per the concept of visual antipriming (e.g., Marsolek, 2008). The replication of a different pattern of effects for the word task and the math task while using a within-subjects design make it unlikely that the effect reported in Experiment 1 was due to participants. Furthermore, it suggests that the *novelty effect* may rely on stimuli providing access to concrete rather than abstract representations. However, there is also a possibility that the non-randomisation of stimuli to condition played a role in the manifestation of this effect, this idea is investigated in Experiment 3.

The two methods of analysis again led to different patterns of results. Under the absolute-difference method there was evidence of a perceptual congruency effect in the word task, with longer RTs when the colour changed at test than when it remained the same (See Appendix B, Table 2B). This difference suggests functional binding between the identity and colour of words. In the math task, the effect of colour was absent (see Appendix B, Table 2B), suggesting no functional binding between colour and the identity of equations. These results can, at least partially, be attributed to not including measures of RT from the study blocks when using the absolute-difference method of analysis.

The absolute-difference method is built on the assumption that performance is comparable across all novel items. The finding of the *novelty effect*, which was only revealed when the data were analysed in the relative-difference method, shows that this assumption is not correct, at least in the present paradigm. Because the absolute-difference method excludes the data from the study blocks, this method cannot detect potential improvements (from repetition) or potential impairments (from novelty) in performance from study to test. As a result, the magnitude of RP could be exaggerated when data were analysed in the absolute-difference method.

It is possible that the non-randomisation of stimuli to condition could contribute to the pattern of data found in Experiments 1 and 2 due to some stimulus-specific differences across conditions. To eliminate this possibility, it is necessary for an additional layer of randomisation to be introduced in the experimental design. In Experiment 3, all stimuli were randomised to condition for each participant at the beginning of the experiment.

3.3 Experiment 3

Experiments 1 and 2 showed consistently greater relative improvements for equations than for words, a consistent detrimental effect on processing when presenting novel words but not when presenting novel equations in test blocks, inconsistent effects of the colour manipulation for equations and words, and inconsistencies between analysing the results using the relative-difference and absolute-difference methods. The purpose of Experiment 3 is to investigate the influence that randomisation of stimuli may have on these findings.

Method

Participants. Thirty one new participants between the ages of 17 and 44 years (M = 20.8 years, SD = 5.7 years) were recruited from the University of Canterbury (9 males and 22 females) in return for course credit.

Apparatus and stimuli. Both the apparatus and stimuli were the same as those used in Experiments 1 and 2.

Design and procedure. The experiment used a 2 (task: word vs. math) x 2 (feature: ID vs. colour) x 2 (condition: same vs. change) within-subjects design. At the beginning of each experiment, stimuli were randomised to condition ensuring equal numbers of words that referred to an animal or object and that were long or short, and equal numbers of equations that had an odd answer or an even answer, that began with a single or a double digit, and that were rated easy or difficult. All other aspects of the procedure were the same as those in Experiment 2. The entire experiment took about 40 minutes to complete.

Results and discussion

The data were treated in the same way as that in Experiments 1 and 2, and this excluded less than 2% of the data. Data from five participants was also excluded due to the average error rate exceeding 25% in one or more condition in the math task, and data from three further participants was excluded due to their median RT being more than 3SD above the average of the median RTs for all participants (one in the math task and two whose median RT exceeded the average in both tasks). The means of median RTs and error rates for the remaining participants are shown in Table 3A for the word task and in Table 3B for the math task.

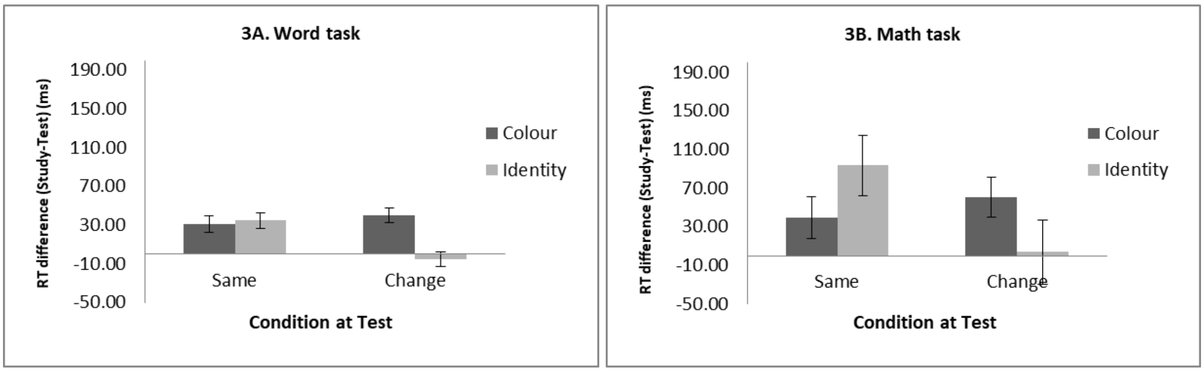
Table 3A. *Means of median reaction times, expressed in milliseconds (ms), and percentage of errors (%) for the classification of words in Experiment 3 (N=23).*

	Reaction Time (ms)				Percentage Error (%)			
	Study		Test		Study		Test	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ID-Same	619	77	584	58	3.2	5.1	3.5	4.1
ID-Change	620	81	625	77	3.7	4.6	5.1	6.7
Colour-Same	614	63	584	62	3.5	3.5	3.2	3.8
Colour-Change	632	81	592	56	5.1	5.2	4.9	5.2

Table 3B. *Means of median reaction times, expressed in milliseconds (ms), and percentage of errors (%) for the classification of equations in Experiment 3 (N=23).*

	Reaction Time (ms)				Percentage Error (%)			
	Study		Test		Study		Test	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ID-Same	1,186	211	1,093	233	8.5	5.9	6.8	5.9
ID-Change	1,209	241	1,205	276	9.6	6.9	7.8	5.1
Colour-Same	1,135	259	1,095	264	8.5	6.9	7.7	6.4
Colour-Change	1,142	251	1,081	255	9.4	6.5	4.8	3.6

Figures 3A and 3B show the mean of the RT difference score in each condition for the word and the math task, respectively. A 2 (task: word vs. math) x 2 (feature: ID vs. colour) x 2 (condition: same vs. change) repeated-measures ANOVA was conducted on the RT data (i.e., the difference scores between the study and the test blocks). The results showed that the intercept was significantly different from zero, $F(1, 22) = 32.87$, $p < .001$, partial $\eta^2 = .60$, indicating that RTs were faster at test than at study. There was also a main effect of task, $F(1, 22) = 4.51$, $p = .045$, partial $\eta^2 = .17$, with a larger RT improvement in the math mask (49 ms) than in the word task (25 ms). In addition, feature and condition interacted, $F(1, 22) = 8.96$, $p = .007$, partial $\eta^2 = .29$. For the ID session, the magnitude of RT difference between the study and test blocks was significantly larger in the ID-Same condition (64 ms) than in the ID-Change condition (-1 ms), indicating RP. For the colour session, there was no significant difference in priming between the Colour-Same condition (35 ms) and the Colour-Change condition (50 ms). The latter result indicates no reduction in priming regardless of whether there was a colour change from the study to the test block. No other effects reached significance.



Figures 3A & 3B. RT differences between the study and test blocks in the word and math classification task as a function of feature and condition in Experiment 3. Error bars show +/- 1 standard deviation of the mean. A positive number indicates faster RT in the test block than in the study block. A negative number indicates slower RT in the test block than in the study block.

A similar ANOVA was conducted on the error rates. The intercept was again significantly different from zero, $F(1, 22) = 5.70$, $p = .026$, partial $\eta^2 = .21$, indicating more accurate responses at test than at study. The main effect of task was also significant, $F(1, 22) = 17.77$, $p < .001$, partial $\eta^2 = .45$, suggesting that the participants made more improvement between the study and test blocks in the math task (2.2%) than in the word task (-0.3%). No other effects were significant.

The most important finding from Experiment 3 is the similarity in results between the math and word tasks when stimuli are randomised to condition for each participant. Specifically, while there was still a slight impairment in accuracy (-1.4%) and RT (-5 ms) when novel words were presented at test neither of these effects was statistically significant (see Appendix A, Tables 3C & 3E). As a consequence, although the absolute improvement in RT from the study to the test block was still larger in the math task (49 ms) than in the word task (25 ms), the percentage change from study to test, a more accurate measure for RP in the present paradigm, was 4% in both tasks. This indicates that the magnitude of RP was comparable in the two tasks. There was also no clear effect of changing the colour of stimuli on the magnitude of priming in either the word task or the math task. This suggests that any advantage of repetition is mediated by efficiencies gained at a semantic level, and if colour does bind with the identity of a stimulus then this binding decays over the intervening lag to the point that it is ineffective. Taken together, these results show there is very little difference in the magnitude of priming, or in the effect of colour on priming, between the word task and the math task once an extra layer of randomisation is built into the experimental design.

In terms of a comparison with the absolute-difference method of analysis, the key finding is that the two methods now show the same pattern of the results. This suggests that where stimuli are carefully randomised to condition for each participant any differences at study are likely to be minimised and so the choice of data analysis strategy is less critical to the results.

3.4 Experiment 4

Experiments 1-3 demonstrated that robust RP could be found in both a word and an arithmetic categorisation task, that the magnitude of priming was comparable in the two tasks, and that changing the task irrelevant feature of colour did not reduce the magnitude of RP in either task. A comparison of the two methods of data analysis (absolute-difference and relative-difference) revealed inconsistencies and indicated that attention needs to be paid to experimental design as well as data analysis method.

In Experiment 4 the binding between an object feature and the hand used to make the response was investigated. Instead of varying ID or colour between the study and test blocks, the response hand was manipulated so that participants either used the same hand to press the response keys in both the study and test blocks, or switched hand between the study and the test block. The goal was to examine the effect of response change on the magnitude of RP, and to see whether such an effect, if found, would be modulated by task type and/or attentional demand. Under the facilitation perspective, no change in RP would be predicted in either task as efficiencies would most likely be gained at a semantic level. Under the episodic retrieval perspective, it could be argued that a response congruency effect would be predicted in the math task (but not the word task) due to computational difficulty and attentional requirements. A second goal was to continue the investigation into what effects the randomisation of stimuli and data analysis strategy may have on RP results.

Method

Participants. Thirty three new participants between the ages of 18 and 31 years (M = 19.5 years, SD = 2.5 years) were recruited from the University of Canterbury (13 males and 20 females) in exchange for course credit.

Apparatus and stimuli. Both the apparatus and stimuli were the same in Experiment 4 as those in Experiments 1-3 (with non-randomisation of stimuli to condition as with Experiments 1 and 2), except for the following three differences. First, the test blocks always contained the same words or equations as the study blocks, and all stimuli were presented in black. Thus, unlike the previous three experiments, there was no change in identity or colour between the study and test blocks. As in Experiments 1 and 2, the stimuli were the same in each mini-block for each participant, and the stimuli within each min-block were presented in a random order. Second, each study-test cycle consisted of one mini-block of eight study items followed by one mini-block of eight test items. This resulted in a maximum possible lag between repeated stimuli being 14 trials and the minimum being zero, with an average lag of 7 intervening trials. The length of the intervening lag was reduced to increase the sensitivity of the experiment, as there is evidence that priming effects are stronger with reduced lags (e.g., Henson et al., 2004). Third, each mini-block was preceded by an instruction display, which informed the participants which hand (left or right) they should use to perform the task in the subsequent trials. In total, each participant was presented with eight 'Hand-Same' study-test cycles interwoven with eight 'Hand-Change' study-test cycles. In the 'Hand-Same' cycle, the same hand was used at study and test (four left hand and four right hand cycles). In the 'Hand-Change' cycle, the response hand was changed from study to test (four left to right and four right to left cycles). These four combinations were randomised and presented an equal number of times to each participant.

Design and procedure. The experiment used a 2 (task: word vs. math) x 2 (condition: same vs. change) within-subjects design. The order of the tasks was counterbalanced, with half the participants completing the word task first, and the other half completing the math task first.

As in Experiments 1-3, each trial began with a fixation display for 1000 ms followed by the task display for 120 ms, and then a blank screen until response. The inter-trial stimulus interval was again 1000 ms. After every eight trials, based on instruction, the participants either used the same hand to respond or switched to the other hand. For right hand responses, the participants pressed the same keys as in Experiments 1-3, i.e., the “o” key for “odd” or “animal”, and the “p” key for “even” or “object”. For left hand responses, they pressed the “w” key for “even” or “object”, and the “e” key for “odd” or “animal”. All the other aspects of the procedure were the same as those in Experiments 1-3. The entire experiment took about 45 minutes to complete.

