

2023

Single-System and Dual-Process Accounts of Explicit and Implicit Memory

Spanton, Rory Wilfrid

<https://pearl.plymouth.ac.uk/handle/10026.1/21760>

<http://dx.doi.org/10.24382/5112>

University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

Copyright Statement

Copyright and Moral rights arising from original work in this thesis and (where relevant), any accompanying data, rests with the Author unless stated otherwise¹.

Re-use of the work is allowed under fair dealing exceptions outlined in the Copyright, Designs and Patents Act 1988 (amended)², and the terms of the copyright licence assigned to the thesis by the Author.

In practice, and unless the copyright licence assigned by the author allows for more permissive use, this means,

- that any content or accompanying data cannot be extensively quoted, reproduced or changed without the written permission of the author / rights holder; and
- that the work in whole or part may not be sold commercially in any format or medium without the written permission of the author/rights holder.

Any third-party copyright material in this thesis remains the property of the original owner. Such third party copyright work included in the thesis will be clearly marked and attributed, and the original licence under which it was released will be specified. This material is not covered by the licence or terms assigned to the wider thesis and must be used in accordance with the original licence; or separate permission must be sought from the copyright holder.

The author assigns certain rights to the University of Plymouth including the right to make the thesis accessible and discoverable via the British Library's Electronic Thesis Online Service (EThOS) and the University research repository, and to undertake activities to migrate, preserve and maintain the medium, format and integrity of the deposited file for future discovery and use.

¹E.g. in the example of third party copyright materials reused in the thesis.

²In accordance with best practice principles such as, Marking/Creators/Marking third party content (2013). Available from: https://wiki.creativecommons.org/wiki/Marking/Creators/Marking_third_party_content [accessed 28th November 2023]



UNIVERSITY OF PLYMOUTH

Single-System and Dual-Process Accounts of Explicit and Implicit Memory

by

Rory Wilfrid Spanton

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

November 29, 2023

Acknowledgements

I would first like to thank Chris Berry for his outstanding supervision during my studies. I have worked with Chris since I was an undergraduate and I owe much of my formative development as a researcher to his mentorship. If it were not for his guidance and encouragement, I may not have even started my PhD, let alone finished it. I am also grateful for the input and support of my second and third supervisors, Tim Hollins and Michael Verde.

Many research apprentices and project students assisted me in my work. Joyita Ashton-Simon, Kyle Honeyman, and Samip Pokharel enabled data collection for Experiment 6 as part of their BSc dissertation project. Meghan Whitehouse and Chloe Stevens assisted with Experiments 4, 8 and 9, and Benjamin Bolt, Gemma Doyle, and Jemima Ajagbe assisted with Experiment 7 as research apprentices. All of these students were of great help to me and have my thanks.

My research would not have been possible without studentship funding from the School of Psychology, University of Plymouth, for which I am very grateful. I would also like to thank the academic community within the School of Psychology. I have met too many wonderful PhD students, postdocs, and staff during my time at Plymouth to list individually, but they have all made the past few years truly enjoyable. I have treasured every lunchtime chat, journal club, and social event we have shared, and I feel fortunate to have been part of such a great community.

Finally, I thank my parents for their encouragement and support, Monty for his playful companionship, and Annabelle for her love and unwavering belief in me. This thesis is dedicated to them.

Author's declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award. Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment. A subset of the data from Experiment 6 of the present thesis was used in BSc (Hons) Psychology dissertation projects at the University of Plymouth. The data reported in Experiment 6 were not subject to the same exclusions and were analysed using different statistical methods.

Relevant scientific seminars and conferences were regularly attended at which work was often presented:

Publications:

Experiments 5-9 were reported in:

Spanton, R. W., & Berry, C. J. (2023). Is implicit memory better for correctly recalled items? Evidence from cued and free recall tasks. Manuscript under review.

Experiment 3 was reported in:

Spanton, R. W., & Berry, C. J. (2022). Does variability in recognition memory scale with mean memory strength or encoding variability in the UVSD model? *Quarterly Journal of Experimental Psychology*, 17470218221136498. <https://doi.org/10.1177/17470218221136498>

Presentations and conferences attended:

Spanton, R. W., & Berry, C. J. (2023). *Is implicit memory better for correctly recalled items? Evidence from cued and free recall tasks* [Conference presentation]. Experimental Psychology Society Plymouth Meeting, University of Plymouth, UK.

Spanton, R. W., & Berry, C. J. (2022). *Does variability in recognition memory scale with mean memory strength or encoding variability in the UVSD model?* [Conference presentation]. MathPsych 2022, Toronto, Canada.

Spanton, R. W., & Berry, C. J. (2022). *Does variability in recognition memory scale with mean memory strength or encoding variability in the UVSD model?* [Conference presentation]. Experimental Psychology Society Stirling Meeting, University of Stirling, UK.

Spanton, R. W. (2021). *Evaluating single-system and dual-process models of recognition memory and repetition priming* [Conference presentation]. 13th Annual School of Psychology Conference, University of Plymouth, UK.

Word count for the main body of this thesis: **41089**

A handwritten signature in black ink, appearing to read "Ron S. Smith", with a horizontal line underneath.

Signed:

Date: November 29, 2023

Rory Wilfrid Spanton — Testing Single-System and Dual-Process Accounts of Explicit and Implicit Memory

While some expressions of memory are accompanied by conscious awareness, others do not elicit this awareness yet still impact task performance. Theorists have long questioned whether these *explicit* and *implicit* forms of memory are governed by separate cognitive systems or a single system. There is much behavioural evidence for multiple-systems and dual-process accounts, mainly focusing on studies of recognition memory and long-term repetition priming. However, experimental results and mathematical modelling have shown the ability of a single-system account to explain the relationship between various forms of explicit and implicit memory. This thesis uses behavioural experiments and mathematical modelling to further investigate the predictions of single-system and dual-process accounts of explicit and implicit memory. Chapter 2 investigates the effect of response speeding in recognition and priming tasks, identifying model-based predictions prompted by this manipulation. Chapter 3 examines the effects of encoding variability on recognition memory, before extending this manipulation to priming to again investigate opposing predictions from single-system and dual-process models. Chapter 4 takes a different approach and investigates the relationship between cued recall and implicit memory in behavioural experiments, with and without the inclusion of a recognition task. Finally, Chapter 5 examines the relationship between free recall and implicit memory in three further experiments. The results of this thesis show that while models of recognition and priming make opposing predictions about the relationship between explicit and implicit memory, these predictions are often hard to test in practice. However, Chapter 4 confirms the relationship between cued recall and priming. Chapter 5 provides evidence

for a similar relationship between free recall and priming. Both of these results align with a single-system view over a strict dual-process (and strict multiple-systems) account. With further experimentation, these results may inform future model development and the understanding of the fundamental relationships between explicit and implicit memory.

Contents

Acknowledgements	i
Author's declaration	iii
Abstract	v
Table of Contents	vii
List of Figures	xi
List of Tables	xiii
1 Introduction	1
1.1 Functional and Neuropsychological Dissociations	3
1.2 Neuroimaging	6
1.3 Fluency	7
1.4 Priming in the Absence of Recognition	8
1.5 Signal Detection Theory Models of Explicit and Implicit Memory	9
1.5.1 A Single-System Model of Recognition and Priming	11
1.5.2 Multiple Systems Models of Recognition and Priming	12
1.5.3 Evidence for Single and Multiple Systems Models	13
1.5.4 Signal Detection Models of Recognition, Priming, and Source Memory .	17
1.6 Outstanding Theoretical Questions Regarding the Memory Systems Debate . . .	18
2 Testing Single-System and Dual-Process Models With Response Speeding Manipulations	20
2.1 Expected Value Simulations	21
2.2 Manipulations of Recollection	23
2.3 Experiment 1	25
2.3.1 Method	25
2.3.2 Results	27
2.3.3 Discussion	32
2.4 Experiment 2	33

2.4.1	Method	34
2.4.2	Results	35
2.4.3	Discussion	38
2.5	General Discussion	39
3	Discriminating SS and DPSD1 Models Using Encoding Variability Manipulations	42
3.1	Experiment 3	46
3.1.1	Method	49
3.1.2	Results	51
3.1.3	Discussion	55
3.2	Experiment 4	56
3.2.1	Expected Value Simulations	56
3.2.2	Method	60
3.2.3	Results	62
3.2.4	Discussion	68
3.3	General Discussion	69
4	Continuity Between Cued Recall and Priming	73
4.1	Experiment 5	75
4.1.1	Method	76
4.1.2	Results and Discussion	79
4.2	Experiment 6	82
4.2.1	Method	83
4.2.2	Results	84
4.2.3	Discussion	84
4.3	General Discussion	85
5	Free Recall and Priming	88
5.1	Experiment 7	89
5.1.1	Method	90
5.1.2	Results	93
5.1.3	Discussion	94
5.2	Experiment 8	96
5.2.1	Method	96
5.2.2	Results and Discussion	98
5.3	Experiment 9	99
5.3.1	Method	100
5.3.2	Results and Discussion	101

5.4	General Discussion	102
6	Discussion	106
6.1	Single-System and Dual-Process Models of Recognition Memory and Priming . .	106
6.2	Continuity Between Recall and Implicit Memory	110
6.3	Considerations For Modelling Recall and Implicit Memory	112
6.4	Alternative Classifications of Memory	115
6.5	Conclusion	117
	Bibliography	121
	Appendices	133
A	Parameter Estimation	134
A.1	Experiments 1 and 2	134
A.2	Experiment 3	136
A.3	Experiment 4	136
B	Parameter Recovery Simulations	137
B.1	Extended UVSD Model	137
B.2	Extended DPSD Model	139
B.3	Extended Models of Recognition and Priming	141

List of Figures

1.1	A visual representation of the unequal-variance signal detection model of recognition memory. This depicts the memory strength distributions for old and new items, their means, and the central decision criterion C . The new item mean is fixed to zero, with the old item mean being a free parameter, and equal to d , the distance between the old and new item distributions. Also labelled are the four types of recognition responses represented in the model: hits (H), misses (M), false alarms (FA) and correct rejections (CR).	10
2.1	Density plots of $M - H$ against parameters σ_o and R , with data points representing the observed values of each RT measure and parameter estimate from Experiments 1 and 2. In each plot, darker colours indicate a greater density and lighter colours indicate a lower density. Data in Panel A was generated using the SS model, and Panel B was generated using the DPSD1 model.	23
3.1	A depiction of the extended UVSD model specification, with parameters set to the mean estimates recovered from Experiment 3.	48
3.2	Raincloud plots of σ parameter estimates and discriminability measures in the UVSD model, with circular points denoting means. The mean of each log-sigma parameter was calculated and then exponentiated.	53
3.3	An example of a fragmented word image stimulus from Experiments 4, 7, 8, and 9.	57
3.4	Density plots showing the extended SS and DPSD1 model predictions about the M-H difference between variability conditions as estimates of σ and R parameters change, respectively. Panel A shows data from the SS model expected value simulation when μ and σ parameters were independently generated. Panel B shows data from this simulation when σ parameters were related according to Equation 3.1. Panel C shows data from the DPSD1 model expected value simulation.	59
3.5	Raincloud plots of σ parameter estimates and discriminability measures in the SS model, with circular points denoting means. The mean of each log-sigma parameter was calculated and then exponentiated.	65
3.6	Bar plots of differences between identification performance for items receiving different types of recognition responses across high and low conditions. H = hit, M = miss, FA = false alarm, CR = correct rejection. New - Old refers to the priming effect. Mean expected values from the SS and the DPSD1 models are denoted by the circular and triangular points for each of these measures. The error bars represent the 95% confidence intervals of each mean. Both models predicted above-zero differences in each comparison.	66

4.1	An example of a fragmented stimulus used in an identification trial from Experiments 5 and 6. Participants were instructed to identify each stimulus at the lowest level of image clarity possible (denoted by labels 1-15). If they could not identify the stimulus, they could press a button to reveal more of the image until it was completely visible at level 15.	77
4.2	Mean fragment identification level per participant for each recall response type in Experiments 5 and 6. The red dots denote the mean of means, and the whiskers the 95% confidence interval.	82
5.1	Mean fragment identification level per participant for recalled and non-recalled items. Panel A shows results from Experiment 7, and panel B shows results from Experiment 8. The red dots denote the grand mean, and the whiskers the 95% confidence interval.	94
5.2	Mean fragment identification level per participant for recalled and non-recalled items from Experiment 9. The red dots denote the mean across participants, and the whiskers the 95% confidence interval.	101

List of Tables

2.1	Generative lower and upper bounds on the uniform distributions used to simulate true parameters.	22
2.2	Mean hit rates, mean false alarm rates and mean identification RTs for each item type from Experiments 1 and 2 (SE in parentheses).	28
2.3	Mean parameter estimates for the SS and DPSD1 models in Experiments 1 and 2 (SDs in parentheses)	30
3.1	The proportion of significant item characteristic regression models and mean R^2 values (standard deviations in brackets) for each condition in Experiments 3 and 4.	52
3.2	Mean parameter estimates for the extended UVSD and DPSD models in Experiment 3, with standard deviations in parentheses.	52
3.3	The lower and upper bounds on the uniform distributions used to generate simulated data from the SS and DPSD1 models. Values of μ_{nh} were fixed to 0 in the SS model simulation, and values of each σ parameter was fixed to $\sqrt{0.5}$ in the DPSD1 model.	58
3.4	Mean parameter estimates for the extended SS and DPSD1 models in Experiment 4, with standard deviations in parentheses. Dashes indicate that a given parameter does not belong to the model in question.	63
B.1	Generative lower and upper bounds a and b on the uniform distributions used to simulate true parameters, the means of the true and estimated parameters given by the extended UVSD model (SDs in parentheses), and the Bayes Factors of the comparisons between true and estimated parameters.	138
B.2	Generative lower and upper bounds a and b on the uniform distributions used to simulate true parameters of the extended DPSD model, the means of the true and estimated parameters given by the model (SDs in parentheses), and the Bayes Factors of the comparisons between true and estimated parameters.	139
B.3	Generative lower and upper bounds a and b on the uniform distributions used to simulate true parameters of the extended SS model, the means of the true and estimated parameters given by the model (SDs in parentheses), and the Bayes Factors of the comparisons between true and estimated parameters.	140
B.4	Generative lower and upper bounds a and b on the uniform distributions used to simulate true parameters of the extended DPSD1 model, the means of the true and estimated parameters given by the model (SDs in parentheses), and the Bayes Factors of the comparisons between true and estimated parameters.	140

Chapter 1

Introduction

Human memory manifests itself in many ways. Since the advent of cognitive psychology, theorists have attempted to draw distinctions between different expressions of memory (for review, see Squire & Dede, 2015). Later, neuropsychological evidence led researchers to hypothesise the existence of two functionally and neurally distinct *explicit* and *implicit* memory systems (Gabrieli, 1998; Squire & Dede, 2015). There is some variation in the way these terms are used in the memory literature. Explicit memory can refer to memory that is accompanied by a conscious awareness of remembering, whereas implicit memory is expressed indirectly, often outside of a participant's awareness. For example, long-term repetition priming (henceforth, *priming*), in which measures of identification, production, or detection are improved after previous exposure to an item, is one expression of implicit memory. Explicit and implicit memory can also refer to different classes of memory tasks on the basis of the level of conscious awareness they elicit in the participant (Roediger & McDermott, 1993). However, not all tasks are considered "pure" expressions of either type of memory. For instance, a free recall task, in which a participant must retrieve as many items as possible from a studied list, is often considered an explicit memory task because it requires a deliberate reinstatement of a study context to complete. However, performance in a recognition task, in which a participant judges whether they have seen an item before in a previous context, can be affected by explicit or implicit memory (Tulving, 1985). The terms *declarative* and *non-declarative* have also been used interchangeably with explicit and implicit memory to describe hypothesised memory stores, systems, or sources in the brain (Squire & Dede, 2015).