Results and discussion

The data were treated in the same way as that in Experiments 1-3, and this excluded less than 2% of the data. Seven participants’ data were not included in further analyses, 6 due to high error rates (exceeding 25% in one or more condition) and 1 due to long RTs (more than 3SD above the average of the median reaction times for all participants). Table 4 shows the means of median RTs and error rates in each condition. As in Experiments 1-3, the difference scores between the study and test blocks were calculated, and Figure 4 shows the means of the RT difference scores.

Table 4. *Means of median reaction times, expressed in milliseconds (ms), and percentage of errors (%) for the classification of words and equations in Experiment 4 (N=26).*

	Reaction Time (ms)				Percentage Error (%)			
	Study		Test		Study		Test	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word Hand-Same	660	123	604	68	4.9	3.5	4.0	4.6
Word Hand-Change	641	96	608	68	5.4	3.9	5.0	3.7
Math Hand-Same	1,237	309	1,158	313	11.7	6.3	9.8	6.6
Math Hand-Change	1,293	364	1,218	366	11.4	6.6	11.5	6.9

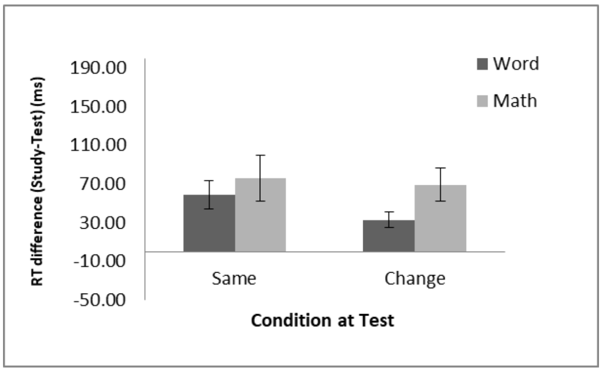


Figure 4. RT differences between the study and test blocks as a function of condition in Experiment 4. Error bars show +/- 1 standard deviation of the mean.

Two 2 (task: math vs. word) x 2 (condition: same vs. change) repeated-measures ANOVAs were conducted, one on the RTs and the other on the error rates, again using the difference scores between the study and the test blocks. The intercept was significantly different from zero for RTs, $F(1, 25) = 46.27$, $p < .001$, partial $\eta^2 = .65$, and marginally significant for accuracy, $F(1, 25) = 3.98$, $p = .057$, partial $\eta^2 = .14$, indicating faster and more accurate performance at test than at study. No effect of task was found in the RTs or error rates, suggesting similar improvements between study and test in both tasks. Importantly, there was no significant main effect of condition in terms of RT, indicating robust RP regardless of a change in motor response at test. However, there was a main effect of condition in terms of error rates, $F(1, 25) = 5.04$, $p = .034$, partial $\eta^2 = .17$, with greater improvements in accuracy between study and test blocks when the same hand was used (1.8%) than when the hand changed (0.13%). No other effects were significant.

There are two key findings from Experiment 4. Firstly, the magnitude of RP is again comparable in the two tasks, a 7% improvement in RT in the word task and a 6% improvement in the math task. Secondly, evidence of a response congruency effect differed by task and dependent measure. There was no response congruency effect in terms of RT, with speed improving in both the word task and the math task regardless of whether the same hand or a different hand was used between study and test. However, in the math task accuracy only improved when the same hand was used in both the study and test blocks.

Increasing the number of study trials has been reported to increase the strength of binding and reveal robust response congruency effects (e.g., Dobbins et al., 2004; Horner & Henson, 2009; Valt et al., 2015). So, it is possible that the single pairing of the stimulus and the motor response at study was insufficient to result in strong functional binding. Furthermore, as RT results are only based on correct responses, it could be argued that the earliest evidence of a response congruency effect would be most likely to be revealed in the accuracy data. A similar difference between dependent measures was also reported in a perceptual object classification study (e.g., Soldan et al., 2012) and in a conceptual face classification study (e.g., Valt et al., 2015). In these cases the difference was interpreted as a possible indicator that the two dependent measures, RT and accuracy, index efficiency gains from different underlying processes. Namely, RT advantages come from facilitation in component processes and accuracy advantages are driven by the retrieval of S-R bindings (Soldan et al., 2012; Valt et al., 2015).

Also of interest is that the improved accuracy was only statistically significant in the math task (see Appendix A, Table 4D), in line with suggestions that high computational demands and long RTs may drive the retrieval of S-R bindings as indexed by measures of accuracy (e.g., Soldan et al., 2012). This apparent boost for accuracy when using the same response hand is analogous to the boost found in the RT data when retaining the same colour on repetition of equations at test in Experiments 1 and 2.

Results from the absolute-difference method of analysis show evidence of a response congruency effect in the math task but not the word task. Responses were faster and, marginally, more accurate when the same hand was used at test compared to when the hand was switched (see Appendix B). It is probable that at least part of this effect is due to the lack of accounting for performance in the study phase of the experiment when calculating the results, exacerbated by the lack of randomisation of stimuli to condition. As such, a replication of Experiment 4 was carried out in which stimuli were randomised to condition for participants at the beginning of each experiment, as in Experiment 3.

3.5 Experiment 5

The purpose of Experiment 5 is to investigate the influence that randomisation of stimuli may have on the findings reported in Experiment 4.

Method

Participants. Twenty seven new participants between the ages of 17 and 59 years (M = 21 years, SD = 8.3 years) were recruited from the University of Canterbury (5 males and 22 females) in return for course credit or a $10 voucher.

Apparatus and stimuli. Both the apparatus and stimuli were the same in Experiment 5 as those in Experiment 4 except that stimuli were randomised to condition for each participant, as in Experiment 3.

Design and procedure. The experiment used a 2 (task: word vs. math) x 2 (condition: same vs. change) within-subjects design. The order of the tasks was counterbalanced, with half the participants completing the word task first, and the other half completing the math task first.

As in Experiment 4, after every eight trials, based on instruction, the participants either used the same hand to respond or switched to the other hand. For right hand responses, the participants pressed the “o” key for “odd” or “animal”, and the “p” key for “even” or “object”. For left hand responses, they pressed the “w” key for “even” or “object”, and the “e” key for “odd” or “animal”. All the other aspects of the procedure were the same as those in Experiment 4. The entire experiment took about 45 minutes to complete.

Results and discussion

The data were treated in the same way as that in Experiments 1-4, and this excluded less than 2% of the data. Five participants’ data were not included in further analyses, 3 due to high error rates (exceeding 25% in one or more condition) and 2 due to long RTs (more than 3SD above the average of the median reaction times for all participants). Table 5 shows the means of median RTs and error rates in each condition for the remaining participants. As in Experiments 1- 4, the results are based on difference scores between the study and test blocks, and Figure 5 shows the means of the difference scores for RT.

Table 5. *Means of median reaction times, expressed in milliseconds (ms), and percentage of errors (%) for the classification of words and equations in Experiment 5 (N=22).*

	Reaction Time (ms)				Percentage Error (%)			
	Study		Test		Study		Test	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word Hand-Same	630	84	590	64	3.9	3.2	3.1	3.6
Word Hand-Change	626	76	591	67	4.8	5.4	3.8	4.7
Math Hand-Same	1,245	312	1,159	273	9.6	6.4	7.8	4.9
Math Hand-Change	1,213	302	1,158	254	9.3	6.1	9.2	5.7

Two 2 (task: math vs. word) x 2 (condition: same vs. change) repeated-measures ANOVAs were conducted, one on the RTs and the other on the error rates, again using the difference scores between the study and the test blocks. The intercept was significantly different from zero in both RTs, $F(1, 21) = 28.46$, $p < .001$, partial $\eta^2 = .58$, and accuracy, $F(1, 21) = 4.96$, $p = .037$, partial $\eta^2 = .19$, indicating faster and more accurate performance at test than at study. The main effect of task in RT was marginally significant, $F(1, 21) = 4.09$, $p = .056$, partial $\eta^2 = .16$, suggesting a larger improvement between the study and test blocks in the math task (70 ms) than in the word task (38 ms). This difference was primarily due to longer RTs in the math task (1194 ms) than in the word task (609 ms). In terms of percentage change from study to test this equated to a 6% improvement in RT in each task. No effect of task was found in the error rates. Importantly, there was no significant main effect of condition, or task by condition interaction, in either RTs or error rates. These results indicate that a change in hand between the study and test blocks did not reduce the magnitude of RP in either the word or the math task.

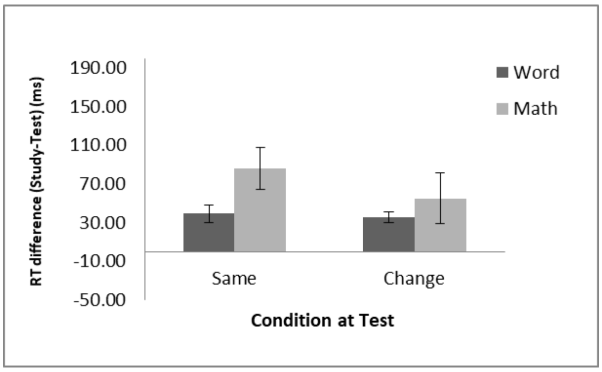


Figure 5. RT differences between the study and test blocks as a function of condition in Experiment 5. Error bars show +/- 1 standard deviation of the mean.

Once again, participants showed significant priming of a comparable magnitude in both tasks. Perhaps of more importance, though, a change in response hand did not affect the magnitude of priming in either task or in either dependent measure, and the results of the absolute-difference method of analysis are in agreement with these findings. Furthermore, these results provide evidence that cohere with the findings of Experiment 3. Once stimuli have been randomised to condition, there is comparable performance between the word and math tasks, there is no effect of manipulating task irrelevant features on either dependent measure, and there is consistency between results from the two methods of analysis.

The absence of a response congruency effect in motor action may appear to be inconsistent with previous studies that showed a reduction in priming when a change in motor action occurred. Using a study-test block design, Horner and Henson (2009, Experiment 6) showed participants pictures of objects, and the task in the test block was to respond whether the object was bigger than a shoebox by pressing one of two response keys for a “Yes” or a “No” answer. In the study block there were three conditions: the same action/same decision condition, in which both the action (i.e., keypress responses) and the decision (i.e., Yes/No answer) were the same as those in the test block; the different action/same decision condition, in which the same Yes/No task as in the test block was performed with verbal responses; and the different action/different decision condition, in which the task was to name the target object verbally. Compared to new stimuli, priming was significantly larger in the same action/same decision condition than in the different action/same decision condition, which in turn was larger than in the different action/different decision condition. Thus, priming was reduced when there was a change in motor action or a change in decision. (See also Dobbins et al., 2004, for a related finding.)

There are many methodological differences between Horner and Henson’s (2009) experiment and Experiment 5 in the present study. These differences include the type of stimuli (objects vs. words/equations), response decision (yes/no vs. animal/object or odd/even), presentation duration (2000 ms vs. 120 ms), motor action (verbal/keypress vs.

keypress only), and the method of data analysis (absolute-difference vs. relative-difference), among others. Although the exact cause for the difference in results is unclear, a possible candidate that might contribute to the different results between the two experiments could concern the manipulation of motor action, which was between verbal and keypress responses in Horner and Henson, but between left and right hand responses in the present experiment. It is possible that a change in motor action requires more attentional resources when the change is large (e.g., from a hand response to a verbal response) than when it is small (e.g., from a left hand response to a right hand response). As attention strengthens binding (Logan, 1990; Treisman, 1992), the S-R binding should be stronger in the former case than the latter case. Consequently, all else being equal, a response congruency effect is more likely to manifest in RP when the change in motor action is between different response modalities as in Horner and Henson's study than when the change is within the same response modality as in the present experiment.

4. General Discussion

The primary goals of this study were to compare RP in a linguistic and an arithmetic categorisation task and to compare the effects on behavioural priming of altering the colour or the response hand for repeated, otherwise identical, stimuli in these tasks. Secondary goals were to investigate the impact of experimental design and data analysis strategy on RP results.

Using a novel arithmetic categorisation task, the results reported here indicate that a single, brief, task relevant experience with an arithmetic equation affects subsequent performance over an average lag of 15 intervening trials. This extends previous findings (e.g., Salimpoor et al., 2010) to a new computationally challenging task within the domain of mathematical cognition.

No previous studies had used a within-subjects design to directly compare the magnitude of RP in a predominantly linguistic task with a task that requires arithmetic computation. Using a within-subjects study-test block design, participants completed both a word and a math task. The results show that the magnitude of RP was comparable between the two types of task, that changing a task irrelevant object feature such as colour or the hand used to respond had a negligible effect on RP in either task, and that the two methods of data analysis (absolute-difference and relative-difference) were only in agreement when full randomisation of stimuli between conditions and participants was used in the experimental design.

A further finding, and one that was unexpected, was the presence of a novelty effect. This effect resembles a previously reported impairment in processing that has been labelled visual antipriming (e.g., Marsolek et al., 2006), and is considered to be the natural antithesis of the benefits that promote RP (Marsolek, 2008). In the experiments reported here, the novelty effect appears only when pseudo-randomisation was used in Experiments 1 and 2. In these experiments, the stimuli within each mini-block were randomly presented and the presentation order of the mini-blocks in the experiment was randomly selected. However, which stimuli were used in a particular mini-block was not randomly selected for each participant. Furthermore, the novelty effect was found only when the data were analysed in the relative-difference method, a method in which performance is first assessed as the difference between study and test blocks and this difference score is then used to compare the magnitude of RP between experimental and control conditions.

Visual antipriming has previously been reported in tasks that required the identification of line drawings of familiar objects (e.g., Marsolek et al., 2006; Marsolek, Deason, Ketz, et al., 2010), and the identification of Chinese characters (e.g., Zhang et al., 2017). As Chinese characters can be considered as a pictorial form of orthography, evidence for antipriming has so far been found only with pictorial stimuli. According to Marsolek and colleagues (Marsolek et al., 2006; Marsolek, 2008), antipriming is a type of interference that results from competitive interactions from overlapping perceptual representations (Marsolek, 2008). Under this account there is an initial assumption that RP results from the strengthening of a representation, i.e., when an image is presented the neural pathways used to process it are potentiated. This strengthening leads to easier access on repetition and speeds processing for repeated stimuli, in line with the facilitation perspective. However, once neural pathways are biased towards strengthening one representation, that representation gains some degree of priority access to those pathways. When novel stimuli that are related to the previous stimuli are presented, the part of the representational structure that overlaps is no longer as easily accessible, and this slows performance down (Marsolek, 2008). In Experiments 1 and 2 of the present study, participants showed a novelty effect (i.e., antipriming) in the word task, and this extends the previous findings of visual antipriming from pictorial stimuli (e.g., Marsolek et al., 2006; Zhang et al., 2017) to a new stimulus domain (words), a new task (conceptual categorisation as opposed to perceptual identification), and a new experimental design (a simple study-test block design using a single task as opposed to a more complex four stage block design using multiple tasks). Hence, this novelty effect may be considered as evidence of conceptual antipriming.

Interestingly, a novelty effect was not found in the math task. This suggests that, while linguistic and arithmetic information may be represented in similar associative networks (e.g., Ashcraft & Battaglia, 1978; Ashcraft & Staykz, 1981) and rely on a common underlying syntactic structure (e.g., Scheepers et al., 2011; Scheepers & Sturt, 2014), there are fundamental differences in the cumulative effects of accessing and updating their representations. This may be related to the fact that arithmetic information is more abstract than linguistic information. Compared with equations, the words used in the present experiments represent more concrete and meaningful concepts that are embedded in a rich network of associations. Furthermore, as visual antipriming has been shown to increase with the degree of visual similarity between images (e.g., Deason, 2008), there would appear to be a strong perceptual component to the effect. It is possible that when participants were presented with words, the process of categorising them as an animal or an object prompted the activation of visual pathways associated with processing pictures of the animals/objects. This perceptual component may have enhanced the effect of interference from overlapping representations for words. However, the same would not apply to equations, as each equation does not provide access to a concrete representation that is as unique or as richly elaborative as a word stimulus.

Importantly, the conceptual antipriming effect observed in Experiments 1 and 2 was eliminated in Experiment 3 when stimuli were randomly selected to each condition for each participant, suggesting that the antipriming effect may have been an artefact of the experimental design. Recall that in Experiments 1 and 2, stimuli were randomly allocated to mini-blocks, which were then randomly allocated to conditions when the experiment was designed. However, the same allocations were subsequently used for all participants. In other words, for each participant, while the order of presentation of the conditions, mini-blocks, and stimuli within each mini-block were randomised, the same stimuli were used in each condition. This design bares strong similarities to that used in the visual antipriming studies, in which the same Chinese characters were used in each condition for each participant (e.g., Zhang et al., 2017). It is also similar to the pseudo-random assignment of stimuli to condition that has been used with visual object stimuli, although in this case there was also a degree of counterbalancing of stimuli between conditions and participants (e.g., Deason, 2008; Marsolek et al., 2006; 2010). Given that the novelty effect did not appear in Experiment 3, in which there was full randomisation of stimuli to condition for each participant, it is possible that in Experiments 1 and 2 the words randomly allocated to the study and test lists in the novel condition were highly related in a way that was optimal for providing the necessary overlapping representations that antipriming stems from. In Experiment 3, the stimuli to condition randomisation may have resulted in a randomisation of the relatedness of stimuli in the study and test blocks, and this reduced the likelihood of overlapping representations causing interference. It remains to be seen whether the visual antipriming effects in previous studies (e.g., Deason, 2008; Marsolek et al., 2006; 2010; Zhang et al., 2017) are retained under careful randomisation.

In terms of the effect of task irrelevant features on RP, there is some evidence that repeating the colour of equations could boost performance (Experiments 1 and 2), and retaining the same motor response in the math task could also boost performance (Experiment 4). It has previously been suggested that attention (Henson et al., 2014; Logan, 1988; 1990) and computational demands (Soldan et al., 2012) encourage S-R binding and retrieval. As the math task was computationally challenging and required a high degree of attention, the boost in performance in the math task (but not in the word task), which indicated binding and retrieval, would not be surprising. Importantly, though, the randomisation of stimuli to condition in Experiment 3 (colour manipulation) and in Experiment 5 (response hand manipulation) eliminated these effects, suggesting they may have been artefacts. That said, it is worth noting that both antipriming and the evidence of feature/response binding were task specific in that the former appeared only in the word task while the latter appeared only in the math task. The evidence of antipriming in the word task was perhaps due to the fact that the words used in the present study represented concrete concepts that are likely to invoke corresponding visual images. The evidence of feature/response binding in the math task was perhaps due to its difficulty and the high requirement for attentional resources. Hence, while non-randomisation was undoubtedly a factor in the manifestation of both the conceptual antipriming effect and the boost to RP for colour and hand repetition, it is perhaps not a complete explanation and further investigation is warranted.

Many previous priming studies have used study-test block designs and reported evidence of RP based on comparing performance between study and test blocks (e.g., Bentin & Moscovitch, 1988; Biederman & Cooper; 1991; 1992; Fiser & Biederman, 2001; Horner & Henson, 2009; Sciama et al., 1999), i.e., the results were calculated by using the relative-difference method of analysis. However, in those same studies, the effects on RP from manipulating various features at test, when otherwise identical stimuli are repeated, is based on a comparison between test blocks only (e.g., Biederman & Cooper; 1991; 1992; Fiser & Biederman, 2001; Horner & Henson, 2009; Sciama et al., 1999; but see Bentin & Moscovitch, 1988), i.e., the results were calculated by using the absolute-difference method of analysis. The potential consequences of using the latter method can be seen if we compare the results calculated under both methods of analysis (relative-difference and absolute-difference) for the experiments reported here. When careful randomisation of stimuli to condition was not undertaken such as in Experiment 1, 2 or 4, the results computed under the absolute-difference method differed from those under the relative-difference method. Specifically, in Experiments 1 and 2, under the absolute-difference method there was evidence of colour binding in the word task but not the math task, whereas under the relative-difference method, the pattern was reversed (i.e., evidence for colour binding in the math task but not the word task). In Experiment 4, there was evidence of response binding in both the RT and accuracy data of the math task under the absolute-difference method, but evidence of response binding was found only in the accuracy data under the relative-difference method. These differences indicate that by leaving out the performance in the study block, one runs the risk of computing spurious results. In contrast, when full randomisation to condition was undertaken such as in Experiments 3 and 5, the two methods of analysis yielded the same pattern of results, with neither method showing evidence of a perceptual congruency effect (Experiment 3) or a response congruency effect (Experiment 5). Furthermore, use of the absolute-difference method has the additional issue of exaggerating the reported RP effect, through conflating the positive and negative effects from study to test. These findings show the advantage of using the relative-difference method of analysis, and they also underscore the importance of employing true randomization in experimental design.

The lack of a colour congruency effect in Experiment 3 can be explained in one of several ways. First, priming occurred at an abstract semantic level where perceptual features were immaterial. This is in line with the proposal that priming operates at the level of processing at which the task is directed (Schacter, 1990). Thus, for a perceptual discrimination task, perceptual features will be primed, while for a conceptual categorisation task, semantic features will be primed (Jacoby, 1983; Kirsner, Milech, & Stumpfel, 1986; Schacter, 1990). Second, colour did not bind with the identity of the stimulus because it was task irrelevant (Hommel, 1998). The sense that colour was irrelevant was likely to be enhanced by the use of 12 different colour values in these experiments, and this led to the observed absence of a perceptual congruency effect in colour. It has been suggested that binary manipulations elicit a mutual inhibition effect, where the activation of one feature serves to inhibit the activation associated with the opposing feature (Hommel, 2009). According to this account, the use of two colours (i.e., red and green) creates a competitive interaction where repeating a word in red results in a behavioural advantage from both the facilitation of red and the inhibition of green, and this interaction is conducive to the manifestation of RP effects. In the present study, such a mutual inhibition effect was not elicited, as there were a dozen different colour values.

It is also possible that colour and identity did not bind because colour was simply ignored. As the strength with which features are bound depends on attention (Henson et al., 2014; Logan, 1988;1990; Treisman, 1992), an ignored feature may be precluded from the binding process, especially when stimulus displays are presented very briefly, which was the case (120 ms of target presentation duration) in these experiments. Finally, colour did bind but was not available upon retrieval. Whereas the identity of stimuli was retained over multiple intervening trials, the delay may have been too long for the binding of colour and identity to survive, in line with the claim that a delay in the repetition of an S-R pairing can lead to decay in the strength of binding (Hommel, 1998). Although each of the above interpretations can account for the absence of the perceptual congruency effect in Experiment 3, it is likely that the observed result was caused by more than one of the reasons stated above.

Previous research using numbers as stimuli has shown that changing task irrelevant features can affect the magnitude of priming in some situations but not in others. Naccache and Dehaene (2001) reported no reduction in priming in a masked prime-probe number categorisation task (probe smaller or larger than “5”) regardless of whether the prime and probe matched in notation (Arabic digit vs. word form). Using a study-test block design that required participants to perform a two-operand arithmetic task, Sciama et al. (1999) found attenuated priming in the notation change condition compared to the notation same condition when the numerals in the test blocks were presented in atypical notations such as in word form or dot array. These results were said to be due to atypical notations requiring additional attentional resources for arithmetic tasks, and the demand of attention in turn enhanced ‘form-specific’ associations. In the present study, Arabic digits were used in both the study and the test blocks. The results reported here are thus generally consistent with previous studies.

The factors discussed in relation to colour change in Experiment 3 can also apply to the results concerning hand change in Experiment 5. One may wonder whether response hand can be considered as a task irrelevant feature. After all, the use of the correct response hand was an important requirement in the experiments. That being said, it is important to remember that the task in these experiments was still about the semantic concept of the target stimuli, that the same hand was used within each mini-block of 8 trials, that the participants had as much time as they needed to prepare the hand change before each mini-block, and most important, that the participants would no longer need to be concerned about which hand to use once a mini-block of trials started. This is because they would already be using the correct hand to respond at the beginning of any mini-block of trials as they first had to press one of the correct response keys designated for the correct hand (left or right) to terminate the response instruction display so that the experiment could proceed. Taking all of these into account, it seems reasonable to consider response hand as being a task irrelevant feature in Experiments 4 and 5, similar to colour being a task irrelevant feature in Experiments 1-3. If this reasoning is correct, then the lack of a response congruency effect once stimuli were carefully randomised to condition (Experiment 5) could be due to priming occurring at an abstract semantic level in accordance with the behavioural goal, the failure of binding between stimulus identity and motor action, and/or the decay of the binding from study to test.