Other memory processes that are accompanied by different levels of awareness have been theorised. One such subdivision is that between recollection and familiarity — two functionally and neurally separable processes that are tied to different phenomenological experiences of memory (Yonelinas

et al., 2022; Yonelinas, 2002). Recollection is a memory process that comes with a clear and immediate sense of remembering details from a previous study context. Familiarity refers to memory that does not include strong contextual details, where the participant relies on a more abstract feeling of "knowing" a stimulus (Mandler, 1980). The experience of familiarity differs from the colloquial definition; Yonelinas et al. (2022) give the example that a participant may be aware that a stimulus in front of them is a dog (and so are "familiar" with it), but that this is distinct from the episodic familiarity that leads a participant to know they encountered that specific dog previously. Recollection and familiarity are assumed to be conditionally independent processes, in that the occurrence or absence of recollection implies nothing about an item's familiarity value (Yonelinas, 1994). This theoretical account of memory is often referred to as "dual-process theory" (Wixted, 2007).

While recollection and familiarity pertain to different levels of awareness in memory, they differ from explicit or implicit memory. While recollection is seen as a reflection of explicit memory, familiarity is often also considered an explicit process (Addante, 2015; Voss et al., 2007). This is perhaps because it is accompanied by a conscious sense of remembering and is invoked in direct tests of memory, despite offering the participant less insight into the factors that influence their memory decision. Despite this, familiarity has been suggested as a contributor to performance in priming (Jacoby & Dallas, 1981) and has been represented as such in mathematical models (Berry et al., 2012). It is important to note that even while making this latter assumption, recollection and familiarity still differ from the explicit and implicit systems of multiple-systems theory. Advocates of the dual-process and multiple-systems accounts have proposed different neural mechanisms underpinning each set of systems. Dual-process theorists state that recollection is related to hippocampal and parahippocampal activity and that familiarity is associated with activity in the perirhinal cortex (Yonelinas et al., 2022). In the multiple-systems view, however, memories that elicit feelings of recollection and familiarity are both driven by the same medial temporal lobe system (Squire & Zola-Morgan, 2015), and are not subdivided by different neural processes. While there is variation in the way in which recollection and familiarity have been previously defined, it is widely acknowledged that recollection is an expression of memory that is accompanied by a conscious awareness of remembering (Yonelinas et al., 2022). In this thesis, I discuss dual-process models and theoretical accounts assuming that familiarity can be responsible for implicit memory tasks such as priming. This reflects the outlook of existing models of recognition and priming and allows the predictions of dual-process and multiple-systems views to be more easily reconciled.

Over the course of decades, evidence has accumulated that appears to support the existence of

independent explicit and implicit memory systems (Squire & Dede, 2015), and for the separate contributions of recollection and familiarity to memory. Much of this evidence has involved demonstrations that tasks involving explicit memory are unrelated to measures that reflect more indirect expressions of memory. For instance, many studies have focused on the distinction between recognition memory and long-term repetition priming. However, such evidence has also been routinely questioned on statistical and methodological grounds, and in some instances, has failed to replicate (Berry et al., 2006a, 2010). Based on this criticism and evidence from mathematical models, some researchers have proposed that tasks with explicit and implicit components are governed by a single cognitive system (Berry et al., 2012). This view opposes both the multiple-systems and dual-process accounts and has proved a subject of controversy and debate. I now review the evidence that has been taken to support each account, before establishing novel tests of dual-process and single-system accounts of explicit and implicit memory.

1.1 Functional and Neuropsychological Dissociations

A functional dissociation is a result that shows an experimental variable has different effects on explicit and implicit memory measures. Such results have previously been taken as evidence for the independence of explicit and implicit memory systems because they appear to show that expressions of these systems behave according to different rules or principles (Berry et al., 2012). Neuropsychological dissociations arise when patients with specific neurological impairments show different levels of performance in explicit and implicit memory tasks. This outcome is often interpreted to show that each system relies on different brain regions, and so are neurally independent. Such dissociations from studies of amnesic individuals led to the initial identification of the explicit/implicit dichotomy in human memory and are often regarded as compelling support for a multiple-systems account (Squire & Dede, 2015).

Much of the functional dissociation evidence for a multiple-systems view of explicit and implicit memory focuses on recognition and priming. For instance, Jacoby & Dallas (1981) found that a levels-of-processing manipulation in which participants answered questions about the semantic meaning of an item during the study phase increased recognition memory, but not subsequent perceptual identification performance. A similar benefit of semantic encoding was also found for free recall, cued recall, and word recognition by Graf & Mandler (1984), who did not observe this effect for an implicit word-stem completion task. Similar dissociation evidence has shown other levels of processing manipulations improve recognition memory but have no effect on priming

(Richardson-Klavehn et al., 1999). Manipulations of item modality between study and test phases have also shown effects on priming with only small effects on recognition (Craik et al., 1994; Jacoby & Dallas, 1981). Finally, Jacoby (1983) reported a double dissociation in which a single experimental manipulation had opposite effects on recognition and priming. In this study, priming was greater when words were read during the study phase, rather than being generated from their antonym (for example, "hot" being generated from "cold"). However, recognition performance improved in this latter condition, showing that performance in each task was improved by opposing experimental conditions.

However, much of this dissociation evidence is undermined by conflicting results. There is evidence to suggest that levels of processing manipulations have inconsistent effects across a variety of explicit and implicit memory tasks, including recall measures (Challis et al., 1996). Although item modality changes have been shown to affect priming more than recognition, Mulligan & Osborn (2009) found a reliable effect on recognition as a result of this manipulation. Despite one conceptual replication of Jacoby (1983) in the auditory modality (Dew & Mulligan, 2008), other item generation manipulations have not been shown to reduce priming (Masson & MacLeod, 1992, 2002; Mulligan & Dew, 2009). A crossover dissociation between recognition and priming was also previously reported by Voss & Gonsalves (2010), who showed that studying items for 2000 ms versus 250 ms increased subsequent recognition yet led to a smaller priming effect. However, Berry et al. (2017) found no evidence for this dissociation across seven experiments. Instead, increased study duration led to either no detectable effect on priming, or a comparable increase in priming and recognition, consistent with a single-system view of the two tasks. These results cast doubt on the generality of many functional dissociations taken as evidence that explicit and implicit memory measures are differently affected by common experimental manipulations.

Behavioural and neurological dissociations have also been observed between recall and implicit memory. Paller (1990) found that both free and cued recall performance were improved by explicit instructions to remember rather than to forget certain studied words. By contrast, word stem priming was not affected by this directed forgetting manipulation. EEG patterns for these tasks also differed, with greater differences in ERP activation being observed for correct and incorrect free and cued recall trials, compared to correct and incorrect stimuli in the priming task. Hunt & Toth (1990) tested the effect of orthographic similarity on memory performance, finding that free recall was better for orthographically distinctive words, but perceptual identification improved for orthographically common words. However, fragment completion, a measure of conceptual priming, was also better for distinctive than common words, suggesting some crossover between explicit and

implicit memory. Roediger & Challis (1992) found that viewing exact repetitions of a studied word (e.g. "elephant", "elephant") improved free recall performance, though only had a small effect on primed word fragment completion performance. Repeating a conceptually related word at study (e.g. "elephant", "tusk") also improved free recall, yet had no effect on word fragment completion. There is also evidence that clinically depressed participants freely recall more negatively valenced than positively valenced words, yet display equivalent conceptual priming for both word classes (Denny & Hunt, 1992). Further evidence has disputed the lack of emotional bias in conceptual priming, though found a lack of evidence for this effect in perceptual priming (Watkins, 2002). Although some of these results find a modest overlap between recall and priming, most provide evidence that recall-based tasks and measures of implicit memory can be dissociated experimentally. Such results are often interpreted as evidence of multiple memory systems influencing each task (Hunt & Toth, 1990).

There have also been many demonstrations that patients with amnesia, often as a result of damage to the hippocampus or medial temporal lobes (MTL), have impaired explicit memory but relatively intact implicit memory. Several studies have shown such individuals have impaired recognition but normal levels of priming (Cermak et al., 1985; Graf et al., 1984). Gabrieli et al. (1995) also showed that individuals with damage to the right occipital lobe have impaired visual priming but above-chance recognition (but see Yonelinas et al., 2001). Taken together, these results form a double dissociation, showing that recognition and priming depend upon different neural regions. Neurological case studies of individuals with profound amnesia have also provided dissociation evidence. The first such evidence followed studies of H.M., an individual who underwent a bilateral resection of the MTL following severe epilepsy, removing much of his hippocampus and parahippocampal gyrus. Researchers found that although H.M. had very poor episodic memory, he was able to learn perceptual and procedural skills (mirror-reading and mirror-writing), within three days (Squire, 2009). Similarly, Clive Wearing, an individual with profound retrograde and anterograde amnesia, has some intact procedural memory despite experiencing profound episodic and semantic memory deficits (Wilson & Wearing, 1995). Multiple studies of E.P., another profoundly amnesic individual with bilateral MTL lesions, showed their priming performance to be comparable with healthy control subjects (Conroy et al., 2005), despite chance-level recognition (Stefanacci et al., 2000).

Although dissociations between explicit memory and priming show that performance in each task can be differentially affected, this does not necessarily imply that each task is a product of a different memory system. Single dissociations, where a variable has an effect on one measure but

no detectable effect on another, can be explained by factors that are not related to memory. For instance, priming has been shown to have lower measurement reliability than recognition (Buchner & Wippich, 2000), and so experimental manipulations are less likely to produce detectable priming effects than recognition effects. Single dissociations between recognition and priming can therefore be accounted for without assuming multiple memory processes. Indeed, mathematical models that assume recognition and priming emerge from a single memory strength source can account for single dissociations with parameters that reflect measurement error (Berry et al., 2008a,b).

Single dissociations also rely on the acceptance of the null hypothesis, which is challenging in practice. Firstly, the size of a true effect can be small, and if so, might not be detected without a large sample size. Researchers may therefore make type-II errors in small-sample studies by accepting the null hypothesis where an effect is present in the population. This may lead them to falsely conclude that an experimental manipulation has no effect on a given variable, contributing to an erroneous single dissociation. As Berry et al. (2012) highlight, this potential for error also applies to instances where two opposing single dissociations are combined to form a double dissociation. Secondly, many studies that claim to show single dissociations use frequentist statistical methods that quantify the strength of evidence for the alternative hypothesis against a point null hypothesis. It is a commonly-held view that any statistical method is incapable of accepting a point null hypothesis (for instance, that an independent variable has precisely zero effect on a dependent variable) at an arbitrary level of precision (Cohen, 1994; Morey & Rouder, 2011). As such, statistical methods that are able to quantify the strength of evidence for an interval null, such as confidence intervals and Bayesian methods, are better suited for assessing the absence of effects in data. Any dissociations concluded from frequentist tests using the point null are therefore undermined by a lack of conclusive statistical evidence for the null hypothesis. Taken together, these concerns cast doubt on the support given by dissociation evidence in favour of a multiple systems account of explicit and implicit memory.

1.2 Neuroimaging

Further evidence for a multiple-systems account of explicit and implicit memory comes from neuroimaging studies. Functional imaging studies have shown that recognition memory and priming have different neural correlates at encoding in fMRI (Schott et al., 2006) and different ERP effects (Woollams et al., 2008; Park & Donaldson, 2016). There is also evidence that cued recall in the form of word stem completion with studied stimuli activates different brain regions from

priming (Schott et al., 2005). Together, these studies show clear evidence that explicit and implicit memory tasks activate separate brain regions, and are therefore likely to rely on different neural systems. However, some recent evidence has also suggested that explicit and implicit memory activate similar brain areas, or have other neural dependencies. Despite finding that recognition and priming gave distinct neural signals, Park & Donaldson (2016) found that short-term priming sped the onset of old/new ERP effects in the left parietal cortex by 300 ms, compared with trials that were not primed. Since the left parietal cortex is related to recollection, Park & Donaldson (2016) concluded that this was novel evidence that short-term priming could have effects on explicit memory processes, yet it is not clear whether long-term repetition priming may have similar effects. Leading from analyses of amnesic patients, it has been claimed that the hippocampus is necessary for explicit memory but not for implicit memory (Squire, 2009). Yet, Addante (2015) found that amnesia patients with MTL damage limited to the hippocampus showed implicit memory impairments compared with a control group. These impairments could not be accounted for by explicit contamination or other control factors, leading Addante (2015) to conclude that explicit and implicit memory both depend upon the hippocampus (see also Hannula & Greene, 2012), in contrast with the multiple-systems view that the hippocampus drives explicit, but not implicit memory (Squire & Dede, 2015). Kim (2019) conducted a meta-analysis of functional neuroimaging studies that examined explicit and implicit memory tasks. Assessing neural signals of encoding and retrieval effects, they concluded that explicit and implicit encoding is driven by largely overlapping neural regions. However, the regions that are responsible for explicit and implicit retrieval are mostly segregated, suggesting some separation between the two expressions of memory. Although most neuroimaging evidence has found some degree of separation between the neural correlates of explicit and implicit memory, this more recent evidence shows there are some reliable commonalities between these subdivisions of memory.

1.3 Fluency

There are claims that fluency — the speed at which an item is processed — is responsible for associations between recognition and priming. For instance, increased fluency can lead to faster identification times in a priming task, the speed of which may be detected by the participant. The unexpected fluency at identification may then be attributed to having seen the stimulus previously, resulting in recognition of that stimulus. This aligns with a dual-process account where familiarity is considered a product of fluency, which can influence both recognition and priming (Jacoby &

Dallas, 1981; Mandler, 1980). Conroy et al. (2005) tested this idea by having amnesic individuals and healthy control participants complete a continuous identification with recognition (CID-R) procedure (Feustel et al., 1983; Stark & McClelland, 2000). In a typical CID-R trial, participants complete a perceptual identification task and then immediately make a recognition judgement for the stimulus they just identified. For example, they might see a word repeatedly presented in fast alternating succession with a mask, as in Stark & McClelland (2000). In the first instance, the word might be presented for 17 ms, and the mask presented for 233 ms, for a presentation block lasting for a total of 250 ms. In the next repetition of this stimulus-mask block, the stimulus duration would increase to 33 ms, and the mask duration would decrease to 217 ms. The exposure duration of the word continues to increase until the participant identifies the word, or until the exposure of the word takes up the whole presentation block. Immediately after submitting their identification response, participants then make a recognition judgement for the word they just identified. This procedure allows for recognition judgements and identification RTs that provide a measure of implicit memory to be collected for each stimulus, enabling the conjoint modelling of recognition and priming.

Conroy et al. (2005) found that amnesic individuals showed impaired recognition but comparable priming to healthy control participants in this task. Furthermore, both amnesic and control participants gave faster RTs for items judged old than those judged new, regardless of their true status as studied or unstudied. This was taken as evidence that amnesic individuals experience fluency, but cannot use it to aid their recognition judgements as part of their deficit, explaining the dissociation between their recognition and priming performances. However, this result can be explained by a single-system model of explicit and implicit memory (Berry et al., 2008a). Such a model also gave a better quantitative fit to the majority of the data from Conroy et al. (2005) than alternative models that assumed multiple memory systems (Berry et al., 2012). This shows that results surrounding the role of fluency in explicit and implicit memory are not constrained to supporting a dual-process or multiple-systems account, and may also be compatible with a single-system view.

1.4 Priming in the Absence of Recognition

Reports of priming effects in experimental conditions where recognition memory is at chance have been taken to support the existence of separate explicit and implicit memory systems. This hinges upon the logic that if implicit memory is not accessible to awareness, then it should be possible for individuals to show implicit memory without reliable explicit memory performance. However, several attempts to replicate results that demonstrate priming without recognition have

failed (Berry et al., 2006b, 2010). Other results show that the magnitude of priming effects lessens when recognition approaches chance (Berry et al., 2006a; Moscovitch & Bentin, 1993), suggesting continuity between these expressions of memory. Stark & McClelland (2000) studied identification RTs for items that received certain recognition responses; namely, "miss" trials where old items were incorrectly judged new, and "correct rejection" trials where new items were correctly judged new. Stark & McClelland (2000) reported that RTs for misses were faster than those for correct rejections in a CID-R task. This was taken as evidence of a priming effect within the set of items that were not recognised, implying a contribution of implicit memory in the absence of explicit memory and supporting multiple systems. However, a single-system model of recognition and priming assumes items that receive miss responses have a greater level of memory strength than those that are correctly rejected (Berry et al., 2008a). Since greater memory strength results in faster identification RTs in this model, this result is also compatible with a single system account. Results that purport to show priming without recognition therefore do not necessitate the existence of separate explicit and implicit memory systems.

1.5 Signal Detection Theory Models of Explicit and Implicit Memory

Important evidence in the memory systems debate has come from formal models that represent explicit and implicit memory in the same model. As a technique for validating theoretical positions, formal models have many advantages over theoretical positions that are verbally specified (Guest & Martin, 2021). In order to be formally specified, a model must formalise auxiliary assumptions and mechanisms that verbal theories may leave vague or omit completely. This allows unambiguous predictions to be derived from model specifications, leading to direct tests of the model in question. Some of these predictions may even be counter-intuitive and not immediately apparent from a verbal inspection. Formal predictions are also less flexible than verbal predictions. A single verbal prediction can often be interpreted in different ways, and is therefore more difficult to test conclusively than a formal equivalent (Farrell & Lewandowsky, 2018). These advantages make formal modelling an ideal way of comparing multiple theoretical positions (Guest & Martin, 2021). The goodness of quantitative fit of different models to data can be used to evaluate their ability to represent phenomena of interest. Unique qualitative predictions made by different models can also be tested to provide strong support for and against candidate models. This method of strong inference can lead to swift scientific progress, constraining theoretical positions by ruling out alternative explanations (Platt, 1964). In this way, signal-detection theory models representing

single-system and multiple-systems views of explicit and implicit memory have made theoretically significant contributions to the memory systems debate.

Signal detection theory (SDT; Green et al., 1966; Macmillan & Creelman, 2004) is a prominent modelling framework in the study of human memory. First used to model recognition memory (Egan, 1958; Green et al., 1966), its assumptions in this domain have been widely supported by several modelling exercises and critical tests (Kellen et al., 2021). It has since been extended to represent other memory judgements in conjoint models, such as priming (Berry et al., 2012), source memory (Lange et al., 2019), source attribution (Marsh & Landau, 1995), and meta-cognitive memory judgements (Jang et al., 2012b). Signal detection theory provides useful descriptions of many aspects of memory performance, and, when extended to form conjoint models, can be used to test theoretical positions on the memory systems debate.

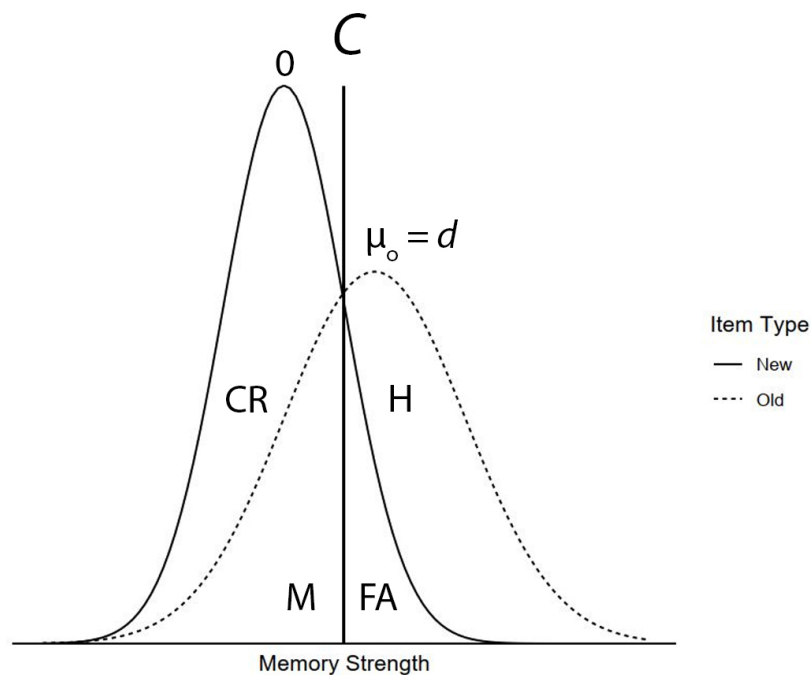


Figure 1.1: A visual representation of the unequal-variance signal detection model of recognition memory. This depicts the memory strength distributions for old and new items, their means, and the central decision criterion C . The new item mean is fixed to zero, with the old item mean being a free parameter, and equal to d , the distance between the old and new item distributions. Also labelled are the four types of recognition responses represented in the model: hits (H), misses (M), false alarms (FA) and correct rejections (CR).

In a standard signal detection model of recognition memory, the memory strength for old and new items is represented by two Gaussian distributions along a unidimensional continuum (see Figure 1.1). Because old items are studied, the mean of the old item strength distribution is greater than that of the new item strength distribution. The difference between these old and new item means (d)

is taken as a measure of recognition performance; the level of discriminability between old and new items. To model recognition judgements, each item's memory strength value can be compared with decision criteria along the memory strength continuum. Simple binary recognition judgements are modelled with one criterion. If an item's strength value exceeds the value of the criterion, then it is judged "old"; if it is less than the value of the criterion, it is judged "new". In the case of confidence judgements with more decision categories, additional criteria are introduced, with the total number of criteria being one less than the number of decision categories.

The first attempts to model recognition memory in this way assumed that the variances of the old and new item distributions (σ_o and σ_n) were equal (Egan, 1958). However, this model is unable to account for benchmark results found in recognition memory data. Most notably, the model cannot explain patterns in the z -ROC curve; the z -transformed plot of the hit rate (proportion of correctly recognised old items) against the false-alarm rate (the proportion of new items incorrectly recognised) at different levels of the response criterion (Spanton & Berry, 2020). Many analyses have found that z -ROC curves from recognition memory data are linear, and have a slope that is consistently less than 1 (Glanzer et al., 1999). As the z -ROC slope is considered a proxy for the ratio σ_n/σ_o , an equal variance model can only produce slopes equal to 1. To address this, σ_o is often made a free parameter, usually taking a value greater than σ_n , which is fixed to 1. This unequal variance signal detection (UVSD) model is favoured over an equal variance model when applied to recognition memory data (Egan, 1958; Green et al., 1966; Rotello, 2017; Wixted, 2007).

1.5.1 A Single-System Model of Recognition and Priming

To represent both explicit and implicit memory, the basic signal detection model described above has been extended to account for priming. Berry et al. (2006a) proposed a single-system model in which recognition and priming share a common memory strength signal (see also Berry et al., 2008a,b). This model was later expanded into a framework that included additional models assuming different levels of separation between the strength signals driving recognition and priming (Berry et al., 2012). In the resultant single-system (SS) model, recognition memory and priming arise from the same memory strength signal, a Gaussian random variable $f \sim \mathcal{N}(\mu_I, \sigma_I)$ where I represents each old or new item type. Because prior exposure during the study phase facilitates an increase in memory strength for studied items, it is assumed that the mean f for old items (μ_o) is greater than the mean f for new items (μ_n). Despite support for the unequal variance assumption in recognition memory, Berry et al. (2012) chose to equate the values of σ_o and σ_n in their initial implementation

of the model for computational simplicity. Although this decision made the model unable to account for z -ROCs from their data, they verified that their equal variance assumption did not change any of the model's key predictions. In line with the recommendation of Berry et al. (2012) and the evidence in support of the unequal variance assumption, implementations of the SS model in the present work assume σ_o is free to be greater than σ_n . This assumption has also been formalised in more recent extensions of the SS model (Lange et al., 2019).

In the SS model, recognition memory strength is calculated by summing f with a Gaussian noise variable $e_r \sim \mathcal{N}(0, \sigma_r)$. From this, we obtain the recognition strength variable, J_r , where

$$J_r = f + e_r. \quad (1.1)$$

Binary old or new recognition judgements are then represented by comparing J_r to a strength criterion C , as in classic signal detection models of recognition memory. Values of $J_r > C$ represent items judged old, and values of $J_r < C$ represent items judged new. J_r can also be compared to multiple criteria as with the strength signal in a traditional signal detection model. To model priming in an identification task, the same value of f used to generate recognition memory strength is summed with another Gaussian noise variable, $e_p \sim \mathcal{N}(0, \sigma_p)$. The resultant identification measure is then given by the equation

$$ID = b - sf + e_p \quad (1.2)$$

where b and s are scaling parameters. b represents the identification RT intercept and is equal to the expected identification RT for new items. s represents the rate of change in RT with f . Therefore, a high value of f will likely result in a high recognition memory strength value (and thus a high confidence "old" judgement) and a high-performance identification measure.

1.5.2 Multiple Systems Models of Recognition and Priming

Berry et al. (2012) also defined two models that assume recognition and priming are driven by multiple strength signals. In their more stringent multiple systems (MS1) model, recognition strength and identification performance are derived from separate, uncorrelated strength sources; f_r and f_p , respectively. In the more flexible MS2 model, the correlation (w) between f_r and f_p , is a free parameter, allowing a range of possible associations between memory strength in recognition

and priming. Substituting the relevant f distribution, J_r and ID are calculated in the same way as in the SS model. This similarity and the inclusion of w nests the SS and MS1 underneath the MS2 model. The SS model is equivalent to the MS2 model when $w = 1$ and $\mu_r = \mu_p$, and the MS1 model is equivalent to the MS2 model when $w = 0$.

Outside of this family of models, Berry et al. (2012) also defined a dual process signal detection (DPSD1) model of recognition and priming, based on the previous dual-process signal detection (DPSD) model by Yonelinas (1994). In the DPSD model, recognition decisions can result from one of two processes. The model proposes a probability, R , that a given old item is recollected. Recollected items receive the highest confidence "old" recognition judgement. If not recollected, recognition judgements are made on the basis of familiarity, which is represented as an equal variance signal detection process. Berry et al. (2012) extended this specification to include priming. In the DPSD1 model, recognition is also assumed to be driven by either a probabilistic recollection process or a continuous familiarity signal. Priming is driven exclusively by this same familiarity strength signal and is not influenced by recollection. This means that the recollection processes are separate; recollection does not affect priming, and the contributions of recollection and familiarity to recognition memory are conditionally separable. This means that the DPSD1 model can be seen as a "multiple-systems model" (Berry et al., 2012).

1.5.3 Evidence for Single and Multiple Systems Models

Many behavioural studies and model fitting exercises have given support to single-system model predictions about empirical data. The earliest single-system model proposed by Berry et al. (2006a) was found to predict several trends in empirical data across four experiments that manipulated attention during study. In all four experiments, priming was greater for words that were attentionally cued during study. There was also no reliable priming effect for uncued words when recognition performance was also at chance. Berry et al. (2006a) also found a correlation between priming and recognition performance in their Experiment 4. Each of these results was predicted by their single-system model, although they did not find reliable evidence for this latter correlation in their Experiments 2 and 3. This absence of a correlation was also found by Berry et al. (2008a) in their replication of Stark & McClelland (2000). On balance, however, the single-system model was able to predict most of the patterns of results observed in the data.

Berry et al. (2008a) identified predictions previously thought to be consistent with a multiple-systems account and demonstrated the single-system model's ability to explain these results.

Replicating Stark & McClelland (2000), they found evidence that priming reaction times in a CID-R task were faster for items that received miss responses than those that received correct rejection responses. Stark & McClelland (2000) asserted that this effect was evidence for a multiple-systems account. However, simulations by Berry et al. (2008a) demonstrated that the single-system model predicts the result, as a consequence of the shared memory strength signal for recognition and priming being greater for misses than correct rejections. Berry et al. (2008a) also demonstrated the single-system model's ability to account for patterns of priming reaction times in Johnston et al. (1985) and the finding that amnesic patients showed comparable priming with control participants despite impaired recognition (Conroy et al., 2005). This latter result aligns with other research that suggests a signal detection model with a single strength source for both recognition and priming can predict dissociation evidence (Shanks & Perruchet, 2002; Kinder & Shanks, 2003; Berry et al., 2008b, 2014). Each of these original results was previously taken to be evidence for multiple memory systems without reference to a formal model. However, Berry et al. (2008a) showed each result to be consistent with a single-system model. This demonstrates the utility of the single-system account, and the ability of formal models to make seemingly counter-intuitive predictions that may guide theoretical development.

Going further, Berry et al. (2012) implemented both the single and multiple systems accounts as mathematical models in a common framework, alongside the DPSD1 model. This allowed them to assess model-derived predictions reflecting both the single and multiple systems views and to compare the quantitative fit of the models to experimental data for the first time. Berry et al. (2012) found their SS model predicted many results observed across three experiments and in a re-analysis of data from Conroy et al. (2005). Strong support was found for four of the SS model's five predictions in this study, with the MS1 model being found insufficient to explain the observed results regarding all of these predictions. The other prediction of the SS model — that a reliable priming effect will never be observed when recognition performance is at chance — did not discriminate between the models. Since both the SS and MS1 models are nested mathematically under the MS2 model, these results were not able to provide strong evidence for the SS model against the MS2 model. The MS2 model was also able to predict priming in the absence of recognition for one amnesic patient (E.P.) in a re-analysis of Conroy et al. (2005), which the SS model could not. However, in order to do so, the MS2 model's parameters changed in a way that prevented it from accounting for other aspects of E.P.'s data. With this in mind, neither the SS or MS2 model proved superior on the grounds of their qualitative predictions.