With regard to whether the RP observed in these experiments is best explained in terms of a facilitation account or a retrieval based S-R binding account, the results are equivocal. On the one hand, a facilitation account appears to be sufficient in accounting for the present results. The lack of a congruency effect once stimuli were randomised to condition in Experiments 3 and 5 is consistent with the notion that RP in the present paradigm was driven by faster stimulus identification of repeated words or equations at a semantic level irrespective of task irrelevant object features or motor action. There was no evidence of an interaction between different stimulus features or between a stimulus feature and response. The improved performance for repeated stimuli can thus be the result of facilitated neural processing, which may manifest as an overall reduction in the amplitude of neuronal activation (Dragoi, Sharma, Miller, & Sur, 2002; Kohn & Movshon, 2003), a sharpening of response tuning in local networks of neurons (Freedman, Riesenhuber, Poggio, & Miller, 2006; Wiggs & Martin, 1998), and/or more rapid onset of neural activation (James & Gauthier, 2006).

On the other hand, the lack of a response congruency effect cannot rule out an S-R binding account. As discussed before, it was possible that binding between stimulus identity and motor action had occurred but did not survive the intervening trials. This could be due to the flexibility of S-R bindings (Henson et al., 2014). It may be that while S-R bindings are hypothesised to require only one pairing (Dobbins et al., 2004; Henson et al., 2014) there are constraints on that assumption. Repeated exposures are always considered as an important component of learning, as emphasised by Logan (1988) in his instance theory of automatisation. Although facilitation in component processes can occur as a result of only one presentation, the formation of robust functional bindings between a stimulus and response seems to require additional presentations (e.g., Horner & Henson, 2009; Valt et al., 2015). Perhaps the requirement for additional instances provides a means to avoid indiscriminate learning that could tie up resources and limit future learning potential (Henson et al., 2014). As such, the accumulation of repetitions may somehow provide a signal of behavioural importance that changes the initially transient nature of the bindings into a more permanent state. Hence, the lack of congruency effect may be a result of transience in the initial binding due to limited exposure to identical S-R pairings prior to a response change.

It was also possible that stimulus identity was simultaneously bound to multiple response codes. Horner and Henson (2009) identified three levels of response codes: the motor action level (e.g., left or right hand used in response), the decision level (e.g., a yes or no response), and the task specific classification level (e.g., a bigger than or a smaller than question). They showed that each level can contribute to S-R binding. Only the motor action was manipulated in this study and so it was conceivable that the binding between identity and the other two levels of response code remained intact between the study and test blocks. Hence, no reduction in RP was found.

As evidence of a response congruency effect has been shown to be sensitive to the number of repetitions (e.g., Horner & Henson, 2009; Valt et al., 2015), a limitation of this study is that there was only one study and one test presentation. However, with this single repetition comparable RP effects are found for a linguistic and an arithmetic task. Hence, it seems that the capacity for facilitation and/or memory retrieval is the same for conceptual processing of words and arithmetic equations, at least initially. In the future, it will be important to add additional study presentations to these two tasks in order to determine three things: (1) Does the magnitude of RP remain comparable between the tasks with additional repetitions? (2) Do additional repetitions lead to response congruency effects in these tasks? And if so, (3) is the magnitude of a response congruency effect comparable for linguistic and arithmetic categorisation tasks?

Additionally, in Experiment 3 with an average lag of 15 intervening trials the RT improvement was 4% for both tasks and when that lag was halved in Experiment 5 the RT improvement increased to 6% for both tasks. Logan (1988) reported that RT decreased as a power function of the number of repetitions in a lexical decision and an alphabet arithmetic task. A question worth attempting to answer is whether the relative magnitude of RP is also able to be modelled as a function of the number of intervening items, and whether this function is comparable for the linguistic and arithmetic tasks.

RP has been shown to be impervious to perceptual changes when the task is conceptual in nature and so does not require close attention to perceptual features (Graf & Ryan, 1990; Jacoby, 1983; Kirsner, Milech, & Stumpfel, 1986; Schacter, 1990). Hence another limitation of Experiments 1 to 3 is that it could be argued that the use of colour as a feature manipulation in these tasks was unlikely to affect RP. However, binding between stimulus features has been claimed to increase due to task relevance (Hommel, 1998), attention (Henson et al., 2014; Logan, 1988; 1990; Treisman, 1992), and computational demands (Soldan et al., 2012). Hence, it was important to use a feature manipulation that was irrelevant and unlikely to result in effects in the relatively effortless word task in order to ascertain whether the same would be true in the computationally demanding math task. The lack of perceptual congruency in both tasks now allows these tasks to be used with feature manipulations that require additional attention to decode (for example: blurred font as per Masson, 1986; or ‘atypical’ notation as per Sciama et al., 1999) in order to determine whether the effect on the two tasks remains comparable. It will also be interesting to see if using only two colours in these tasks (rather than the 12 used in this study) leads to the engagement of a mutual inhibition effect (as per Hommel, 2009) and results in a perceptual congruency effect, and whether such an effect differs by stimulus domain and task difficulty.

A further limitation is the use of a single response component manipulation in Experiments 4 and 5. Classification, decision, and motor manipulations have been shown to have independent and additive effects on RP (e.g., Horner & Henson, 2009; Valt et al., 2015). Hence, perhaps a response congruency effect would be more likely with manipulations directed at every level. A way to investigate this in the future would be to add an orthogonal response manipulation (i.e., in the math task: “is the answer odd/even?” changes to “is the answer greater/less than 30?”) to the current design and so enable the classification, decision, and/or motor components to be reversed simultaneously.

In terms of conceptual antipriming, in future studies it will be important to replicate this effect for word stimuli using the designs under which visual antipriming was originally reported (e.g., Marsolek et al., 2006; 2010; Zhang et al., 2017). It will also be important to manipulate the semantic relationships between the stimuli to see if this interacts with conceptual antipriming in the same way that visual similarity interacts with visual antipriming (e.g., Deason, 2008). Furthermore, it is of interest that Ashcraft and Stazyk (1981) predicted that if numerical information is represented in networks that are analogous to those of linguistic information then as related information becomes activated interference would slow RTs. This type of interference seems to lie at the heart of antipriming and so future studies should also carefully manipulate the relatedness of arithmetic equations to test this prediction.

In summary, the experiments reported here showed a remarkable similarity in the pattern of priming between a word and a math task, both when there was a colour change and a response hand change. These results extended previous findings in mathematical cognition, and provided supporting evidence for the proposal that RP does not vary with stimulus domain (Ward et al., 2013) and that mathematical and linguistic information relies on common representative structures (e.g., Ashcraft & Battaglia, 1978; Ashcraft & Stzyk, 1981; Scheepers et al., 2011; Scheepers & Sturt, 2014). As RP is regarded as an early indicator of learning (Henson et al., 2014; Logan, 1988; 1990), the comparable relative improvement in the math and the word task suggests that, on the basis of a single study trial, learning capacity for numerical and linguistic information may be equivalent. The results also extended previous reports of visual antipriming to a novel conceptual task and raised the question of whether the phenomenon was limited to a specific type of experimental design. In terms of data analysis method, the present study indicates that the relative-difference method provides a more complete picture of repetition effects than the absolute-difference method, and that the choice of method should, at least partly, be driven by considerations of experimental design. Finally, with respect to the mechanisms that give rise to the RP phenomenon, it seems that the best way forward is to continue investigations into the integration of facilitation and retrieval perspectives.

References

Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science*, *11*(2), 118-124.

Abrams, R. L., Klinger, M. R., & Greenwald, A. G. (2002). Subliminal words activate semantic categories (not automated motor responses). *Psychonomic Bulletin & Review*, *9*(1), 100-106.

Ashcraft, M. H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(5), 527-538.

Ashcraft, M. H., & Stazyk, E. H. (1981). Menatal addition: A test of three verification models. *Memory & Cognition*, *9*(2), 185-196.

Bentin, S., & Moscovitch, M. (1988). The time course of repetition effects for words and unfamiliar faces. *Journal of Experimental Psychology General*, *117*, 148-148.

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, *20*(5), 585-593.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 121-133.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407-428.

Deason, R. G. (2008). *Masked antipriming: A behavioral and event-related potential examination of short-term antipriming effects* (Doctoral dissertation, University of Minnesota).

Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J. B., Le Bihan, D., & Cohen, L. (2004). Letter binding and invariant recognition of masked words behavioral and neuroimaging evidence. *Psychological Science*, *15*(5), 307-313.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597-600.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences*, *93*(24), 13494-13499.

Dobbins, I. G., Schnyer, D. M., Verfaellie, M., & Schacter, D. L. (2004). Cortical activity reductions during repetition priming can result from rapid response learning. *Nature*, *428*(6980), 316-319.

Dragoi, V., Sharma, J., Miller, E. K., & Sur, M. (2002). Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nature neuroscience*, *5*(9), 883-892.

Fiser, J., & Biederman, I. (2001). Invariance of long-term visual priming to scale, reflection, translation, and hemisphere. *Vision Research*, *41*(2), 221-234.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex*, *16*(11), 1631-1644.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815-836.

Ghuman, A. S., Bar, M., Dobbins, I. G., & Schnyer, D. M. (2008). The effects of priming on frontal-temporal communication. *Proceedings of the National Academy of Sciences*, *105*(24), 8405-8409.

Gilbert, J. R., Gotts, S. J., Carver, F. W., & Martin, A. (2010). Object repetition leads to local increases in the temporal coordination of neural responses. *Frontiers in human neuroscience*, *4*, 30.

Graf, P., & Ryan, L. (1990). Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 978-992.

Grainger, J., & Jacobs, A. M. (1993). Masked partial-word priming in visual word recognition: Effects of positional letter frequency. *Journal of experimental psychology: human perception and performance*, *19*(5), 951-964.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in cognitive sciences*, *10*(1), 14-23.

Gotts, S. J., Chow, C. C., & Martin, A. (2012). Repetition priming and repetition suppression: Multiple mechanisms in need of testing. *Cognitive neuroscience*, *3*(3-4), 250-259.

Henson, R. N., Eckstein, D., Waszak, F., Frings, C., & Horner, A. J. (2014). Stimulus–response bindings in priming. *Trends in cognitive sciences*, *18*(7), 376-384.

Henson, R. N., Mouchlianitis, E., Matthews, W. J., & Kouider, S. (2008). Electrophysiological correlates of masked face priming. *Neuroimage*, *40*(2), 884-895.

Henson, R. N., Rylands, A., Ross, E., Vuilleumeir, P., & Rugg, M. D. (2004). The effect of repetition lag on electrophysiological and haemodynamic correlates of visual object priming. *Neuroimage*, *21*(4), 1674-1689.

Hillstrom, A. P. (2000). Repetition effects in visual search. *Attention, Perception, & Psychophysics*, *62*(4), 800-817.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*(1-2), 183-216.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in cognitive sciences*, *8*(11), 494-500.

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research PRPF*, *73*(4), 512-526.

Horner, A. J., & Henson, R. N. (2009). Bindings between stimuli and multiple response codes dominate long-lag repetition priming in speeded classification tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 757-779.

Horner, A. J., & Henson, R. N. (2011). Stimulus–response bindings code both abstract and specific representations of stimuli: Evidence from a classification priming design that reverses multiple levels of response representation. *Memory & Cognition*, *39*(8), 1457-1471.

Horner, A. J., & Henson, R. N. (2012). Incongruent abstract stimulus–response bindings result in response interference: fMRI and EEG evidence from visual object classification priming. *Journal of Cognitive Neuroscience*, *24*(3), 760-773.

Huang, L., Holcombe, A. O., & Pashler, H. (2004). Repetition priming in visual search: Episodic retrieval, not feature priming. *Memory & Cognition*, *32*(1), 12-20.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 485-508.

James, T. W., & Gauthier, I. (2006). Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Human brain mapping*, *27*(1), 37-46.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & Dr Davies (eds.), Varieties of attention (pp. 29-61).

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, *24*(2), 175-219.

Kirsner, K., Milech, D., & Stumpfel, V. (1986). Word and picture identification: Is representational parsimony possible? *Memory & Cognition*, *14*(5), 398-408.

Kohn, A., & Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron*, *39*(4), 681-691.

Kunde, W., Kiesel, A., & Hoffmann, J. (2003). Conscious control over the content of unconscious cognition. *Cognition*, *88*(2), 223-242.