Despite this, Berry et al. (2012) found strong evidence for the SS model in analyses of quantitative

fit. In all of their experiments, the SS model fitted better to data than the MS1 or MS2 models according to comparisons of the Akaike information criterion (AIC). The SS model also fitted best to pooled data from all experiments and all groups in the data from Conroy et al. (2005) except the case of patient E.P., which the MS2 model fitted best. The DPSD1 model also fit better than the SS model to data from Experiment 2, providing the best fit for that experiment. The superiority of the SS model's quantitative fits may be in part due to it having 9 free parameters, while the MS2 model has 10 with the addition of w . This makes the SS model more parsimonious than the MS2 model, allowing it to explain the key trends in the observed data with a less complex specification. This quality also makes the SS model more diagnostic than the MS2 model, as its fixed w parameter and $\mu_r = \mu_p$ constraint forces it to make a priori predictions about a wide range of phenomena. While the MS2 model can, by mathematical necessity, account for these predictions, its generality means that it is not constrained to make them. Berry et al. (2012) also found that in several cases, the MS2 model predicted a strong correlation between f_p and f_r , with w often taking values close to 1. This shows that the MS2 model often mimics the SS model, despite being more flexible. In all, Berry et al. (2012) not only showed the SS model can make accurate qualitative predictions but also give better quantitative fits with a simpler, more constrained specification than its multiple-systems counterparts.

The SS model has since been applied to explain other memory phenomena. Ward et al. (2013b) studied the decline in explicit and implicit memory resulting from normal ageing through the lens of single-system and dual-process accounts. Previous research has stated that priming remains intact in older people while recognition memory performance reduces. However, Ward et al. (2013b) found small but reliable reductions in priming in older adults compared with younger adults, but only when pooling data from two experiments (the same comparisons in each separate experiment were not significant). Fitting the SS, MS1, and MS2 models to this data, Ward et al. (2013a) found some support for the predictions of the SS model over the MS1 model, although some predicted differences were not reliable despite showing numerical trends. Although the results of Ward et al. (2013b) showed only a marginally significant decrease in priming with age, a highly powered study by Ward et al. (2020) gave further support for this result. Recruiting a sample of 1072 participants, they found that age was a significant predictor of both explicit and implicit memory decline for attended items. This gives evidence for an association between recognition and priming performance, and so is consistent with a single-system view of explicit and implicit memory.

The SS model has also been shown to predict memory trends in special populations. Berry et al. (2014) compared recognition memory and priming in patients with Korsakoff's amnesia and healthy

control participants. They found that the SS model predicted various trends in these data, including the tendency for items judged old in a recognition test to be identified at greater levels of obscurity regardless of whether they were actually present during encoding or not. The model also correctly predicted that amnesic patients would have deficits in both recognition and priming compared to a control group, contrasting previous evidence that amnesic patients have intact priming (Cermak et al., 1985; Gabrieli et al., 1995; Squire, 2009). Rothen et al. (2020) also applied the SS model to examine memory performance advantages in individuals with synesthesia. Their results showed synaesthetes had improved recognition of word stimuli over control participants. In regards to priming and fluency, there were only numeric trends toward a synaesthetic memory advantage, with no significant difference being found in these measures against controls. However, despite the SS, MS1 and MS2 models each predicting a synesthetic advantage in recognition, priming, and fluency, the SS model was able to account for these results with the most parsimonious specification. Once more, these results exemplify the SS model's explanatory power as a model of recognition and priming, compared with more complex multiple-systems models.

Compared with the SS, MS1, and MS2 models, the DPSD1 model Berry et al. (2012) has seen relatively little application. In Experiment 2 of Berry et al. (2012), the DPSD1 model predicted an improvement in identification performance as recognition confidence increased, but that the mean identification RT for items receiving the highest confidence rating would not be as short as that predicted by the SS model. This prediction was observed in empirical data, and the DPSD1 model gave a better quantitative fit to the data than the SS model in this experiment. However, in Experiment 3 of Berry et al. (2012), it was found that the mean identification RT of old items that received "remember" responses in a remember-know task were shorter than those for both items that received "know" responses and the mean identification RT for all old items. This conflicts with the DPSD1 model's prediction that each of these measures should be approximately equal on the basis that priming is determined solely by familiarity, and not recollection, which determines "remember" judgements (but see Parks & Yonelinas, 2007; Rotello et al., 2005). The SS model predicted this result successfully and also gave the best quantitative fit to data from Experiment 3; the DPSD1 model gave the worst fit of all the models considered.

Despite multiple accounts of the SS model's success in explaining recognition and priming data, the DPSD1 model has not been fit to data or directly compared with other models in published work since Berry et al. (2012). This is in spite of the enduring popularity of dual-process theory in the recognition memory literature (Cha & Dobbins, 2021; Wixted & Mickes, 2010). The MS1 model has been shown to consistently fall short of the SS model, and the MS2 model is hard to

disambiguate from the SS model given its flexibility. However, the DPSD1 model has scope for making novel predictions that oppose those of the SS model. Such competing predictions may allow for strong tests of single-system and dual-process accounts of recognition and priming in a promising new avenue of research.

1.5.4 Signal Detection Models of Recognition, Priming, and Source Memory

Adding to extensive tests of the SS, MS1, and MS2 models, Lange et al. (2019) developed single-system and multiple-systems models that also encompassed source memory. Source memory, like recognition memory, requires a participant to judge whether they have previously seen an item in a certain context. However, this context is not simply temporal; participants might judge whether an item at test was presented in a certain location at study, or against a particular background. Lange et al. (2019)'s single-system model assumed that one strength signal should drive priming, recognition, and source memory, and therefore predicted two main continuities between source memory and priming. Firstly, priming should be greater for items that receive correct, versus incorrect source decisions, and secondly, priming should increase with confidence in the source decision. These predictions were observed in the data, contrary to those of a multiple-systems model. However, analysis of quantitative fit put the models on closer footing. The single-system model fit best to data from Lange et al. (2019)'s first two experiments, with the multiple-systems model fitting better to their Experiments 3A and 3B.

Lange & Berry (2021) replicated the association between source memory and priming, before investigating the role of recognition and fluency in driving this association. They first removed the recognition test from their Experiment 2 procedure, and then removed any new items from their Experiment 3 to ensure participants did not covertly use recognition to influence their source judgements. Precluding overt and covert recognition judgements in this way, they found that the association between source memory and priming persisted, as predicted by a single-system model. In their Experiment 4, they also prevented fluency from influencing subsequent source memory decision by separating the source and identification test trials into blocked phases. Again, the association between source memory and priming was observed. Evidence from Lange et al. (2019) and Lange & Berry (2021) therefore weighs in favour of a single-system account where source memory and priming share a common strength signal, and that the association between performance in these tasks is not caused by a third variable.

Although Lange et al. (2019) formalised single-system and multiple-systems accounts of source

memory, they did not do so for a dual-process view. While source memory can be completed by an explicit reinstatement of a study context, it is also possible that familiarity or guesswork can aid in making a source judgement. This possibility is crucial for the dual-process explanation of the relationship between source memory and priming. If familiarity is assumed to aid source memory, a dual-process signal detection model of priming and source would account for this relationship in the same way as a single-system model (Lange et al., 2019). Early dual-process accounts of source memory assumed that familiarity could only contribute to source decisions if source attributes had different encoding strengths (Yonelinas, 1999). Later work also found evidence for a greater contribution of familiarity to source decisions, depending on task factors (Diana et al., 2008). However, neuroimaging evidence supporting this involvement is mixed. Some studies point towards the involvement of the hippocampus in source memory (Ekstrom & Bookheimer, 2007; Slotnick & Thakral, 2013) which is thought to reflect recollection (Yonelinas et al., 2022). However, there is some evidence that the hippocampus can also play a role in familiarity (Hannula & Greene, 2012; Addante, 2015), and some studies point toward the involvement of other regions in source memory (Kirwan et al., 2008).

Huang & Shanks (2021) also investigated the possibility that fluency, a familiarity-related process, influences the accuracy of multidimensional source memory decisions. They found reliable associations between identification RT speed and source correctness across multiple experiments, despite participants having no access to source attributes at test that could have prompted familiarity. Huang & Shanks (2021) concluded that the persistence of the relationship between priming and source memory in these conditions challenged dual-process models that assume familiarity does not contribute to source memory. Taken together, these results demonstrate that a strict dual-process account of source memory and priming is not sufficient to explain associations between performance in the two tasks. For the dual-process view to explain this association, familiarity must contribute to source memory, yet the extent of this contribution is subject to debate. By contrast, a single-system account of source memory and priming provides a valid and more parsimonious explanation of the results observed.

1.6 Outstanding Theoretical Questions Regarding the Memory Systems Debate

Although much research has investigated the relationship between explicit and implicit memory systems, theoretically significant questions remain unanswered. While some evidence was found by

Berry et al. (2012) in support of the SS model over the DPSD1 model, these models have not been directly compared since. Dual-process theory remains a popular framework through which memory can be conceptualised, and so is important to test in the context of recognition and priming. By comparing new predictions of the DPSD1 and SS models, a single-system account of recognition and priming can be evaluated against a dual-process alternative. The result of this evaluation can therefore provide new evidence for either a single-system or dual-process account of explicit and implicit memory, in the context of recognition and priming.

Compared with recognition and priming, and even source memory and priming, there has also been relatively little evidence concerning the relationship between implicit memory and recall. As a task that involves a complete stimulus retrieval from memory, recall is often thought to require conscious effort to complete, making it a relatively pure expression of explicit memory. Indeed, dissociations have been reported between recall and implicit memory measures (Paller, 1990; Hunt & Toth, 1990; Roediger & Challis, 1992). Yet, other evidence has shown that recall can either be associated with implicit memory performance (Mazancieux et al., 2020) or that implicit memory can influence recall measures (Ozubko et al., 2021). This sets a precedent for further tests of the association between performance in cued and free recall tasks and implicit memory. If indeed there are continuities between recall and implicit memory, such results would add support to a single-system account against strict multiple-systems or dual-process alternatives, informing future theory and model development.

I investigate these questions in the following chapters. In Chapter 2, I identify unique predictions about priming RTs in a CID-R task made by the SS and DPSD1 models. Experiments 1 and 2 test these predictions. Chapter 3 focuses on discriminating single-system and dual-process models using manipulations of variability in recognition memory. Experiment 3 validates a manipulation of encoding variability using the UVSD and DPSD models of recognition memory. This design is then extended in Experiment 4 to encompass priming, again prompting the SS and DPSD1 models to make opposing predictions about identification measures. Chapter 4 investigates the continuity between cued recall and implicit memory. Experiment 5 does so in a method that tests identification, recognition, and cued recall, and Experiment 6 simplifies this method by focusing solely on identification and cued recall. Finally, Experiments 7, 8, and 9 in Chapter 5 investigate the relationship between implicit memory and free recall performance. Taken together, these results provide evidence that contributes toward validating theoretical stances on the memory systems debate, ultimately favouring a single-system account over a strict dual-process alternative.

Chapter 2

Testing Single-System and Dual-Process Models With Response Speeding Manipulations

Since the SS and DPSD1 models (Berry et al., 2012) represent contrasting theoretical accounts, they have important differences in their specifications. Both the SS and DPSD1 models assume that recognition judgements and identification responses can arise from a common memory strength source. However, the DPSD1 model also assumes that a proportion of recognition judgements depend on a threshold recollection process. This recollection process has no bearing upon priming, in contrast to the familiarity process that underlies all other recognition responses and identification performance for all items. This inclusion of recollection, a separate, wholly explicit memory process, is the crucial difference between the SS and DPSD1 models. It is also one of the most important for discriminating the two models, as it can prompt them to make opposing predictions about priming in the face of certain recognition conditions. In this chapter, I identify one such pair of opposing predictions, verify them by way of simulation and mathematical analysis, and present two experiments to test them using the CID-R paradigm.

As previously established, the DPSD1 model includes a free parameter, R , that corresponds to the probability that a studied item is recollected. Recollected items are given high-confidence "old" recognition ratings and are assumed to have very high recognition memory strength (Parks & Yonelinas, 2007). Yet, R can increase independently of μ_0 , the mean of the distribution of familiarity strength for studied items. As R increases, the relationship between recognition confidence and identification performance for studied items in a CID-R procedure weakens, as fewer recognition judgements are determined by the familiarity signal that drives priming. As a result, the mean identification RTs for items that receive hit or miss responses converge towards the mean identification RT across all studied items. This is because recollected hits do not necessarily have

high familiarity strength and therefore shorter identification RTs. This leads the DPSD1 model to make a seemingly counter-intuitive prediction; that the difference between the mean RTs for miss and hit responses, $M - H$, will slightly decrease as R increases. In formal terms, the DPSD model predicts that $M - H$ is a decreasing function of R within the interval of the parameter's bounds $[0,1]$.

By contrast, the SS model assumes that the same strength signal, f , drives both recognition and priming. As with the univariate signal detection models of recognition memory that preceded it, the SS model does not treat strong, highly confident memory for studied items as a separate phenomenon from other types of recognition memory. Because "recollected" items receive high confidence recognition judgements based on very high memory strength, greater recollection signifies greater overall memory strength for old items. Greater recollection also results in old items having a greater spread of strength values at the higher end of the strength continuum, increasing the variability in recognition memory strength for these items. So, data that leads to greater recollection in the DPSD1 model are represented in the SS model by increases in the parameters μ_o and σ_o , the mean and variance of the old item memory strength distribution. Unlike in the DPSD1 model, however, this parameter adjustment also has an effect on priming. Specifically, increases in μ_o and σ_o increase the difference between mean identification RTs for hits and misses. So, in the face of increasing "recollection", the SS model predicts that $M - H$ increases as σ_o increases. This opposes the DPSD1 model's qualitative prediction, allowing for the models to be discriminated by a common experimental manipulation.

2.1 Expected Value Simulations

To confirm that the models make opposing predictions, I conducted simulations using the expected value functions of each model. These simulations, and all the other analyses in this thesis, were conducted using the statistical programming language R (Version 4.2.0; R Core Team, 2021). All Bayesian statistics in this and the following analyses were conducted using the BayesFactor package (Morey & Rouder, 2018). First, 100,000 sets of parameter estimates were randomly sampled from uniform distributions. The bounds on these distributions reflected a psychologically plausible parameter space typical of a CID-R experiment (see Table 2.1). Expected values of identification RTs for hit and miss responses were analytically calculated from the SS and DPSD1 model specifications using each set of parameter estimates. Expected values of identification reaction times conditional on a binary recognition judgement, Z , for the SS model were calculated

Table 2.1: Generative lower and upper bounds on the uniform distributions used to simulate true parameters.

Model	Parameter	Lower Bound	Upper Bound
SS	σ_o	$\sqrt{0.5}$	2.5
DPSD1	R	0	1
Shared	μ_o	0	4
	C	0	2
	b	1200	3500
	s	20	350

using the function

$$\lambda(Z, I) = b - s\mu_I + \frac{sw\sigma_f^2}{\sigma_{J_r}} \frac{\phi\left(\frac{C_j - \mu_I}{\sigma_{J_r}}\right) - \phi\left(\frac{C_{j-1} - \mu_I}{\sigma_{J_r}}\right)}{\Phi\left(\frac{C_j - \mu_I}{\sigma_{J_r}}\right) - \Phi\left(\frac{C_{j-1} - \mu_I}{\sigma_{J_r}}\right)} \quad (2.1)$$

where $\sigma_{J_r} = \sqrt{\sigma_f^2 + \sigma_r^2}$, $C_0 = -\infty$, $C_1 = C$, $C_2 = \infty$, and where I denotes the old or new item distribution. For hits, $Z = \text{"old"}$, $j = 2$ and $I = \text{old}$, and for misses, the value of $Z = \text{"old"}$, $j = 1$ and $I = \text{old}$. For the DPSD1 model, the expected identification RTs for misses were also given by Equation 2.1 where $Z = \text{"old"}$, $j = 1$ and $I = \text{old}$. Expected identification RTs for hits in the DPSD1 model were calculated with the function

$$E[RT | Z = \text{"old"}, I = \text{old}] = \frac{[1 - R][1 - \Phi(C - \mu_I)]\lambda(Z = \text{"old"}, I = \text{old}) + R(b - s\mu_I)}{[1 - R][1 - \Phi(C - \mu_I)] + R} \quad (2.2)$$

where λ is Equation 2.1. The $M - H$ measure was then calculated for each set of parameters using these expected values. 2D density plots of this measure against values of the SS model's σ_o and the DPSD1 model's R parameters can be found in Figure 2.1.

To confirm each model's predictions, linear models predicting $M - H$ with relevant parameter values were fitted. For the SS model, values of σ_o reliably predicted $M - H$, $R^2 = .26$, $F(1, 99998) = 34242.03$, $p < .001$, $BF = 5.65 \times 10^{6391}$. A model including both σ_o and μ_o as coefficients gave a slightly stronger positive linear association with $M - H$, $R^2 = .27$, $F(2, 99997) = 18907.14$, $p < .001$, $BF = 6.54 \times 10^{6959}$. Adding values of μ_o to this model reliably improved its fit, $BF = 1.16 \times 10^{568}$. This model predicted that $M - H$ should increase by 303 ms with each unit increase of σ_o and that $M - H$ should also increase by 37 ms with each unit increase in μ_o within the parameter space studied. These results demonstrate that as old item variance in the SS model increases, so does the predicted difference between mean identification measures for hits and misses. Furthermore,

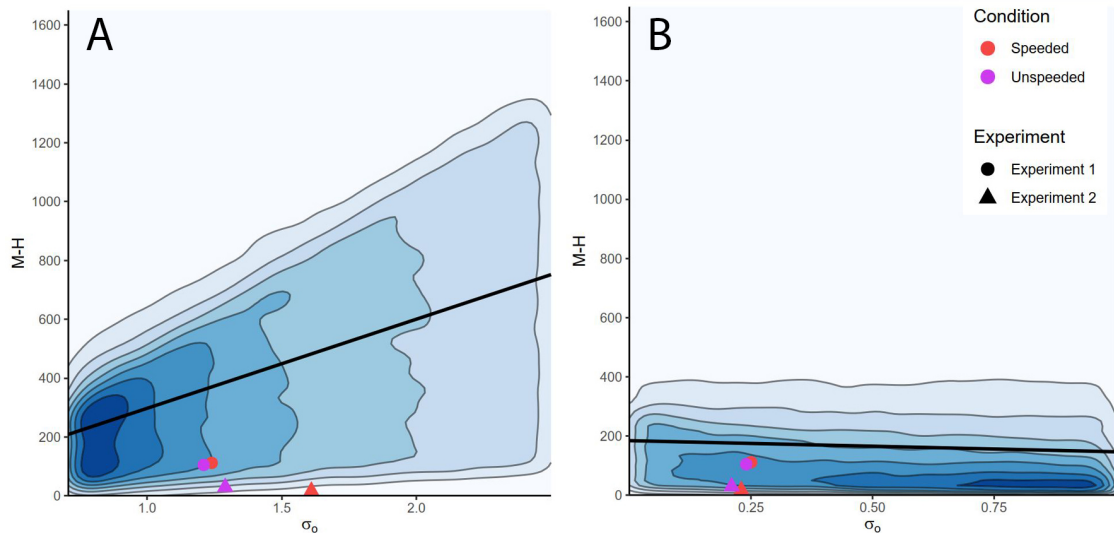


Figure 2.1: Density plots of $M - H$ against parameters σ_o and R , with data points representing the observed values of each RT measure and parameter estimate from Experiments 1 and 2. In each plot, darker colours indicate a greater density and lighter colours indicate a lower density. Data in Panel A was generated using the SS model, and Panel B was generated using the DPSD1 model.

this relationship is enhanced by the addition of mean strength as a predictor in the linear model, although the change when adding this predictor R^2 was small. This confirms that if estimates of old item variance scale and mean memory strength increase, the $M - H$ difference will also increase.

In the DPSD1 model, R was a significant predictor of $M - H$, $R^2 = .01$, $F(1, 99998) = 760.90$, $p < .001$, $BF = 2.70 \times 10^{162}$. The model predicts a decrease of 37 ms in $M - H$ following a unit increase in R . These results show that the DPSD1 model predicts a decrease in $M - H$ as the probability of recollection increases. From the R^2 value and the coefficient of this model, it is likely that this effect is very small in real terms and may not even be detected in a typically powered experiment. Regardless, this prediction is still qualitatively different from that of the SS model, which predicts an increase in $M - H$ in response to the same experimental conditions. Given a common experimental method that elicits both of these predictions, they can be evaluated against the trend observed in an experimental dataset, providing evidence for one of the two models.

2.2 Manipulations of Recollection

Given that these different model behaviours depend upon levels of recollection in the DPSD1 model and corresponding parameter estimates in the SS model, it is logical to investigate manipulations of recollection in recognition memory. There have been many attempts to experimentally dissociate the contributions of recollection and familiarity to recognition memory (for review, see Yonelinas,

2002; Yonelinas et al., 2022). For the purpose of testing the present predictions, it is also worth considering which of these experimental manipulations fit well into a CID-R procedure. One option would be to focus on manipulations of variables during encoding, as these would be less likely to confound consecutive measurements of recognition and priming at test. However, encoding manipulations such as increasing study duration or study elaboration have generally been shown to have similar effects on recollection and familiarity (Yonelinas et al., 2022).

By contrast, some retrieval manipulations are reported to decrease recollection but not familiarity (Yonelinas, 2002). One such manipulation is the act of speeding recognition responses (Koen et al., 2013; Koen & Yonelinas, 2011). Response speeding refers to giving the participants only a short duration during which to make their recognition response, rather than an unlimited response window. For example, Koen et al. (2013) gave participants in their speeded response condition 1500 ms to make a recognition judgement, after which a buzzer would sound prompting them to make a response if they hadn't already. Participants in their unspeeded condition had no time constraints within which to make their response. Koen et al. (2013) found that estimates of the recollection parameter R in the DPSD model were decreased by response speeding, with no effect on estimates of familiarity. Dual process theorists have stated that this selective effect is because familiarity is a fast process, and so its contribution to recognition memory is preserved even within a tight response deadline (Yonelinas, 2002). The contribution of the recollection process is thought to be relatively slow by contrast, and so is lessened by forcing fast responses.

Response speeding is also more compatible with the CID-R paradigm than other retrieval manipulations thought to have similar selective effects on recollection and familiarity. For example, Koen et al. (2013) had participants study objects presented against background scenes. They showed that context reinstatement of background scenes at retrieval increases estimates of recollection and not familiarity for studied objects. However, manipulating backgrounds during a CID-R test could also affect identification performance, confounding the conjoint measure of recognition and priming. Similarly, although Koen et al. (2013) found the same effect on recollection by dividing attention with a concurrent task during retrieval, this would also interfere with the identification task preceding each recognition judgement. By contrast, response speeding would not alter the characteristics of the stimuli within the identification component of the CID-R task, making it preferable to attempt within a CID-R test phase.

In the following experiments, participants completed a CID-R paradigm with two conditions. In an experimental condition, recognition responses were restricted to a 1500 ms response window

compared to a control condition where participants had an unlimited recognition response window. Under the expectation that response speeding impairs recollection and leaves familiarity intact, this manipulation was intended to elicit the previously identified predictions about $M - H$ from the SS and DPSD1 models. Specifically, the SS model predicts that $M - H$ will be greater in the unspeeded than the speeded condition, alongside greater estimates of its parameters σ_0 and μ_0 . The DPSD1 model predicts the opposite; that the RT index will be the same or slightly greater in the speeded condition, given greater values of R . Whichever direction the $M - H$ takes in the data will therefore give support to one of the two models.

2.3 Experiment 1

2.3.1 Method

Participants

40 Psychology undergraduates (4 males, 34 females, 2 other/did not say) with a mean age of 22.50 ($SD = 7.07$) from the University of Plymouth took part in this experiment. They participated in return for points that contributed to a pass/fail course component. The participants were all native English speakers. The sample size allowed the detection of a minimum effect size $d_z = 0.45$ with 80% power in a within-subjects t -test. This effect size is relatively conservative compared with previous manipulations of recollection by recognition response speeding, such as Koen et al. (2013) who reported an effect size of $d_z = 0.99$.

Materials

The stimuli were 240 five-letter English nouns selected from the SUBTLEX-UK database (Van Heuven et al., 2014). Words were presented in 40px white monospaced font against a black background. The experiment was created using the OSWeb functionality of OpenSesame (Mathôt et al., 2012) and hosted on a JATOS server (Lange et al., 2015). Due to the experiment being conducted online, participants used their own hardware to take part. Only laptops and desktop computers were permitted, with Chromebooks, tablets, and mobile devices being disallowed.

Procedure

The order of each within-subjects condition was randomised to reduce the possibility of order effects. Participants completed the experiment while in a Zoom call with the experimenter.

After giving informed consent, participants completed eight practice trials where they acclimatised to the CID procedure. In each CID trial, participants first saw a mask “#####” presented for 500 ms, after which they viewed alternating presentations of a word stimulus and a mask. Each of these word/mask presentation blocks was 250 ms in length, with the word being presented at first for 16.67 ms (one monitor refresh rate at 60 Hz) and the mask for the remainder of the 250 ms block duration. Every two blocks, the duration of the stimulus would be incremented by 16.67 ms, with the mask duration being decremented by the same amount. This pattern of presentation repeated until the stimulus made up the entire duration of the presentation block. During this presentation routine, the participant was instructed to press the Space key as soon as they could identify the word being presented. After doing so, they were asked to type the word into a text box in the centre of the screen and submit their response with the Space key, at which point the next trial would begin. If they did not respond in time, a screen instructing them to respond faster was shown for 2000 ms before the next trial began.

After the practice trials, they began the first experimental condition. The speeded and unspeeded conditions each consisted of a study phase, a brief interval, and a test phase. Both study phases were 60 trials long; in each trial, a fixation point (“+”) was presented in the centre of the screen for 500 ms, followed by a word for 2500 ms, and then a blank screen for 500 ms. Participants were instructed to pay attention to each word, as their memory of these words would be tested later. After the study phase, participants viewed a 60-second countdown timer in a retention interval before viewing the test phase instructions.

In the test phase, participants made responses to all 60 studied words, randomly intermixed with 60 new words in a total of 120 trials. At the beginning of each test trial, participants completed a CID procedure identical to that in the practice phase. After submitting their CID response, the test word was presented in the centre of the screen and participants were asked to judge whether the test word was old or new. Participants were instructed to press the “f” key to respond “New” and the “j” key to respond “Old”; this response key was presented as a static prompt below the title “New or Old?” near the bottom of the screen. In the speeded condition, participants had a 1500 ms window in which to make their recognition response. If they did not respond in this window, they viewed a message for 2000 ms that notified them of this and instructed them to respond faster

in successive trials, before starting the next trial. In the unspeeeded condition, participants were given an unlimited recognition response window. Following a successful recognition response, participants were asked to rate their level of confidence in their recognition response on a 3-point scale: “1 = Low Confidence, 2 = Moderate Confidence, 3 = High Confidence”. They pressed the number key that corresponded to their judgement, at which point the next test trial began. After completing both conditions, participants logged their age and gender in the experimental program before being debriefed.

2.3.2 Results

Five participants were excluded from all the following analyses. One did not make any miss responses during the unspeeeded condition, and as such key measures could not be calculated from their data. One other participant’s data produced very large outlying parameter estimates from the SS model in the unspeeeded condition. Three other participants’ data produced similar outlying parameter estimates from the DPSD1 model in one of the two conditions. Before model fitting and analysis, trials where identification RTs were either greater than 3.5 standard deviations above the sample mean, less than 200 ms, or where responses were missing as a result of the speeding manipulation were excluded. Trials with incorrect identification responses were also excluded. Together, these cases made up 3.79% of all trials.

Manipulation Check

To verify that our recognition speeding manipulation increased the speed of binary recognition responses, we compared mean recognition response RTs in each condition. As expected, mean recognition RTs in the speeded condition ($M = 447.91$, $SE = 5.19$) were faster than those in the unspeeeded condition ($M = 1201.95$, $SE = 26.22$), $t(35) = -5.22$, $p < .001$, 95% CI [-1040.31, -457.82], $BF = 2336.72$. This confirmed that restricting the response window forced participants to make faster recognition judgements.

Task Performance

To evaluate participants’ recognition memory and priming effects, I tested measures from both memory tasks. Mean hit and false alarm rates and mean identification RTs for each item type from each condition can be found in Table 2.2. A 2×2 within-subjects ANOVA on the hit and false alarm rates was conducted, with factors Response Type (hit, false alarm) and Condition (speeded,

Table 2.2: Mean hit rates, mean false alarm rates and mean identification RTs for each item type from Experiments 1 and 2 (SE in parentheses).

Measure	Experiment 1		Experiment 2	
	Speeded	Unspeeded	Speeded	Unspeeded
$P(H)$	0.66 (0.02)	0.69 (0.02)	0.67 (0.02)	0.64 (0.02)
$P(FA)$	0.31 (0.02)	0.34 (0.03)	0.51 (0.02)	0.51 (0.03)
$RT(H)$	2119.82 (75.48)	2101.36 (69.51)	1781.12 (80.07)	1744.34 (66.16)
$RT(M)$	2256.16 (88.52)	2240.13 (80.63)	1820.06 (87.48)	1800.45 (74.39)
$RT(FA)$	2248.02 (91.50)	2200.51 (76.35)	1885.11 (85.24)	1892.52 (71.25)
$RT(CR)$	2335.64 (86.44)	2311.91 (80.03)	1882.84 (91.77)	1877.98 (71.65)

unspeeded). There was a marginally significant main effect of Condition, although Bayesian analyses indicated evidence for the null, $F(1, 34) = 4.48$, $p = .04$, $\eta_p^2 = .12$, $BF = 0.24$. This suggests that there was mixed evidence for a difference between hit and false alarm rates between conditions. There was a reliable main effect of Response Type, $F(1, 34) = 110.07$, $p < .001$, $\eta_p^2 = .76$, $BF = 1.07 \times 10^{29}$, suggesting a much greater proportion of hits than false alarms in each condition. There was no reliable interaction, $F(1, 34) = 0.01$, $p = .91$, $\eta_p^2 < .001$, $BF = 0.36$.