Lamy, D., Zivony, A., & Yashar, A. (2011). The role of search difficulty in intertrial feature priming. *Vision research*, *51*(19), 2099-2109.

Lander, K., Bruce, V., Smith, E., & Hancock, P. J. (2009). Multiple repetition priming of faces: Massed and spaced presentations. *Visual Cognition*, *17*(4), 598-616.

Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: neuronal adaptation or stimulus expectation?. *Cerebral Cortex*, *22*(3), 567-576.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological review*, *95*(4), 492-527.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms?. *Cognitive Psychology*, *22*(1), 1-35.

Maljkovic, V., & Martini, P. (2005). Implicit short-term memory and event frequency effects in visual search. *Vision Research*, *45*(21), 2831-2846.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & cognition*, *22*(6), 657-672.

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Attention, Perception, & Psychophysics*, *58*(7), 977-991.

Marsolek, C. J. (2008). What antipriming reveals about priming. *Trends in Cognitive Sciences*, *12*(5), 176-181.

Marsolek, C. J., Deason, R. G., Ketz, N. A., Ramanathan, P., Bernat, E. M., Steele, V. R., ... & Schnyer, D. M. (2010). Identifying objects impairs knowledge of other objects: A relearning explanation for the neural repetition effect. *NeuroImage*, *49*(2), 1919-1932.

Marsolek, C. J., Schnyer, D. M., Deason, R. G., Ritchey, M., & Verfaellie, M. (2006). Visual antipriming: Evidence for ongoing adjustments of superimposed visual object representations. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(3), 163-174.

Masson, M. E. (1986). Identification of typographically transformed words: Instance-based skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 479-488.

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*(6), 1013-1052.

Mesulam, M. (2008). Representation, inference, and transcendent encoding in neurocognitive networks of the human brain. *Annals of neurology*, *64*(4), 367-378.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, *16*(5), 519-533.

Naccache, L., & Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen stimuli. *Cognition*, *80*(3), 215-229.

Neill, W. T. (1997). Episodic retrieval in negative priming and repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1291-1305.

Race, Elizabeth A., Shanti Shanker, and Anthony D. Wagner. "Neural priming in human frontal cortex: multiple forms of learning reduce demands on the prefrontal executive system." *Journal of Cognitive Neuroscience* 21.9 (2009): 1766-1781.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79-87.

Reynvoet, B., Caessens, B., & Brysbaert, M. (2002). Automatic stimulus-response associations may be semantically mediated. *Psychonomic Bulletin & Review*, *9*(1), 107-112.

Roediger, H. L., & Blaxton, T. A. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory & cognition*, *15*(5), 379-388.

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. *Handbook of neuropsychology*, *8*, 63-63.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental psychology: General*, *104*(3), 192-233.

Sackur, J., & Dehaene, S. (2009). The cognitive architecture for chaining of two mental operations. *Cognition*, *111*, 187-211.

Salimpoor, V. N., Chang, C., & Menon, V. (2010). Neural basis of repetition priming during mathematical cognition: repetition suppression or repetition enhancement?. *Journal of cognitive neuroscience, 22*(4), 790-805.

Sciama, S. C., Semenza, C., & Butterworth, B. (1999). Repetition priming in simple addition depends on surface form and typicality. *Memory & cognition, 27*(1), 116-127.

Schacter, D. L. (1990). Perceptual representation systems and implicit memory. *Annals of the New York Academy of Sciences*, *608*(1), 543-571.

Scheepers, C., & Sturt, P. (2014). Bidirectional syntactic priming across cognitive domains: From arithmetic to language and back. *The Quarterly Journal of Experimental Psychology*, *67*(8), 1643-1654.

Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., & Viskupova, I. (2011). Structural priming across cognitive domains: From simple arithmetic to relative-clause attachment. *Psychological Science*, *22*(10), 1319-1326.

Soldan, A., Clarke, B., Colleran, C., & Kuras, Y. (2012). Priming and stimulus–response learning in perceptual classification tasks. *Memory*, *20*(4), 400-413.

Stevens, W. D., Kahn, I., Wig, G. S., & Schacter, D. L. (2011). Hemispheric asymmetry of visual scene processing in the human brain: evidence from repetition priming and intrinsic activity. *Cerebral Cortex*, *22*(8), 1935-1949.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *121*(1), 15–23.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science*, *314*(5803), 1311-1314.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature neuroscience*, *11*(9), 1004-1006.

Summerfield, C., Wyart, V., Johnen, V. M., & De Gardelle, V. (2011). Human scalp electroencephalography reveals that repetition suppression varies with expectation. *Frontiers in Human Neuroscience*, *5*(67), 192-193.

Thomson, D. R., & Milliken, B. (2012). Perceptual distinctiveness produces long-lasting priming of pop-out. *Psychonomic bulletin & review*, *19*(2), 170-176.

Thomson, D. R., & Milliken, B. (2013). Contextual distinctiveness produces long-lasting priming of pop-out. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 202-215.

Treisman, A. (1992). Perceiving and re-perceiving objects. *American Psychologist*, *47*(7), 862-875.

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of experimental psychology: learning, memory, and cognition*, *8*(4), 336-342.

Valt, C., Klein, C., & Boehm, S. G. (2015). Dissociation of rapid response learning and facilitation in perceptual and conceptual networks of person recognition. *British Journal of Psychology*, *106*(3), 375-396.

Ward, E. J., Chun, M. M., & Kuhl, B. A. (2013). Repetition suppression and multi-voxel pattern similarity differentially track implicit and explicit visual memory. *Journal of Neuroscience*, *33*(37), 14749-14757.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current opinion in neurobiology*, *8*(2), 227-233.

Wong, E. K. (2000). Dissociative prime-probe contextual effects on negative priming and repetition priming: A challenge to episodic retrieval as a unified account of negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1411.

Zhang, F., Fairchild, A. J., & Li, X. (2017). Visual antipriming effect: Evidence from Chinese character identification. *Frontiers in Psychology*, *8*, 1791.

Appendix A

Appendix A contains tables of all inferential results from each ANOVA and t-test calculated under the relative-difference method of analysis and reported in the experiment section of the thesis. All t-tests were two-tailed, unless otherwise stated, and compared the change from study to test against zero for each condition.

Experiment 1

Table 1A. *Results of a 2 (task) x 2(feature) x 2(condition) mixed ANOVA calculated on the RT data and reported for Experiment 1 (df=39).*

	F	*p*	*partial* η^2
Intercept	13.70	*0.001*	0.26
Task	5.74	*0.021*	0.13
Feature	2.84	*0.100*	0.07
Condition	1.05	*0.311*	0.03
Task x Feature	0.46	*0.502*	0.01
Task x Condition	0.16	*0.691*	0.00
Feature x Condition	0.10	*0.759*	0.00
Task x Feature x Condition	2.61	*0.114*	0.06

Table 1B. *Results of a 2 (task) x 2(feature) x 2(condition) mixed ANOVA calculated on the error data and reported for Experiment 1 (df=39).*

	F	*p*	*partial* η^2
Intercept	0.22	*0.639*	0.01
Task	0.02	*0.896*	0.00
Feature	0.21	*0.647*	0.01
Condition	0.71	*0.406*	0.02
Task x Feature	2.40	*0.130*	0.06
Task x Condition	1.21	*0.277*	0.03
Feature x Condition	0.73	*0.398*	0.02
Task x Feature x Condition	5.69	*0.022*	0.13

Table 1C. *Results of a 2(feature) x 2(condition) mixed ANOVA calculated on the error data and reported for the word task in Experiment 1 (df=20).*

	F	*p*	*partial* η^2
Intercept	0.37	*0.550*	0.02
Feature	1.01	*0.326*	0.05
Condition	2.97	*0.100*	0.13
Feature x Condition	6.32	*0.021*	0.24

Table 1D. *Results of a 2(feature) x 2(condition) mixed ANOVA calculated on the error data and reported for the math task in Experiment 1 (df=19).*

	F	*p*	*partial* η^2
Intercept	0.04	*0.848*	0.00
Feature	1.38	*0.254*	0.07
Condition	0.02	*0.878*	0.00
Feature x Condition	0.99	*0.333*	0.05

Table 1E. *Mean RT difference (ms) between study and test for each condition in the word task of Experiment 1, and inferential statistics from single sample t-tests (df=20).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	21.33	8.33	2.56	*0.019*	(3.96 , 38.71)
ID-Change	-18.26	6.17	-2.96	*0.008*	(-31.12 , -5.40)
Colour-Same	18.57	5.94	3.12	*0.005*	(6.17 , 30.97)
Colour-Change	32.69	5.68	5.76	*0.000*	(20.85 , 44.53)

Table 1F. *Mean RT difference (ms) between study and test for each condition in the math task of Experiment 1, and inferential statistics from single sample t-tests (df=19).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	30.00	47.32	0.63	*0.534*	(-69.05 , 129.05)
ID-Change	40.53	35.15	1.15	*0.263*	(-33.04 , 114.09)
Colour-Same	126.00	52.59	2.40	*0.027*	(15.92 , 236.08)
Colour-Change	57.48	36.06	1.59	*0.127*	(-18.01 , 132.96)

Table 1G. *Mean error difference (%) between study and test for each condition in the word task of Experiment 1, and inferential statistics from single sample t-tests (df=20).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	1.37	1.18	1.17	*0.257*	(-1.08 , 3.82)
ID-Change	-3.03	1.22	-2.47	*0.022*	(-5.58 , -0.47)
Colour-Same	-0.41	0.77	-0.53	*0.601*	(-2.02 , 1.20)
Colour-Change	0.68	0.92	0.74	*0.470*	(-1.25 , 2.61)

Table 1H. *Mean error difference (%) between study and test for each condition in the math task of Experiment 1, and inferential statistics from single sample t-tests (df=19).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	-0.07	1.27	-0.05	*0.960*	(-2.73 , 2.60)
ID-Change	1.46	1.74	0.83	*0.415*	(-2.20 , 5.11)
Colour-Same	-0.55	1.66	-0.33	*0.744*	(-4.03 , 2.93)
Colour-Change	-1.63	1.64	-0.99	*0.335*	(-5.06 , 1.81)

Experiment 2

Table 2A. *Results of a 2 (task) x 2(feature) x 2(condition) repeated-measures ANOVA calculated on the RT data and reported for Experiment 2 (df=17).*

	F	*p*	*partial* η^2
Intercept	16.70	*0.001*	0.50
Task	6.89	*0.018*	0.29
Feature	1.03	*0.325*	0.06
Condition	9.30	*0.007*	0.35
Task x Feature	0.62	*0.441*	0.04
Task x Condition	1.79	*0.199*	0.10
Feature x Condition	3.61	*0.075*	0.18
Task x Feature x Condition	13.36	*0.002*	0.44

Table 2B. *Results of a 2 (task) x 2(feature) x 2(condition) repeated-measures ANOVA calculated on the error data and reported for Experiment 2 (df=17).*

	F	*p*	*partial* η^2
Intercept	0.34	*0.566*	0.02
Task	0.03	*0.872*	0.00
Feature	2.19	*0.157*	0.11
Condition	0.26	*0.617*	0.02
Task x Feature	5.23	*0.035*	0.24
Task x Condition	0.28	*0.604*	0.02
Feature x Condition	4.99	*0.039*	0.23
Task x Feature x Condition	0.40	*0.538*	0.02

Table 2C. *Results of a 2(feature) x 2(condition) mixed ANOVA calculated on the RT data and reported for the word task in Experiment 2 (df=17).*

	F	*p*	*partial* η^2
Intercept	39.01	*0.000*	0.70
Feature	5.23	*0.035*	0.24
Condition	28.81	*0.000*	0.63
Feature x Condition	25.37	*0.000*	0.60

Table 2D. *Results of a 2(feature) x 2(condition) mixed ANOVA calculated on the RT data and reported for the math task in Experiment 2 (df=17).*

	F	*p*	*partial* η^2
Intercept	11.47	*0.004*	0.40
Feature	0.01	*0.918*	0.00
Condition	5.35	*0.033*	0.24
Feature x Condition	8.27	*0.011*	0.33

Table 2E. *Mean RT difference (ms) between study and test for each condition in the word task of Experiment 2, and inferential statistics from single sample t-tests (df=17).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	32.89	5.14	6.40	*0.000*	(22.05 , 43.73)
ID-Change	-19.92	5.76	-3.46	*0.003*	(-32.08 , -7.76)
Colour-Same	25.89	6.56	3.95	*0.001*	(12.06 , 39.79)
Colour-Change	21.11	5.23	4.04	*0.001*	(10.08 , 32.14)