Measures of d' were calculated for each participant in each condition using the formula $d' = \Phi[P(H)] - \Phi[P(FA)]$ where Φ is the inverse cumulative normal distribution function and $P(H)$ and $P(FA)$ are the corrected hit and false alarm rates using the correction given by Snodgrass & Corwin (1988). These corrected proportions were $P(H) = \frac{(|H|+0.5)}{(|Old|+1)}$ for hits and $P(FA) = \frac{(|FA|+0.5)}{(|New|+1)}$ for false alarms. One-sample t -tests showed that d' was significantly greater than zero in both the speeded condition ($M = 1.00$, $SE = 0.11$), $t(34) = 9.09$, $p < .001$, 95% CI [0.77, 1.22], $BF = 81632388$, $d_z = 1.54$, and in the unspeeded condition ($M = 1.02$, $SE = 0.14$), $t(34) = 7.50$, $p < .001$, 95% CI [0.75, 1.30], $BF = 1249552$, $d_z = 1.27$. This indicates that recognition memory performance was significantly above chance in both conditions, showing that participants could reliably distinguish between old and new items.

Regarding priming, we tested for differences between the mean identification RTs for old and new items. In the speeded condition, RTs for old items were faster than those for new items, $t(34) = 6.06$, $p < .001$, 95% CI [97.72, 196.37], $BF = 23449.83$, $d_z = 1.02$. This difference was found in the unspeeded condition, $t(34) = 5.88$, $p < .001$, 95% CI [96.34, 198.24], $BF = 14078.87$, $d_z = 0.99$. This means that there was a priming effect in both conditions. There was no difference between the priming effect in the speeded condition ($M = 142.08$, $SE = 24.10$) and the unspeeded condition ($M = 143.14$, $SE = 24.72$), $t(34) = -0.01$, $p = .99$, 95% CI [-68.50, 68.01], $BF = 0.18$, $d_z < -0.01$.

Parameter Estimates

The SS and DPSD1 models were fitted to the data (see Appendix A for a detailed description of the model fitting procedure and likelihood functions for each model). To test whether key parameter estimates from the SS and DPSD1 models were affected by our response speeding manipulation, we conducted t -tests on these estimates. Mean parameter estimates from each model in each condition are found in Table 2.3. SS model estimates of μ_0 did not differ between speeded and unspeeded conditions, $t(34) = -0.01$, $p = .99$, 95% CI [-0.44, 0.43], BF = 0.18, $d_z < -0.01$. Neither did estimates of σ_0 , $t(34) = 0.77$, $p = 0.45$, 95% CI [-0.18, 0.40], BF = 0.24, $d_z = 0.13$. The DPSD1 model's estimates of μ_0 also did not differ between conditions, $t(34) = -0.10$, $p = 0.92$, 95% CI [-0.21, 0.19], BF = 0.18, $d_z = -0.01$. Likewise, estimates of R did not differ between conditions, $t(34) = -0.13$, $p = 0.90$, 95% CI [-0.10, 0.09], BF = 0.18, $d_z = -0.02$. These results indicate that although our response speeding manipulation was successful in making recognition responses faster, it did not have any effect on the mean or variance of memory strength in the SS model, nor recollection and familiarity in the DPSD1 model.

Identification Reaction Time Analysis

In each condition, t -tests were conducted to determine whether there were differences between RTs for misses and hits. There was a significant difference in the speeded condition, $t(34) = 4.15$, $p < .001$, 95% CI [69.64, 203.05], BF = 129.47, $d_z = 0.70$, and the unspeeded condition, $t(34) = 3.80$, $p = .001$, 95% CI [64.58, 212.97], BF = 52.32, $d_z = 0.64$. The $M - H$ difference was then calculated per participant, per condition. Following from the null effects of our response speeding manipulation on parameter estimates, $M - H$ did not differ between the speeded ($M = 136.34$, $SE = 32.82$) and unspeeded ($M = 138.77$, $SE = 36.51$) conditions, $t(34) = -0.05$, $p = .96$, 95% CI [-98.72, 93.87], BF = 0.18, $d_z = -0.01$.

Goodness of Fit Comparisons

As an exploratory analysis, we compared the quantitative fit of each model to the observed data. As specified here, the SS and DPSD1 models both have 10 free parameters. However, for parity with previous research, we compared the models with the $AIC = -2\ln(L) + 2p$, which accounts for both a model's logarithmic likelihood value $\ln(L)$ and its number of free parameters p . AICs were calculated for each SS and DPSD1 model fit to each participant's data in each condition; the

Table 2.3: Mean parameter estimates for the SS and DPSDI models in Experiments 1 and 2 (SDs in parentheses)

Model, Parameter	Experiment 1		Experiment 2	
	Speeded	Unspeeded	Speeded	Unspeeded
SS				
μ_0	1.36 (1.13)	1.36 (1.21)	1.19 (1.45)	1.12 (1.36)
σ_0	1.31 (0.84)	1.20 (0.55)	1.46 (1.43)	1.29 (0.86)
σ_p	621.13 (119.39)	658.43 (142.56)	497.18 (198.55)	496.45 (174.66)
b	2295.11 (492.19)	2286.14 (454.89)	1886.43 (531.00)	1866.37 (433.29)
s	99.44 (81.04)	115.02 (95.03)	245.36 (366.01)	164.48 (220.55)
C_1	-1.07 (0.79)	-1.09 (0.71)	-2.03 (2.21)	-1.70 (1.85)
C_2	0.12 (0.57)	0.02 (0.59)	-0.44 (1.61)	-0.23 (0.97)
C_3	0.58 (0.49)	0.50 (0.56)	0.32 (0.37)	0.41 (0.40)
C_4	0.89 (0.53)	0.89 (0.67)	0.50 (0.46)	0.62 (0.52)
C_5	1.59 (0.54)	1.64 (0.67)	0.98 (0.47)	1.14 (0.65)
DPSDI				
μ_0	0.64 (0.52)	0.65 (0.62)	0.42 (0.44)	0.48 (0.39)
R	0.26 (0.21)	0.26 (0.23)	0.24 (0.25)	0.21 (0.25)
σ_p	616.67 (116.74)	656.27 (142.98)	569.03 (181.13)	540.50 (148.04)
b	2288.03 (490.27)	2273.46 (449.00)	1863.10 (519.34)	1843.62 (437.18)
s	149.37 (136.43)	151.55 (130.47)	91.71 (117.48)	79.20 (82.85)
C_1	-0.99 (0.70)	-1.04 (0.66)	-1.60 (1.37)	-1.54 (1.67)
C_2	0.11 (0.55)	0.01 (0.57)	-0.19 (0.65)	-0.16 (0.98)
C_3	0.55 (0.46)	0.46 (0.51)	0.36 (0.40)	0.45 (0.38)
C_4	0.85 (0.50)	0.85 (0.61)	0.54 (0.45)	0.67 (0.49)
C_5	2.06 (1.96)	2.24 (2.95)	1.05 (0.43)	2.29 (6.83)

model with the minimum AIC was selected as the best-fitting model. In the speeded condition, the SS model gave the best fit in 55.56% of cases, whereas the DPSD1 model gave the best fit in the remaining 44.44%. In the unspeeded condition, the SS model fitted best in 72.22% of cases, with the DPSD1 model fitting best in the remaining cases (27.78%).

To determine the extent that each model fitted best on a participant level, AIC differences were calculated for each model (Burnham & Anderson, 1998). For a given model i , the AIC difference was calculated as $\Delta_i = \text{AIC}_i - \text{AIC}_{\min}$. In the speeded condition, the mean of Δ_{SS} ($M = 0.93$) did not significantly differ from the mean Δ_{DPSD1} ($M = 1.26$) in a Wilcoxon signed rank test¹, $V = 263$, $p = .40$. In the unspeeded condition, there was also no significant difference between Δ_{SS} ($M = 0.45$) and Δ_{DPSD1} ($M = 1.12$), $V = 213$, $p = .10$. It is also of note that in both conditions, at least 77% of AIC differences for both models were less than or equal to two, meaning that both models had strong support for their quantitative fit, relative to one another (Burnham & Anderson, 1998). Indeed, across both conditions, only one fit from the DPSD1 model had an AIC difference greater than 10, indicating essentially no support for the model; the SS model had no such cases. This shows that both models gave a roughly equivalent quantitative fit, despite the SS model fitting best to the majority of participant-level datasets.

Exploratory Reaction Time Comparisons

Since the primary experimental manipulation was not successful in eliciting the unique RT predictions simulated prior to this experiment, exploratory analyses were conducted. Berry et al. (2012) investigated three other predictions about mean RTs associated with certain recognition responses. Specifically, the SS model predicts that $RT(\text{FA}) < RT(\text{CR})$, $RT(\text{M}) < RT(\text{CR})$, and $RT(\text{New}) - RT(\text{Old}) > RT(\text{CR}) - RT(\text{M})$. These predictions were assessed in each condition of the present experiment with t -tests.

In the speeded condition, $RT(\text{FA})$ was marginally faster than $RT(\text{CR})$, $t(34) = 2.06$, $p = .05$, 95% CI [1.04, 174.21], BF = 1.18, $d_z = 0.35$. A significant difference was observed in the unspeeded condition, $t(34) = 2.48$, $p = .01$, 95% CI [20.19, 202.61], BF = 2.58, $d_z = 0.42$. In the speeded condition, values of $RT(\text{M})$ were significantly lower than values of $RT(\text{CR})$, despite inconclusive Bayesian evidence, $t(34) = 2.30$, $p = .03$, 95% CI [9.10, 149.86], BF = 1.80, $d_z = 0.39$. Values of $RT(\text{M})$ were marginally lower than those of $RT(\text{CR})$ in the unspeeded condition, however, the Bayes Factor for this test was also inconclusive $t(34) = 2.02$, $p = .05$, 95% CI [-0.56, 144.13], BF =

¹A non-parametric test was used because of positively skewed distributions of Δ_i .

1.10, $d_z = 0.34$. Finally, in the speeded condition, the overall priming effects $RT(\text{New}) - RT(\text{Old})$ ($M = 147.04$, $SE = 24.27$) were reliably greater than $RT(\text{CR}) - RT(\text{M})$, the priming effects for new items ($M = 79.48$, $SE = 34.63$), $t(34) = 2.68$, $p = .01$, 95% CI [16.32, 118.81], $BF = 3.86$, $d_z = 0.45$. The same was also true of $RT(\text{New}) - RT(\text{Old})$ ($M = 147.29$, $SE = 25.07$) and $RT(\text{CR}) - RT(\text{M})$ ($M = 71.78$, $SE = 35.60$) in the unspeeded condition, $t(34) = 2.48$, $p = .02$, 95% CI [13.67, 137.34], $BF = 2.58$, $d_z = 0.42$, although the Bayes Factor for this comparison was inconclusive.

2.3.3 Discussion

Although manipulating the time available for participants to make recognition judgements successfully shortened recognition decision RTs, estimates of recollection or equivalent parameters did not change in either model. Estimates of R in the DPSD1 model did not reliably differ between speeded and unspeeded conditions, and neither did estimates of σ_o or μ_o , with Bayesian analyses giving strong evidence for the null in each comparison. Accordingly, $M - H$ did not differ between conditions either. This means that the SS and DPSD1 models cannot be discriminated in this experiment on the basis of their predictions about the $M - H$ difference. Both models gave an equivalent quality of quantitative fit to the data too, meaning one cannot be favoured over the other on this basis either.

Exploratory analyses revealed mixed support for the SS model. Its prediction that $RT(\text{FA}) < RT(\text{CR})$ was observed, but in the absence of conclusive Bayesian evidence. There was some evidence that $RT(\text{M}) < RT(\text{CR})$ and $RT(\text{New}) - RT(\text{Old}) > RT(\text{CR}) - RT(\text{M})$; significant differences between these RT indices were found in both conditions, however, the Bayes Factors for these comparisons were also inconclusive, save for one which supported the alternative hypothesis. This demonstrates partial evidence for model predictions previously shown to demonstrate the SS model's superiority over a strict multiple systems model (Berry et al., 2012). However, these results also do not discriminate between the SS and DPSD1 models, as the latter can also make these predictions.

The null effect of response speeding on the model parameters contrasts the results of previous research where this manipulation lowered estimates of recollection (Yonelinas, 2002; Koen et al., 2013; Koen & Yonelinas, 2011). However, these previous results were observed studying recognition memory alone, without an added identification task in a CID-R procedure. Although the recognition response window was shortened successfully, it is possible that the effect of this ma-

nipulation on recollection was confounded by the presence of the identification task immediately beforehand in each test trial. That is, participants might have covertly decided their recognition response as they identified each item, rather than only using information about the stimulus obtained in the recognition response window. This is particularly problematic because identification RTs vary, meaning that the time between the stimulus onset in the identification task and the following recognition judgement may not have been controlled. If participants made covert recognition decisions during the identification window, estimates of recollection and familiarity would not be expected to differ between conditions. Experiment 2 aimed to resolve this confound.

2.4 Experiment 2

To address the problem of a variable identification response window impacting the response speeding manipulation, the identification and recognition components of the CID-R procedure can be presented to participants in separate blocks of trials. Rather than each test trial prompting an identification response immediately followed by a recognition decision, participants could complete a block of identification test trials, followed by a separate block of recognition test trials. Although the identification phase might still provide cues that aid in the later recognition of stimuli, the separation of these phases eliminates the variable time interval directly preceding each recognition judgement. This means that participants will be forced to use only the recognition response window to reinstate their memory for each item, and will not be able to make covert judgements immediately before each recognition trial.

In the present experiment, participants judged three types of items across the blocked identification and recognition test phases. In the identification phase, participants judged all 80 old items they had studied in the preceding study phase, intermixed with 60 items they had not yet seen. These *partially new* items were included for several reasons. To measure the priming effect, the mean identification RT for old items must be subtracted from that for new items, and so new items must appear in the identification phase. The presence of new items is also necessary to get identification measures for all four recognition response types. Finally, it is important that participants make their recognition decision on the basis of memory from the study phase. If partially new items were not included during identification, participants could feasibly make their recognition judgements on the basis of item repetition or familiarity with the old items from the identification phase. To ensure that some diagnostically new items remained in the recognition test phase, 20 *completely new* items that did not appear in any other part of the experiment were presented in this phase. This

was intended to help old and new items remain reliably discriminable, so as to avoid floor effects that might hinder the ability to observe differences in recognition memory parameters. This also allowed for equal numbers of old and new items in each condition and ensured that identification RTs were still collected for most new items.

To fit this new design, the SS and DPSD1 models were adapted, with the mean of the partially new item memory strength distribution, μ_n , being a free parameter. This reflected the fact that partially new items had been seen twice by the recognition test phase, and so could accumulate additional memory strength. Implementing this design and model extension, the hypotheses remained the same as those in Experiment 1.

2.4.1 Method

Participants

40 participants (17 males, 22 females, 1 other/non-binary) with a mean age of 22.95 ($SD = 4.89$) took part in this experiment. Participants were either undergraduates from the University of Plymouth participating in return for participation points or members of the public participating in return for Amazon vouchers worth £7.50. All participants were native English speakers who had not taken part in Experiment 1. The sample size was justified by the same power calculation as in Experiment 1.

Materials and Procedure

The stimuli were 320 five-letter English nouns from the SUBTLEX-UK database (Van Heuven et al., 2014). The rest of the apparatus and materials were the same as in Experiment 1. As in Experiment 1, the order of each condition was randomised across participants, and all participants completed the study while on a Zoom call with the experimenter. Participants first completed a practice phase identical to that in Experiment 1 to acclimatise them to the CID procedure. In each subsequent condition, participants first viewed 80 old words in a study phase, with procedures identical to those in Experiment 1. Afterwards, participants completed a CID phase in which they identified the 80 old words that appeared during the most recent study phase, intermixed with 60 partially new words that did not appear in the study phase. The trial-level procedure in this CID phase was identical to that of the practice phase.

After the CID phase, participants completed a recognition memory test. Participants judged 80 old

words, 60 partially new words (those that first appeared in the CID phase) and 20 completely new words intermixed in a random trial order. This mixture of partially new and completely new words during the test phase was intended to keep the ratio of items requiring old and new decisions equal and helped to ensure that “old” judgements were made on the basis of an item’s presentation in the study phase. In each trial, participants saw a word and were asked to make a binary recognition memory judgement. They responded "j" for “old” if they judged an item to have appeared in the study phase, and "f" for “new” if not, encompassing both the partially new and completely new words. They had 1500 ms to do this in the speeded condition, and an unlimited response window in the unspeeded condition. If no judgement was made within the response window in the speeded condition, a prompt was displayed for 2000 ms encouraging a faster response in the next trial, after which the next trial began. Upon a successful binary recognition judgement, participants were then asked to rate their level of confidence in their judgement on the same 3-point scale as in Experiment 1, after which point the next test trial would begin. After completing both conditions, the participants were asked their age and gender before being debriefed.

2.4.2 Results

Four participants were excluded from the following analyses due to having large outlying parameter estimates from the SS and DPSD1 models. Identification RTs were also subject to the same exclusion criteria as in Experiment 1; 4.15% of trials were excluded on this basis.

Manipulation Checks

As in Experiment 1, we compared mean recognition response times in the speeded and unspeeded conditions to determine whether our manipulation of recognition response time was successful. Due to a technical fault in the experiment program, recognition response times were only recorded for 13 out of 36 participants. A *t*-test showed a marginally significant difference between recognition response times in the speeded condition ($M = 872.95$, $SE = 31.14$) and those in the unspeeded condition ($M = 1189.17$, $SE = 182.05$), $t(12) = -1.79$, $p = .10$, 95% CI [-700.96, 68.51], $BF = 0.97$, $d_z = -0.50$. This means that recognition response times were shorter in the speeded condition, showing that the manipulation was effective.

Task Performance

To examine recognition performance in terms of hit and false alarm rates, a 2 (condition) \times 3 (response type) within-subjects ANOVA was conducted with a Greenhouse-Geisser sphericity correction. No significant main effect of condition was found, $F(1, 35) = 0.83$, $p = .37$, $\eta_p^2 = .02$, $BF = 0.17$. There was a significant main effect of response type, $F(1.37, 47.93) = 57.06$, $p < .001$, $\eta_p^2 = .62$, $BF = 8.76 \times 10^{26}$. No interaction was observed, $F(1.98, 69.24) = 0.55$, $p = 0.58$, $\eta_p^2 = .02$, $BF = 0.10$. Bonferroni-corrected pairwise comparisons revealed that false alarm rates for new items were significantly lower than those for partially new items, $t(35) = -7.67$, $p < .001$, and significantly lower than hit rates $t(35) = -8.16$, $p < .001$. False alarm rates for partially new items were also reliably lower than hit rates, $t(35) = -3.98$, $p = .001$. This means that partially new items were more likely to be incorrectly judged "old" than new items, likely due to their appearance in the identification phase before the recognition test phase. However, a greater proportion of old items were judged old than partially new items, indicating participants were able to discriminate between these item classes.

A one-sample t -test on estimates of d' derived from the experimental data revealed a significant difference from 0 in the speeded condition ($M = 0.45$, $SE = 0.07$), $t(35) = 6.23$, $p < .001$, 95% CI [0.30, 0.59], $BF = 42128.01$, $d_z = 1.04$. This difference was also observed in the unspeeded condition ($M = 0.37$, $SE = 0.08$), $t(35) = 4.49$, $p < .001$, 95% CI [0.20, 0.54], $BF = 323.47$, $d_z = 0.75$. We also tested for a priming effect by comparing the mean identification RTs for old and partially new items. The mean identification RT for old items ($M = 1800.922$, $SE = 80.21$) was faster than that for partially new items ($M = 1899.99$, $SE = 84.92$) in the speeded condition, $t(35) = 4.03$, $p < .001$, 95% CI [49.06, 148.74], $BF = 95.92$, $d_z = 0.67$. In the unspeeded condition, there was also a significant difference between the mean identification RT for old ($M = 1762.67$, $SE = 68.52$) and partially new ($M = 1892.73$, $SE = 68.83$) items, $t(35) = 7.01$, $p < .001$, 95% CI [91.15, 165.50], $BF = 373701.50$, $d_z = 1.17$.

Parameter Estimates

We conducted t -tests to test whether or not our manipulation had any impact on key parameter estimates from the SS and DPSD1 models. There was no effect of response speeding on the SS model's estimates of μ_0 , $t(35) = 0.47$, $p = .64$, 95% CI [-0.24, 0.38], $BF = 0.20$, $d_z = 0.08$, nor on estimates of σ_0 , $t(35) = 0.90$, $p = .38$, 95% CI [-0.22, 0.57], $BF = 0.26$, $d_z = 0.15$. Similarly, there was no effect of response speeding on the DPSD1's estimates of μ_0 , $t(35) = -0.92$, $p = .36$,

95% CI [-0.20, 0.07], BF = 0.27, $d_z = -0.15$. There was also no effect of response speeding on R , $t(35) = 0.91$, $p = .37$, 95% CI [-0.04, 0.10], BF = 0.26, $d_z = 0.15$. This means that, as in Experiment 1, our response speeding manipulation did not have an effect on the characteristics of memory strength in the SS model, nor recollection or familiarity processes in the DPSD1 model.

Identification Reaction Time Analysis

As in Experiment 1, t -tests were conducted on RTs for misses and hits within each condition. No reliable difference was found in the speeded condition $t(35) = -1.40$, $p = .17$, 95% CI [-26.43, 4.85], BF = 0.44, $d_z = -0.23$. No significant difference was observed in the unspeeded condition either, $t(35) = -1.12$, $p = .27$, 95% CI [-21.17, 6.13], BF = 0.32, $d_z = -0.19$. $M - H$ was then calculated for each participant and compared between conditions. No significant difference between $M - H$ in the speeded condition ($M = -10.79$, $SE = 7.70$) and the unspeeded condition ($M = -7.52$, $SE = 6.72$) was found, with Bayesian evidence for the null, $t(35) = -0.31$, $p = .76$, 95% CI [-24.62, 18.07], BF = 0.19, $d_z = -0.05$.

Goodness of Fit Comparisons

AICs were used to compare the fit of the SS and DPSD1 models to participant data from both conditions. In the speeded condition, the SS model fitted best to 77.78% of cases, with the DPSD1 model fitting best to the remaining 22.22%. In the unspeeded condition, the SS model fitted best to 75.00% of cases, with the DPSD1 model fitting best to the remaining 25.00%. This shows that the SS model provided the best quantitative fit to participant data in the majority of cases, in both conditions. Analyses of the AIC differences in the speeded condition revealed that Δ_{SS} ($M = 0.38$) was significantly lower than Δ_{DPSD1} ($M = 1.79$), $t(35) = -2.80$, $p = .008$, 95% CI [-2.43, -0.39], BF = 4.96, $d_z = -0.47$. Δ_{SS} ($M = 0.32$) was also significantly lower than Δ_{DPSD1} ($M = 2.04$) in the unspeeded condition, $t(35) = -3.86$, $p < .001$, 95% CI [-2.63, -0.82], BF = 62.41, $d_z = -0.64$. The AIC difference was greater than 10 for one DPSD1 model fit, and for no SS model fits, indicating that neither model should be rejected purely on the grounds of poor quantitative fit relative to the other model (Burnham & Anderson, 1998). However, these analyses show that the SS model fitted the data reliably better than the DPSD1 model.

Exploratory Reaction Time Comparisons

The same RT comparisons assessed in Experiment 1 were also tested in the present experiment as an exploratory analysis. Values of $RT(FA)$ did not significantly differ from values of $RT(CR)$ in the speeded condition, $t(35) = 0.68$, $p = .50$, 95% CI [-11.02, 22.00], $BF = 0.22$, $d_z = 0.11$. This was also the case in the unspeeded condition, $t(35) = -1.92$, $p = .06$, 95% CI [-32.83, 0.84], $BF = 0.94$, $d_z = -0.32$. Values of $RT(M)$ were significantly lower than $RT(CR)$ in the speeded condition, $t(35) = 1.46$, $p = .048$, 95% CI [0.16, 44.61], $BF = 1.14$, $d_z = 0.34$. However, $RT(M)$ was not reliably faster than $RT(CR)$ in the unspeeded condition, $t(35) = 0.13$, $p = .89$, 95% CI [-15.64, 17.84], $BF = 0.18$, $d_z = 0.02$. The overall priming effect was significantly greater than that of the priming effect for new items in the speeded condition, $t(35) = 3.09$, $p = .004$, 95% CI [26.21, 126.81], $BF = 9.46$, $d_z = 0.51$. This difference was also significant in the unspeeded condition, $t(35) = 6.17$, $p < .001$, 95% CI [85.35, 169.09], $BF = 35355.17$, $d_z = 1.03$. Taken together, these results mostly fail to support the predictions of the SS model identified by Berry et al. (2012).

2.4.3 Discussion

The present experiment used a blocked CID-R design to manipulate recollection through recognition response speeding. However, as in Experiment 1, this manipulation of recollection was unsuccessful. As a result, the relevant parameter estimates from the SS and DPSD1 models were unaffected. Since recognition response times were not recorded for a number of participants, it is not certain whether this was due to the effectiveness of the manipulation in reducing recognition RTs, or another artefact of the design. Although the analysis of 13 out of 36 participants showed reduced recognition RTs in the speeded condition, this was not guaranteed for the other 23 participants, meaning the manipulation may not have been strong enough to affect the recollection process in recognition. However, it is also possible that introducing a continuous identification task at any point before a recognition judgement nullifies the effects of speeding that judgement on recollection. This could be because, regardless of the blocked design, viewing items for an uncontrolled length of time during identification could still have provided participants with information that influenced later recognition decisions. It is hard to gauge the extent of this confounding effect in the present design. However, researchers conducting further work with similar manipulations should take care to avoid contamination between identification and recognition tasks.

In either case, the present experiment was unable to elicit differences in the RT indices of interest and so could not discriminate the SS and DPSD1 models on this basis. The SS model gave a better

quantitative fit to the majority of participant-level data in both conditions when compared to the DPSD1 model. However, many of the exploratory analyses did not confirm the results of Berry et al. (2012) in favour of the SS model, meaning that the present results give only mixed support for it. Although the models are still hard to discriminate based on the present results, the predictions they make about RTs for hits and misses still stand. However, it is clear that a different experimental manipulation would be needed to successfully elicit them.

2.5 General Discussion

There is an ongoing debate as to whether performance in recognition and priming tasks is governed by a single memory system or multiple systems. The SS and DPSD1 models (Berry et al., 2012) formalise these theoretical positions and make opposing predictions about identification RTs for different types of recognition responses while recollection increases. However, the present attempts to manipulate recollection by speeding recognition response times were unsuccessful, and the models could not be discriminated on the basis of their predictions. In Experiment 1, manipulation of the recognition response window in a CID-R test phase was not successful in affecting estimates of recollection from either model, and the difference between mean RTs for misses and hits was unaffected. An adaptation to this method with blocked identification and recognition test phases in Experiment 2 also did not elicit effects upon parameter estimates from either model or the $M - H$ difference. This meant that neither model's prediction about changes in the $M - H$ difference could be validated. Although the models could not be distinguished on this basis, the SS model provided a better quantitative fit for the majority of participants in Experiment 2.

Although the response speeding manipulation did not elicit decreases in recollection as intended, it could be improved for better results in the future. The present choice of a 1500 ms response deadline was based upon previous successful manipulations of recollection (Koen et al., 2013). However, there is evidence that responses associated with neural correlates of recollection can be made faster than this. In an ERP analysis, Park & Donaldson (2016) found that unprimed recollection-related trials had peak activation around 500-800 ms from stimulus onset. Furthermore, they found that priming targets directly before a recognition judgement sped the onset of these recollection-related old/new ERP effects by approximately 300 ms. This provides evidence that neural markers of recollection-based responses could still occur within a short time interval, especially when implicit memory for those items is present. It could be argued that additional time would be necessary for participants to make their recognition response, and so the responses given in the present

speeded conditions could still be expected to be mostly based upon familiarity. However, the mean response times in the unspeeded conditions in Experiments 1 ($M = 1201.95$) and 2 ($M = 1189.17$) were less than the 1500 ms response deadline in the speeded conditions. If responses under 1500 ms are expected to reflect familiarity and not recollection, it is possible that many responses in the unspeeded condition were not influenced by recollection, explaining the lack of reliable difference in R parameters between conditions. Future response speeding manipulations could address this possibility by forcing participants to respond faster in the speeded condition and delay their recognition response for longer than the speeded response deadline (2000 ms, for example) in the unspeeded condition. This would increase the difference between the mean recognition response times in the speeded and unspeeded conditions further, making the responses in each condition more likely to be based on familiarity and recollection, respectively. Such a method may have a better chance of affecting estimates of R in the DPSD1 model and eliciting its unique predictions.