Table 2F. *Mean RT difference (ms) between study and test for each condition in the math task of Experiment 2, and inferential statistics from single sample t-tests (df=17).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	64.53	30.32	2.13	*0.048*	(0.56 , 128.50)
ID-Change	70.44	23.24	3.03	*0.008*	(21.41 , 119.48)
Colour-Same	129.53	33.14	3.91	*0.001*	(59.61 , 199.44)
Colour-Change	1.56	23.83	0.07	*0.949*	(-48.72 , 51.83)

Table 2G. *Mean error difference (%) between study and test for each condition in the word task of Experiment 2, and inferential statistics from single sample t-tests (df=17).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	0.90	1.09	0.82	*0.422*	(-1.41 , 3.21)
ID-Change	-2.68	1.01	-2.66	*0.017*	(-4.81 , -0.55)
Colour-Same	0.72	0.60	1.20	*0.245*	(-0.54 , 1.99)
Colour-Change	2.26	0.92	2.46	*0.025*	(0.32 , 4.20)

Table 2H. *Mean error difference (%) between study and test for each condition in the math task of Experiment 2, and inferential statistics from single sample t-tests (df=17).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	1.02	1.19	0.86	*0.403*	(-1.49 , 3.52)
ID-Change	-0.61	1.69	-0.36	*0.723*	(-4.18 , 2.97)
Colour-Same	-0.57	1.54	-0.37	*0.717*	(-3.82 , 2.68)
Colour-Change	0.89	1.04	0.86	*0.404*	(-1.30 , 3.08)

Experiment 3

Table 3A. *Results of a 2 (task) x 2(feature) x 2(condition) repeated-measures ANOVA calculated on the RT data and reported for Experiment 3 (df=22).*

	F	*p*	*partial* η^2
Intercept	32.87	*0.000*	0.60
Task	4.51	*0.045*	0.17
Feature	0.52	*0.480*	0.02
Condition	2.19	*0.153*	0.09
Task x Feature	0.40	*0.532*	0.02
Task x Condition	0.40	*0.536*	0.02
Feature x Condition	8.96	*0.007*	0.29
Task x Feature x Condition	1.32	*0.263*	0.06

Table 3B. *Results of a 2 (task) x 2(feature) x 2(condition) repeated-measures ANOVA calculated on the error data and reported for Experiment 3 (df=22).*

	F	*p*	*partial η²*
Intercept	5.70	*0.026*	0.21
Task	17.77	*0.000*	0.45
Feature	1.49	*0.236*	0.06
Condition	0.53	*0.475*	0.02
Task x Feature	0.01	*0.926*	0.00
Task x Condition	2.49	*0.129*	0.10
Feature x Condition	1.97	*0.174*	0.08
Task x Feature x Condition	0.60	*0.446*	0.03

Table 3C. *Mean RT difference (ms) between study and test for each condition in the word task of Experiment 3, and inferential statistics from single sample t-tests (df=22).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	34.54	8.16	4.23	*0.000*	(17.63 , 51.46)
ID-Change	-5.37	7.37	-0.73	*0.474*	(-20.66 , 9.92)
Colour-Same	30.28	8.59	3.52	*0.002*	(12.46 , 48.10)
Colour-Change	39.83	7.77	5.12	*0.000*	(23.71 , 55.95)

Table 3D. *Mean RT difference (ms) between study and test for each condition in the math task of Experiment 3, and inferential statistics from single sample t-tests (df=22).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	93.52	31.32	2.99	*0.007*	(28.58 , 158.47)
ID-Change	3.76	33.24	0.11	*0.911*	(-65.17 , 72.69)
Colour-Same	39.43	22.01	1.79	*0.087*	(-6.22 , 85.09)
Colour-Change	60.76	20.60	2.95	*0.007*	(18.04 , 103.48)

Table 3E. *Mean error difference (%) between study and test for each condition in the word task of Experiment 3, and inferential statistics from single sample t-tests (df=22).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	-0.29	1.01	-0.29	*0.776*	(-2.38 , 1.80)
ID-Change	-1.35	0.75	-1.80	*0.085*	(-2.90 , 0.20)
Colour-Same	0.29	1.07	0.27	*0.791*	(-1.93 , 2.50)
Colour-Change	0.24	1.05	0.23	*0.823*	(-1.95 , 2.43)

Table 3F. *Mean error difference (%) between study and test for each condition in the math task of Experiment 3, and inferential statistics from single sample t-tests (df=22).*

	M	*SE*	*t*	*p*	*95% CI*
ID-Same	1.70	1.24	1.36	*0.187*	(-0.88 , 4.28)
ID-Change	1.81	1.10	1.64	*0.116*	(-0.48 , 4.10)
Colour-Same	0.80	1.40	0.57	*0.573*	(-2.10 , 3.70)
Colour-Change	4.63	1.35	3.43	*0.002*	(1.83 , 7.44)

Experiment 4

Table 4A. *Results of a 2 (task) x 2(condition) repeated-measures ANOVA calculated on the RT data and reported for Experiment 4 (df=25).*

	F	*p*	*partial η^2*
Intercept	46.27	*0.000*	0.65
Task	1.80	*0.192*	0.07
Condition	0.58	*0.454*	0.02
Task x Condition	0.60	*0.445*	0.02

Table 4B. *Results of a 2 (task) x 2(condition) repeated-measures ANOVA calculated on the error data and reported for Experiment 4 (df=25).*

	F	*p*	*partial η^2*
Intercept	3.98	*0.057*	0.14
Task	0.00	*0.992*	0.00
Condition	5.04	*0.034*	0.17
Task x Condition	2.80	*0.107*	0.10

Table 4C. *Mean RT difference (ms) between study and test in each condtion of Experiment 4, and inferential statistics from single sample t-tests (df=25).*

	M	*SE*	*t*	*p*	*95% CI*
Word Hand-Same	54.23	14.57	3.72	*0.001*	(24.23 , 84.23)
Word Hand-Change	33.75	8.18	4.13	*0.000*	(16.91 , 50.59)
Math Hand-Same	74.06	23.76	3.12	*0.005*	(25.11 , 123.00)
Math Hand-Change	72.92	17.20	4.24	*0.000*	(37.49 , 108.35)

Table 4D. *Mean error difference (%) between study and test in each condition of Experiment 4, and inferential statistics from single sample t-tests (df=25).*

	M	*SE*	*t*	*p*	*95% CI*
Word Hand-Same	1.20	0.72	1.66	*0.109*	(-0.29 , 2.70)
Word Hand-Change	0.75	0.63	1.20	*0.241*	(-0.54 , 2.04)
Math Hand-Same	2.46	1.11	2.21	*0.036*	(0.17 , 4.75)
Math Hand-Change	-0.49	0.70	-0.69	*0.494*	(-1.94 , 0.96)

Experiment 5

Table 5A. *Results of a 2 (task) x 2(condition) repeated-measures ANOVA calculated on the RT data and reported for Experiment 5 (df=21).*

	F	*p*	*partial* η^2
Intercept	28.46	*0.000*	0.58
Task	4.09	*0.056*	0.16
Condition	1.06	*0.316*	0.05
Task x Condition	0.61	*0.442*	0.03

Table 5B. *Results of a 2 (task) x 2(condition) repeated-measures ANOVA calculated on the error data and reported for Experiment 5 (df=21).*

	F	*p*	*partial* η^2
Intercept	4.96	*0.037*	0.19
Task	0.01	*0.943*	0.00
Condition	0.81	*0.379*	0.04
Task x Condition	1.10	*0.306*	0.05

Table 5C. *Mean RT difference (ms) between study and test in each condtion of Experiment 5, and inferential statistics from single sample t-tests (df=21).*

	M	*SE*	*t*	*p*	*95% CI*
Word Hand-Same	39.39	9.11	4.33	*0.000*	(20.45 , 58.32)
Word Hand-Change	35.75	5.34	6.70	*0.000*	(24.65 , 46.85)
Math Hand-Same	85.93	21.52	3.99	*0.001*	(41.18 , 130.69)
Math Hand-Change	54.98	26.22	2.10	*0.048*	(0.45 , 109.50)

Table 5D. *Mean error difference (%) between study and test in each condition of Experiment 5, and inferential statistics from single sample t-tests (df=21).*

	M	*SE*	*t*	*p*	*95% CI*
Word Hand-Same	0.85	0.57	1.50	*0.150*	(-0.33 , 2.04)
Word Hand-Change	0.93	0.94	0.99	*0.335*	(-1.03 , 2.88)
Math Hand-Same	1.84	1.10	1.67	*0.110*	(-0.45 , 4.13)
Math Hand-Change	0.10	1.02	0.09	*0.926*	(-2.02 , 2.21)

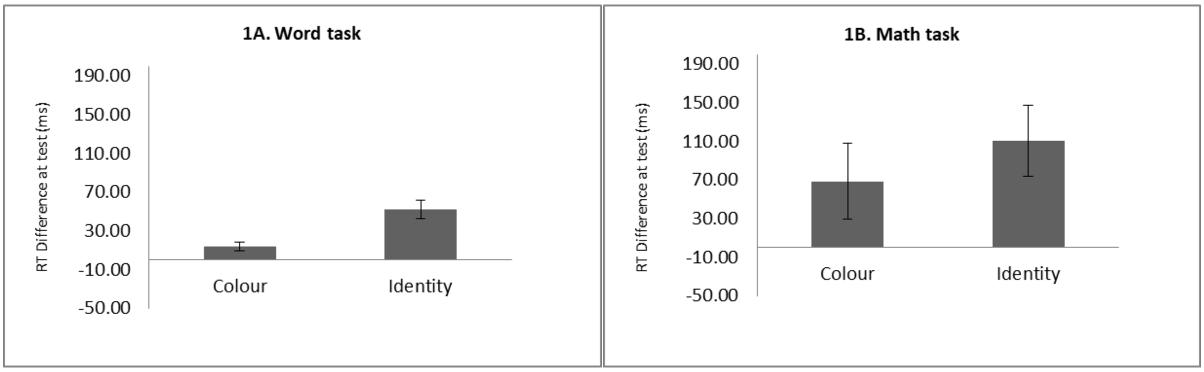
Appendix B

The results reported in Appendix B have been calculated based on the absolute-difference method of analysis. For each participant, the median RT and error rate was calculated in the test blocks of each condition, and these provided the basis for a comparison between conditions. All data exclusions and analytical issues raised under the relative-difference method apply to these analyses. The RT and error rate differences at test (Change – Same) and related paired t-test results are also provided here for completeness.

Experiment 1

Figures 1A and 1B show the mean difference between the median RTs for the test blocks in each session for the word and the math task, respectively. Table 1A shows the results of a 2 (task: word vs. math) x 2 (feature: ID vs. colour) x 2 (condition: same vs. change) mixed ANOVA conducted on the RT data (i.e., the median RT for test blocks).

There was a main effect of task, $F(1, 39) = 52.83$, $p < .001$, partial $\eta^2 = .58$, indicating longer RTs at test in the math task (1309 ms) than the word task (584 ms), and a main effect of condition, $F(1, 39) = 14.59$, $p = .001$, partial $\eta^2 = .27$, indicating shorter RTs at test for repetition (908 ms) than for change (968 ms). No other effects reached significance.



Figures 1A & 1B. RT difference in the test blocks (change-same) in the word and math classification tasks as a function of feature in Experiment 1. Error bars show +/- 1 standard deviation of the mean.

Table 1A. *Results of a 2 (task) x 2(feature) x 2(condition) mixed ANOVA calculated on the RT data and reported for Experiment 1 (df=39).*

	F	*p*	*partial* η^2
Task	52.83	*0.000*	0.58
Feature	0.00	*1.000*	0.00
Condition	14.59	*0.000*	0.27
Task x Feature	0.18	*0.678*	0.00
Task x Condition	3.15	*0.084*	0.07
Feature x Condition	3.81	*0.058*	0.09
Task x Feature x Condition	0.01	*0.928*	0.00

The mean RT difference at test (Change - Same) and related t-test results are provided in Table 1B. In the word task, there was a statistically significant difference between presenting studied words and presenting novel words at test, an indicator of RP under the traditional method of analysis. There was also a significant difference between presenting the studied words at test in the same colour or in a novel colour, an effect of manipulating a task irrelevant perceptual feature that is suggestive of feature binding and retrieval. In the math task, the traditional RP effect was present, as shown by faster RTs for studied equations than for novel equations, and there was no effect of manipulating the colour of repeated equations.

Table 1B. *Mean RT difference (ms) between test blocks (Change - Same) in each session of the word (df=20) and math (df=19) tasks from Experiment 1, and inferential statistics from paired t-tests.*

	M	*SE*	*t*	*p*	*95% CI*
Word - ID	51.88	9.37	5.54	*0.000*	(32.33 , 71.43)
Word - Colour	13.52	4.60	2.94	*0.008*	(3.94 , 23.11)
Math - ID	110.53	37.09	2.98	*0.008*	(32.90 , 188.15)
Math - Colour	68.40	39.34	1.74	*0.098*	(-13.94 , 150.74)

A similar ANOVA was conducted on the error rates and results are provided in Table 1C. There was a main effect of task, $F(1, 39) = 7.39$, $p = .010$, partial $\eta^2 = .16$, indicating higher error rates in the test blocks of the math task (8.9%) than in the test blocks of the word task (5.3%), and a main effect of condition, $F(1, 39) = 19.35$, $p < .001$, partial $\eta^2 = .33$, indicating lower error rates for repetition (5.9%) than for change (8.2%) at test. In addition, task and feature interacted, $F(1, 39) = 4.92$, $p = .033$, partial $\eta^2 = .11$. In the word task, error rates were higher in the test blocks of the ID session (5.7%) than in the test blocks of the colour session (4.9%). In the math task, error rates were higher in the test blocks of the colour session (10.3%) than in the test blocks of the ID session (7.6%). No other effects reached significance.

Table 1C. *Results of a 2 (task) x 2(feature) x 2(condition) mixed ANOVA calculated on the error data and reported for Experiment 1 (df=39).*

	F	*p*	*partial* η^2
Task	7.39	*0.010*	0.16
Feature	1.41	*0.243*	0.03
Condition	19.35	*0.000*	0.33
Task x Feature	4.92	*0.033*	0.11
Task x Condition	1.20	*0.280*	0.03
Feature x Condition	1.34	*0.254*	0.03
Task x Feature x Condition	1.11	*0.298*	0.03

The mean error difference (Change - Same) and related t-test results are provided in Table 1D. In the word task, the traditional RP effect was present as shown by more accurate performance when presenting studied words than when presenting novel words at test, but there was no effect of manipulating the colour of repeated words at test. In the math task, there were no effects on accuracy from manipulating identity or colour at test.

Table 1D. *Mean error difference (%) between test blocks (Change - Same) in each session of the word (df=20) and math (df=19) tasks from Experiment 1, and inferential statistics from paired t-tests.*

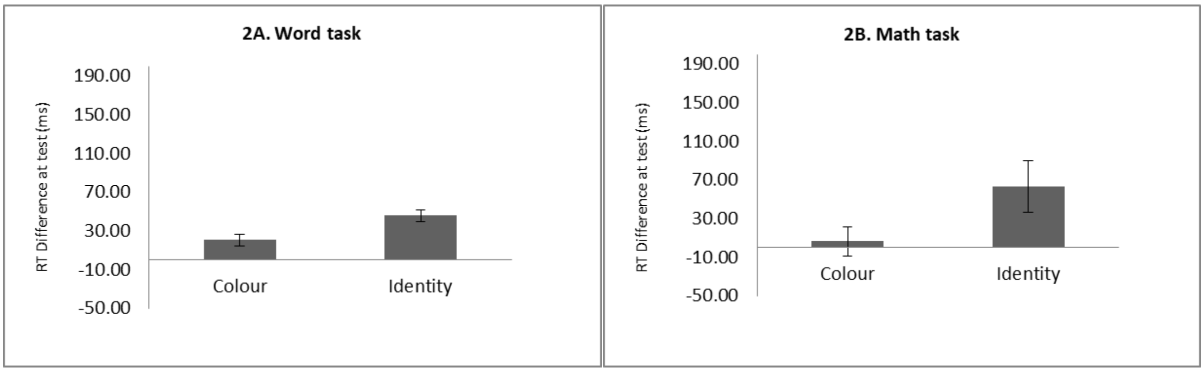
	M	*SE*	*t*	*p*	*95% CI*
Word - ID	4.22	0.84	5.00	*0.000*	(2.46 , 5.98)
Word - Colour	1.40	0.95	1.48	*0.150*	(-0.58 , 3.38)
Math - ID	1.75	1.17	1.50	*0.150*	(-0.70 , 4.21)
Math - Colour	1.63	1.56	1.04	*0.310*	(-1.65 , 4.90)

Experiment 2

Figures 2A and 2B show the mean difference between the median RTs for the test blocks in each session for the word and the math task, respectively. Table 2A shows the results of a 2 (task: word vs. math) x 2 (feature: ID vs. colour) x 2 (condition: same vs. change) repeated-measures ANOVA conducted on the RT data (i.e., the median RT for the test blocks).

There was a main effect of task, $F(1, 17) = 122.05$, $p < .001$, partial $\eta^2 = .88$, indicating longer RTs in the math task (915 ms) than the word task (547 ms), and a main effect of condition, $F(1, 17) = 12.61$, $p = .002$, partial $\eta^2 = .43$, indicating shorter RTs for repetition (714 ms) than for change (748 ms) at test. In addition, feature and condition interacted, $F(1, 17) = 6.48$, $p = .021$, partial $\eta^2 = .28$. In the ID session, RTs were shorter in test blocks in the ID-Same condition (707 ms) than in the ID-Change condition (762 ms), indicating significant RP. In the colour session there was no difference at test between the

Colour-Same (721 ms) and Colour-Change conditions (734 ms), indicating colour had no effect on subsequent processing. No other effects reached significance.



Figures 2A & 2B. RT difference between the test blocks (change-same) in the word and math classification tasks as a function of feature in Experiment 2. Error bars show +/- 1 standard deviation of the mean.

Table 2A. *Results of a 2 (task) x 2(feature) x 2(condition) repeated-measures ANOVA calculated on the RT data and reported for Experiment 2 (df=17).*

	F	*p*	*partial* η^2
Task	122.05	*0.000*	0.88
Feature	0.23	*0.636*	0.01
Condition	12.61	*0.002*	0.43
Task x Feature	1.11	*0.307*	0.06
Task x Condition	0.01	*0.917*	0.00
Feature x Condition	6.48	*0.021*	0.28
Task x Feature x Condition	1.18	*0.293*	0.06

The mean RT difference at test (Change - Same) and related t-test results are provided in Table 2B. In the word task there was again a significant difference between presenting studied words and presenting novel words at test, an indicator of RP under the traditional method of analysis. In addition there was a difference between presenting studied words in the same colour or in a novel colour, indicating feature binding between colour and the identity of words. In the math task the traditional RP effect was present as shown by faster

RTs for studied equations than for novel equations, and there was no effect of manipulating the colour of studied equations, indicating no feature binding.

Table 2B. *Mean RT difference (ms) between test blocks (Change - Same) in each session of the word and math tasks from Experiment 2, and inferential statistics from paired t-tests (df=17).*

	M	*SE*	*t*	*p*	*95% CI*
Word - ID	45.86	6.29	7.30	*0.000*	(32.60 , 59.12)
Word - Colour	20.61	5.96	3.46	*0.003*	(8.05 , 33.18)
Math - ID	63.00	26.75	2.35	*0.031*	(6.56 , 119.44)
Math - Colour	6.36	15.15	0.42	*0.680*	(-25.60 , 38.32)

A similar ANOVA was conducted on the error rates and results are reported in Table 2C. There was a main effect of task, $F(1, 17) = 24.85$, $p < .001$, partial $\eta^2 = .59$, indicating higher error rates in the test blocks of the math task (6.2%) than in the test blocks of the word task (2.8%). No other effects reached significance.

Table 2C. *Results of a 2 (task) x 2(feature) x 2(condition) repeated-measures ANOVA calculated on the error data and reported for Experiment 2 (df=17).*

	F	*p*	*partial η^2*
Task	24.85	*0.000*	0.59
Feature	3.37	*0.084*	0.17
Condition	2.21	*0.156*	0.11
Task x Feature	2.26	*0.151*	0.12
Task x Condition	2.19	*0.158*	0.11
Feature x Condition	3.73	*0.070*	0.18
Task x Feature x Condition	0.67	*0.425*	0.04

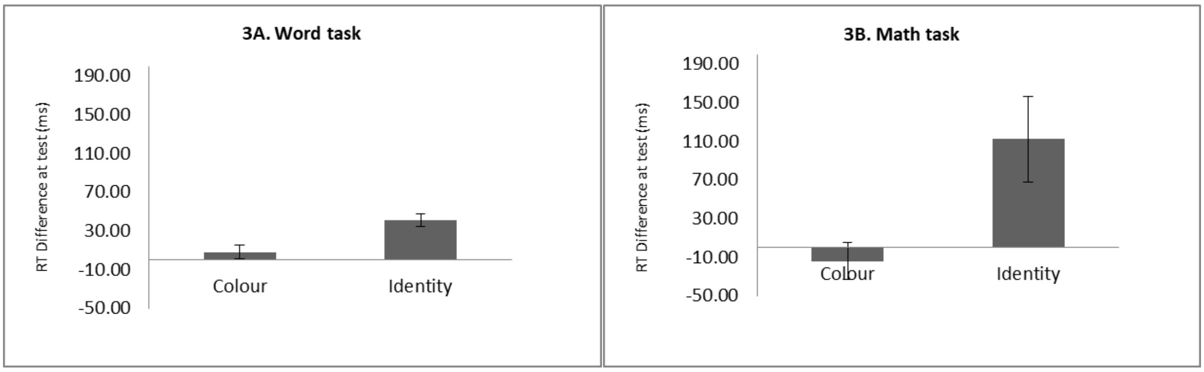
The mean error difference (Change - Same) and related t-test results are shown in Table 2D. In the word the traditional RP effect was present as shown by more accurate performance when presenting studied words than when presenting novel words at test but there was no effect of manipulating colour. In the math task there were no effects on accuracy of manipulating identity or colour.

Table 2D. *Mean error difference (%) between test blocks (Change - Same) in each session of the word and math tasks from Experiment 2, and inferential statistics from paired t-tests (df=17).*

	M	*SE*	*t*	*p*	*95% CI*
Word - ID	3.13	0.96	3.28	*0.004*	(1.12 , 5.15)
Word - Colour	0.86	1.24	0.69	*0.500*	(-1.76 , 3.47)
Math - ID	1.81	1.88	0.96	*0.351*	(-2.17 , 5.78)
Math - Colour	-2.29	1.28	-1.79	*0.091*	(-4.99 , 0.41)

Experiment 3

Figures 3A and 3B show the mean difference between the median RTs for the test blocks in each session for the word and the math task, respectively. Table 3A reports the results of a 2 (task: word vs. math) x 2 (feature: ID vs. colour) x 2 (condition: same vs. change) repeated-measures ANOVA conducted on the RT data (i.e., the median RT for the test blocks).



Figures 3A & 3B. RT difference between the test blocks (change-same) in the word and math classification tasks as a function of feature in Experiment 3. Error bars show +/- 1 standard deviation of the mean.

There was a main effect of task, $F(1, 22) = 143.37$, $p < .001$, partial $\eta^2 = .87$, indicating longer RTs at test in the math task (1119 ms) than the word task (596 ms), a main effect of feature, $F(1, 22) = 6.48$, $p = .018$, partial $\eta^2 = .23$, indicating faster RTs at test in the colour session (838 ms) than in the ID session (877 ms), and a main effect of condition, $F(1, 22) = 7.61$, $p = .011$, partial $\eta^2 = .26$, indicating faster RTs at test for repetition (839 ms) than for change (876 ms). In addition, feature and condition interacted, $F(1, 22) = 9.31$, $p = .006$, partial $\eta^2 = .30$. For the ID session, RTs were shorter at test in the ID-Same condition (839 ms) than in the ID-Change condition (915 ms), an indication of RP. For the colour session, there was no difference between the Colour-Same (840 ms) and the Colour-Change condition (836 ms), indicating no effect on priming from manipulating the colour of repeated stimuli in test blocks. There was also a 3-way interaction of task, feature, and condition, $F(1, 22) = 4.58$, $p = .044$, partial $\eta^2 = .17$. To clarify the interaction, two separate ANOVAs were conducted, one for the word task and the other for the math task.

Table 3A. *Results of a 2 (task) x 2(feature) x 2(condition) repeated-measures ANOVA calculated on the RT data and reported for Experiment 3 (df=22).*

	F	*p*	*partial* η^2
Task	143.37	*0.000*	0.87
Feature	6.48	*0.018*	0.23
Condition	7.61	*0.011*	0.26
Task x Feature	1.56	*0.225*	0.07
Task x Condition	1.06	*0.315*	0.05
Feature x Condition	9.31	*0.006*	0.30
Task x Feature x Condition	4.58	*0.044*	0.17

The results for the word task are reported in Table 3B. There was a significant main effect of feature, $F(1, 22) = 4.64$, $p = .043$, partial $\eta^2 = .17$, indicating RTs were significantly faster in the colour session (588 ms) than in the ID session (605 ms), and a main effect of condition, $F(1, 22) = 38.31$, $p < .001$, partial $\eta^2 = .64$, indicating faster RTs when features were repeated at test (584 ms) than when they changed (608 ms). There was also a significant 2-way interaction of feature and condition, $F(1, 22) = 8.22$, $p = .009$, partial $\eta^2 = .27$. In the ID session, presenting studied words at test led to faster RTs (584 ms) than presenting novel words (625 ms), indicating RP. In the colour session, RT was comparable for studied words regardless of whether colour was repeated (584 ms) or changed (592 ms), indicating no functional binding between the identity and colour of words.

Table 3B. *Results of a 2(feature) x 2(condition) mixed ANOVA calculated on the RT data and reported for the word task in Experiment 3 (df=22).*

	F	*p*	*partial η^2*
Feature	4.64	*0.043*	0.17
Condition	38.31	*0.000*	0.64
Feature x Condition	8.22	*0.009*	0.27

The results for the math task are reported in Table 3C. There was a statistically significant 2-way interaction of feature and condition, $F(1, 22) = 7.31$, $p = .013$, partial $\eta^2 = .25$. In the ID session, presenting studied equations at test led to faster RTs (1093 ms) than presenting novel equations (1205 ms), indicating RP. In the colour session, RT was comparable for studied equations regardless of whether the colour was repeated (1095 ms) or changed (1081 ms), indicating no feature binding. No other effects reached significance.

Table 3C. *Results of a 2(feature) x 2(condition) mixed ANOVA calculated on the RT data and reported for the math task in Experiment 3 (df=22).*

	F	*p*	*partial η^2*
Feature	3.61	*0.070*	0.14
Condition	3.85	*0.063*	0.15
Feature x Condition	7.31	*0.013*	0.25

The mean RT difference at test (Change - Same) and related t-test results for the word and math tasks are shown in Table 3D. In both the word and the math task there was a significant difference between presenting studied words/equations and presenting novel words/equations at test, an indicator of RP. There was no effect in either task of manipulating the colour of repeated stimuli, indicating no feature binding.

Table 3D. *Mean RT difference (ms) between test blocks (Change - Same) in each session of the word and math tasks from Experiment 3, and inferential statistics from paired t-tests (df=22).*

	M	*SE*	*t*	*p*	*95% CI*
Word - ID	40.80	6.72	6.07	*0.000*	(26.86 , 54.75)
Word - Colour	8.09	7.15	1.13	*0.270*	(-6.74 , 22.91)
Math - ID	112.20	44.36	2.53	*0.019*	(20.20 , 204.19)
Math - Colour	-14.22	19.32	-0.74	*0.470*	(-54.29 , 25.86)

The same 2 (task: word vs. math) x 2 (feature: ID vs. colour) x 2 (condition: same vs. change) repeated-measures ANOVA was conducted on the error rates and the results are reported in Table 3E. There was a main effect of task, $F(1, 22) = 11.83$, $p = .002$, partial $\eta^2 = .35$, indicating higher error rates at test in the math task (6.8%) than the word task (4.2%). In addition, task and condition interacted, $F(1, 22) = 4.66$, $p = .042$, partial $\eta^2 = .17$. In the word task, error rates were lower when features remained the same at test (3.3%) than when they changed (5.0%), indicating an advantage for repetition as per the traditional notion of RP. In the math task, there was no significant difference in error rates when features changed at test (6.3%) compared with when they remained the same (7.2%). No other effects reached significance.

Table 3E. *Results of a 2 (task) x 2(feature) x 2(condition) repeated-measures ANOVA calculated on the error data and reported for Experiment 3 (df=22).*

	F	*p*	*partial* η^2
Task	11.83	*0.002*	0.35
Feature	0.97	*0.335*	0.04
Condition	0.21	*0.650*	0.01
Task x Feature	0.58	*0.454*	0.03
Task x Condition	4.66	*0.042*	0.17
Feature x Condition	3.47	*0.076*	0.14
Task x Feature x Condition	2.24	*0.148*	0.09

The mean error difference (Change - Same) and related t-test results for the word and math tasks are shown in Table 3F. In the word task, there was no difference in accuracy when manipulating identity or colour. In the math task there was no effect of manipulating identity but there was an accuracy benefit when changing colour at test, the opposite effect that would be expected from feature binding.

Table 3F. *Mean error difference (%) between test blocks (Change - Same) in each session of the word and math tasks from Experiment 3, and inferential statistics from paired t-tests (df=22).*

	M	*SE*	*t*	*p*	*95% CI*
Word - ID	1.60	1.41	1.13	*0.270*	(-1.33 , 4.53)
Word - Colour	1.65	1.12	1.48	*0.153*	(-0.66 , 3.96)
Math - ID	0.97	1.12	0.87	*0.396*	(-1.35 , 3.29)
Math - Colour	-2.90	1.33	-2.19	*0.040*	(-5.65 , -0.15)

Experiment 4

Figure 4 shows the mean difference between the median RTs for the test blocks in the word and the math task. Table 4A provides the results of a 2 (task: word vs. math) x 2 (condition: same vs. change) repeated-measures ANOVA conducted on the RT data (i.e., the median RT for test blocks).

There was a main effect of task, $F(1, 25) = 82.16$, $p < .001$, partial $\eta^2 = .77$, indicating slower RTs in the math task (1188 ms) than the word task (606 ms), and a main effect of condition, $F(1, 25) = 7.14$, $p = .013$, partial $\eta^2 = .22$, indicating faster RTs when using the same hand at test (881 ms) compared with using the other hand (913 ms). In addition, task and condition interacted, $F(1, 25) = 6.83$, $p = .015$, partial $\eta^2 = .21$. In the word task, there was no difference in RT regardless of whether the response hand was the same (604 ms) or changed (608 ms) when words were repeated in the test block. In the math task, performance was faster when the same hand was used (1158 ms) than when the response hand changed (1218 ms) at test, an indication of S-R binding. No other effects reached significance.

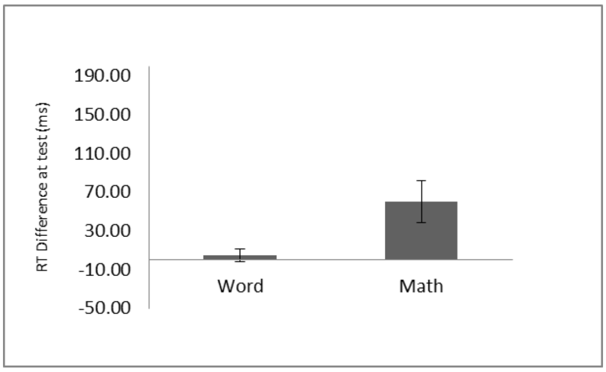


Figure 4. RT difference between test blocks (hand-change – hand-same) in the word and math tasks in Experiment4. Error bars show +/- 1 standard deviation of the mean.

Table 4A. *Results of a 2 (task) x 2(condition) repeated-measures ANOVA calculated on the RT data and reported for Experiment 4 (df=25).*

	F	*p*	partial η^2
Task	82.16	*0.000*	0.77
Condition	7.14	*0.013*	0.22
Task x Condition	6.83	*0.015*	0.21

A similar ANOVA was conducted on the error rates and results are reported in Table 4B. There was a main effect of task, $F(1, 25) = 19.05$, $p < .001$, partial $\eta^2 = .43$, due to higher error rates in the math task (10.6%) than the word task (4.5%), and a main effect of condition,

$F(1, 25) = 7.51$, $p = .011$, partial $\eta^2 = .23$, indicating a lower error rate when the same hand was used (6.9%) than when the hand was changed (8.2%). No other effects reached significance.

Table 4B. *Results of a 2 (task) x 2(condition) repeated-measures ANOVA calculated on the error data and reported for Experiment 4 (df=25).*

	F	*p*	*partial* η^2
Task	19.05	*0.000*	0.43
Condition	7.51	*0.011*	0.23
Task x Condition	0.31	*0.581*	0.01

The mean RT and error rate differences at test (Change - Same) and related t-test results are provided in Table 4C. Evidence of S-R binding and retrieval was absent from the RT and accuracy data in the word task. However evidence of S-R binding was present in the RT data of the math task.

Table 4C. *Mean RT (ms) and error (%) difference between test blocks the word and math tasks from Experiment 4, and inferential statistics from paired t-tests (df=25).*

	M	*SE*	*t*	*p*	*95% CI*
RT - Word	4.40	6.52	0.68	*0.505*	(-9.02 , 17.82)
RT - Math	59.96	21.76	2.76	*0.011*	(15.14 , 104.78)
Error - Word	1.01	0.69	1.46	*0.158*	(-0.42 , 2.44)
Error - Math	1.69	0.86	1.97	*0.060*	(-0.08 , 3.46)

Experiment 5

Figure 5 shows the mean difference between the median RTs for the test blocks in the word and the math task. A 2 (task: word vs. math) x 2 (condition: same vs. change) repeated-measures ANOVA was conducted on the RT (see Table 5A) and accuracy (see Table 5B) data.

In the RT data there was a main effect of task, $F(1, 21) = 114.51$, $p < .001$, partial $\eta^2 = .85$, indicating longer RTs in the test blocks of the math task (1159 ms) than the word task (590 ms). Importantly, there was no main effect of condition and there was no interaction between task and condition. There were no significant effects in the accuracy data.

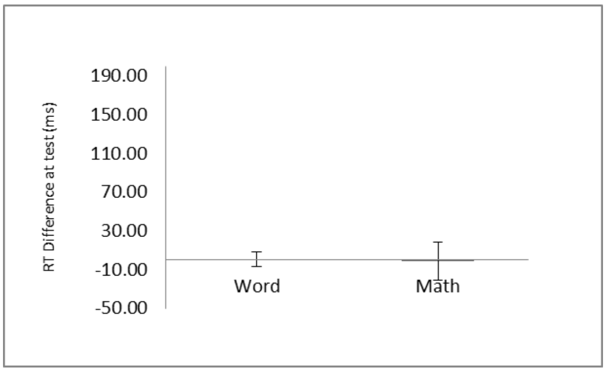


Figure 5. RT difference between test blocks (hand-change – hand-same) in the word and math tasks in Experiment5. Error bars show +/- 1 standard deviation of the mean.

Table 5A. *Results of a 2 (task) x 2(condition) repeated-measures ANOVA calculated on the RT data and reported for Experiment 5 (df=21).*

	F	*p*	partial η^2
Task	114.51	*0.000*	0.85
Condition	0.00	*0.970*	0.00
Task x Condition	0.01	*0.936*	0.00

Table 5B. *Results of a 2 (task) x 2(condition) repeated-measures ANOVA calculated on the error data and reported for Experiment 5 (df=21).*

	F	*p*	partial η^2
Task	0.01	*0.943*	0.00
Condition	0.81	*0.379*	0.04
Task x Condition	1.10	*0.306*	0.05

The mean RT and error rate differences at test (Change - Same) and related t-test results are provided in Table 5C. There is no evidence in either task or in either dependent measure that functional binding took place between stimulus identity and motor response.

Table 5C. *Mean RT (ms) and error (%) difference between test blocks the word and math tasks from Experiment 5, and inferential statistics from paired t-tests (df=21).*

	M	*SE*	*t*	*p*	*95% CI*
RT - Word	0.45	7.54	0.06	*0.952*	(-15.22 , 16.13)
RT - Math	-1.25	19.64	-0.06	*0.950*	(-42.10 , 39.60)
Error - Word	0.07	1.11	0.07	*0.948*	(-2.24 , 2.38)
Error - Math	-1.75	1.42	-1.23	*0.231*	(-4.69 , 1.20)