It is also possible that a different manipulation of recollection could be more effective. For instance, since the SS and DPSD1 models specify different combinations of strength sources, a better test of their predictions could be to manipulate recollection with an encoding manipulation. However, levels of familiarity also need to remain unchanged by such a manipulation to elicit the model predictions identified here. Yonelinas (2002) reviewed many variables that have been shown to have separable effects on recollection and familiarity. Each encoding manipulation they reviewed altered recollection but also had some effect on familiarity, sometimes depending on the materials used in a given study. Besides response speeding, Yonelinas (2002) found only one other retrieval manipulation — dividing attention at test — could influence recollection while maintaining a constant level of familiarity (see also Koen et al., 2013). However, integrating this manipulation into the CID-R paradigm could present challenges. Namely, the preceding identification task could reveal information about the stimulus relevant to the recognition decision, confounding a later manipulation of attention during the recognition judgement.

This confound may threaten any successful manipulation of recollection during retrieval in a CID-R task. Indeed, it may have contributed to the null effects observed in Experiments 1 and 2, where each response speeding manipulation did not influence recollection in both sequential and blocked CID-R test phases. Alternatively, a between-subjects comparison could be used to highlight differences in recollection. Yonelinas (2002) identified special populations that experience disrupted recollection, such as older adults and individuals with amnesia or frontal lobe lesions. However, there is also support for the view that comparable explicit and implicit memory deficits result from normal

ageing (Ward et al., 2013a) and amnesia (Jernigan & Ostergaard, 1993; Ostergaard & Jernigan, 1993; Ostergaard, 1999), meaning that familiarity might be subject to a decline in these groups alongside recollection. Between-subjects comparisons also have less statistical power to detect small effects, so a large sample would be needed to detect differences in $M - H$ predicted by either model. Taking these factors together, the use of special populations may not easily improve upon the present attempts to test SS and DPSD1 model predictions about the $M - H$ difference.

There are, however, conceptual limitations of any attempt to selectively target recollection with experimental manipulation. From the perspective of a single-system theory, memory strength is one continuous latent variable that underlies all recognition memory judgements. The qualitative experiences of recollection and familiarity are not individual processes in their own right. Instead, they are emergent properties of experiencing the feeling of remembering items that have different memory strength values. If this theoretical position is true, it is hard to selectively target recollection versus familiarity with an experimental manipulation because neither "system" exists in its own right. The logic of using response speeding to prevent the "slow" contribution of recollection falls apart if recollection is simply a feeling elicited by high memory strength. For this reason, it may not be useful to target recollection or familiarity specifically in future research comparing single-system and dual-process predictions. Instead, it is useful to consider the behaviour of each model in response to a manipulation that has particular relevance to the SS model specification. In the following chapter, I adopt this approach to test the predictions of single-system and dual-process accounts of recognition and priming further using variability in memory strength — a concept represented in both models, but of particular importance to the SS model.

To conclude, I conducted two experiments to investigate opposing predictions made by the SS and DPSD1 models of recognition and priming. Experiment 1 attempted to restrict recognition response speed in a CID-R test to manipulate recollection, which in turn would cause the models to make opposing predictions about identification RTs in the experiment. However, this manipulation was unsuccessful, possibly due to the speeding manipulation being confounded by the preceding identification judgement in each trial. Experiment 2 attempted this manipulation in a blocked CID-R design, however, this was unsuccessful in affecting estimates of recollection as well. Ultimately, the models were not distinguishable based on their identification RT predictions. Although these predictions could still be tested with other methods, focusing on manipulations that do not attempt to selectively change measures of recollection or familiarity will likely be most beneficial for further research.

Chapter 3

Discriminating SS and DPSD1 Models Using Encoding Variability Manipulations

In the seminal dual-process signal detection model of recognition memory (DPSD; Yonelinas, 1994), the distinction between recollection and familiarity is clear. Items are either recollected as the result of an "all-or-none" threshold process, or they are judged on the basis of a continuous familiarity signal. However, as established in the previous chapter, there are few experimental manipulations of recollection that do not also affect familiarity (Yonelinas, 2002). It is harder still to successfully implement such a manipulation within a CID-R procedure. As the results of Chapter 2 showed, manipulations that affect recollection independently of familiarity in a recognition task may be confounded or otherwise affected by the addition of an identification test. This makes it challenging to test the DPSD1 model's (Berry et al., 2012) unique predictions about identification RTs when its R parameter increases independently of the mean of the old item familiarity distribution, μ_o . In this chapter, I explore another manipulation that may elicit opposing predictions for perceptual identification measures in the SS and DPSD1 models; a manipulation of encoding variability.

The model simulations in Chapter 2 confirmed that increases in the SS model's μ_o and σ_o parameters cause the model to predict an increase in $M - H$; the difference in mean identification for miss and hit responses. While both parameters can independently motivate this effect (see Chapter 2, Expected Value Simulations), σ_o has a much greater effect on values of $M - H$. Increases in σ_o result in a greater spread of identification values for items that receive hit and miss responses, leading to an increased difference between the mean RTs for these response types. As the variability of the old and new item strength distributions are fixed to $\sqrt{0.5}$ in the DPSD1 model, it must represent changes in the variability in memory strength for old items by adjusting both μ_o and R . This would result in a different pattern of RTs than that predicted by the SS model. It is therefore

worth investigating manipulations of old item variance in recognition memory that may be able to elicit these opposing predictions in a recognition and priming experiment.

From the first applications of signal detection models to the study of recognition memory, it has been widely accepted that the variance of the old item strength distribution is greater than the variance of the new item distribution (for a review, see Rotello, 2017). The acceptance of this *old item variance effect* is motivated by analyses of the z -ROC, a z -transformed plot of the hit rate against the false alarm rate at each level of recognition confidence in a given response scale. Most z -ROCs calculated from recognition confidence data are approximately linear, with slopes less than 1 (Glanzer et al., 1999). Since the value of the z -ROC slope has long been presumed to represent the ratio σ_o/σ_n in a traditional Gaussian signal detection model (but see Rabe et al., 2021), a non-unit z -ROC slope necessitates making σ_o a free parameter with a value typically greater than σ_n . This assumption was the defining feature of the UVSD model (Egan, 1958).

Although the UVSD model can account for some commonly observed regularities in the z -ROC slope (Egan, 1958; Yonelinas & Parks, 2007), its unequal variance assumption was created purely for the need to account for observed data, and not with a priori psychological assumptions in mind. A complementary psychological explanation for the unequal variance assumption was later proposed in the form of the encoding variability hypothesis (Jang et al., 2012a; Wixted, 2007). According to this theory, the old item variance effect is caused by the presence of a large number of variables that affect memory strength at encoding. These variables contribute additional strength and variance to memory strength across a set of old items during the study phase, resulting in an increase in σ_o relative to σ_n . Examples of such encoding variables could presumably include the level of attention paid to a stimulus, item characteristics, item-participant interaction, and many others. Stated mathematically, old items have some level of baseline strength, $B \sim \mathcal{N}(\mu_{\text{baseline}}, \sigma_{\text{baseline}})$, which is equivalent to the new item strength distribution (Jang et al., 2012a). In the study phase, B is incremented by an added strength variable $A \sim \mathcal{N}(\mu_{\text{added}}, \sigma_{\text{added}})$ during encoding. The addition of baseline and added strength gives the resulting old item distribution in the formula $O = B + A$. The variance of O is therefore greater than that of new items, being equal to $\sigma_B^2 + \sigma_A^2$.

There have been several attempts to test the encoding variability hypothesis and compare its predictions with those of other accounts. Koen & Yonelinas (2010) first attempted this in a method where items at study were presented for either a fixed duration of 2500 ms, or a mixture of 1000 and 4000 ms durations. It was found that the latter variable encoding condition did not change estimates of σ_o . Instead, the contribution of an additional recollection process was solely responsible for

changes to the z -ROC slope, supposedly constituting evidence against the encoding variability hypothesis in favour of a dual-process model. However, subsequent comments by Starns et al. (2012b) and Jang et al. (2012a) clarified that these results had no bearing on the encoding variability hypothesis. This was because Koen & Yonelinas (2010)'s method mixed two discrete levels of encoding strength, which would be expected to result in a mixture strength distribution rather than a Gaussian as the encoding variability hypothesis predicts. However, Koen et al. (2013) later studied the effects of retrieval manipulations on old item variance, finding that it was possible to induce changes in estimates of σ_0 without manipulating encoding variability. Although this finding does not exclude the possibility that encoding variability may still have some role in determining estimates of σ_0 , it suggests that it is not the only factor that influences old item variance.

More recently, Spanton & Berry (2020) attempted to test the encoding variability hypothesis by manipulating encoding variables directly during the study phase. To avoid the creation of mixture strength distributions that confounded the method adopted by Koen & Yonelinas (2010), encoding variables were manipulated by adding variance along a continuous scale, rather than by mixing two separate conditions of high or low-quality encoding. Across three experiments, attempts to influence σ_0 by manipulating three encoding variables (study duration, attention, and word frequency) were unsuccessful. There were no resultant effects on σ_0 , although each independent variable was assessed to have a weak effect on recognition confidence ratings in manipulation checks. Despite this, both d and σ_0 were found to be significantly greater in the low item characteristic variance condition in Experiment 2, suggesting again that changes in σ_0 may result from factors other than encoding variability. Estimates of d and σ_0 also showed strong positive correlations in every experiment, indicating that old item variance may scale with mean strength. Spanton & Berry (2022) found further evidence for a strength scaling trend in their Experiments 1 and 2, as experimental manipulations of mean strength independently contributed to increases the σ_0 parameter in the UVSD model. By contrast, they did not find strong evidence that variability in item characteristics increased σ_0 . Assuming that manipulations in variability in item characteristics translate to manipulations of encoding variability, these results are not predicted by the encoding variability hypothesis.

The idea that mean memory strength and variance in memory strength are related is evidenced elsewhere in the recognition memory literature. Although some previous research concluded that the z -ROC slope takes a constant value of approximately 0.8 (Ratcliff et al., 1992, 1994), it was later found that in many cases, increases in mean strength generally decrease the z -ROC slope (Glanzer et al., 1999; Parks & Yonelinas, 2007), meaning that mean strength and old item variance increase

with one another in several experimental contexts. The finding that greater strength coincides with greater old item variance has since been observed in other studies (Glanzer & Adams, 1990; Heathcote, 2003; Hirshman & Hostetter, 2000; Koen et al., 2013; but see Starns et al., 2012a; Grider & Malmberg, 2008). More recently, Dopkins et al. (2017) found that a semantic priming manipulation increased the memory strength of new items and the variance of their corresponding confidence ratings at test, as well as the z -ROC slope. This suggests that a form of strength and item variance scaling could apply more generally to both old and new item types – a distribution with a greater mean tends to have greater variance. In sum, this is evidence that σ_o scales as a monotonically increasing function of d in many experimental settings.

Despite several attempts to test the encoding variability hypothesis, no conclusive experimental evidence that encoding variability affects estimates of σ_o in the UVSD model exists. There have however been indications that the mean memory strength parameter d increases alongside old item variance in the UVSD model (Spanton & Berry, 2020, 2022). This prompts a question; can an experimental manipulation of encoding variability increase estimates of σ_o in the UVSD model without simultaneous increases in d ? If so, then there is little doubt that encoding variability can independently determine estimates of old item variance, supporting the theory. As well as relating to the encoding variability hypothesis, this question has implications for single-system and dual-process theory. If the DPSD model were fitted to data where variability in old item strength increased independently of mean memory strength, it would represent this in a different way from the UVSD model. While greater values of the DPSD model's R parameter signify increases in mean strength and old item variance, lower values of the μ_o parameter indicate lower strength and greater old item variance (Spanton & Berry, 2020). The DPSD model could therefore increase the R parameter and decrease the μ_o parameter to represent greater old item variance while controlling overall memory strength. This contrasting behaviour between the two models is of interest when they are extended to represent priming. In the SS model, changes in old item variance for recognition are driven by the σ_o parameter, like in the UVSD model. As shown in the previous chapter, increases in this parameter affect priming, increasing the difference between miss and hit responses. In the DPSD1 model, however, greater values of R do the opposite, decreasing the miss-hit identification difference. If one experimental manipulation can prompt these two opposing predictions, the models can be tested against one another.

The aims of this chapter are twofold. Experiment 3 attempts to provide a new way to manipulate old item variance in the UVSD model while controlling new item variance and reducing the effect of overall strength as much as possible. When fitting the DPSD model to these data, its potential to

make unique predictions on the basis of its recollection process can also be evaluated. Experiment 4 then expands the design of Experiment 3 to include an identification task. The SS and DPSD1 model specifications are extended to fit this new data structure, and their opposing predictions are evaluated. To preface the results, Experiment 3 shows promising parameter estimates from the UVSD and DPSD1 models that may provoke opposing predictions if extended to the SS and DPSD1 models. However, the models were unable to be discriminated on this basis in Experiment 4, despite other support being found for the SS model. These results and their implications for further investigations of explicit and implicit memory are discussed.

3.1 Experiment 3

Previous research has highlighted conceptual difficulties with testing the encoding variability hypothesis. Spanton & Berry (2020) attempted to add Gaussian variability to continuous variables known to affect memory strength during encoding. However, this manipulation only had a weak effect on the variance in recognition confidence ratings, and so was not expected to prompt large changes in σ_0 . This may be because even without experimental manipulation, there are already a very large number of encoding variables that sum to determine levels of added strength in any condition. Therefore, any further attempts to experimentally manipulate a given encoding variable might have a minimal effect on old item variance because added strength already varies greatly. It could also be possible that the effect of any experimentally manipulated encoding variable is partially counteracted by any number of other encoding variables that occur naturally. When manipulating item characteristics, for example, if word frequency and strength are negatively related, whereas concreteness and strength are positively related, then any amount of added strength that a word may receive for having low word frequency may be balanced by a decrement in strength if that word also happens to have low concreteness. There is also likely to be a negative correlation between an item's baseline strength value and the increment of added strength it receives during study (Jang et al., 2012a). This, in conjunction with the aforementioned factors, makes it difficult to establish a strong experimental manipulation of encoding variability (Spanton & Berry, 2020). To mitigate these problems in their Experiments 1 and 2, Spanton & Berry (2022) manipulated multiple item characteristics simultaneously to achieve a greater combined experimental effect upon old item variance. In doing so, they ensured that these characteristics are correlated within a word list, addressing the possibility that manipulated item characteristics may systematically counteract each other. Returning to the example above, word frequency and concreteness would

be less likely to counteract one another if their values were negatively correlated, increasing their summed effect upon the variance of recognition confidence judgements. Spanton & Berry (2022) compared this condition with another wherein item characteristics are constrained to be as low in variance as possible, resulting in low encoding variability. Controlling the mean of each item characteristic measure to be equal across word lists in both high and low variability conditions also allowed them to control the overall memorability of stimuli in each set. This control allowed Spanton & Berry (2022) to establish an orthogonal manipulation of memory strength in their first two experiments. In the present experiment, it allows memory strength to be controlled, increasing the likelihood that any effects on old item variance are the result of encoding variability alone.

The present experiment attempted to test the encoding variability hypothesis by including these low and high item characteristic variance stimulus conditions within a single test phase. That is, half of the items in each list (studied, and new) had high Gaussian variability in four item characteristics, with the other half having low Gaussian variability. The following characteristics were chosen for having previously been shown to effect memory strength. Word frequency, a measure of how commonly a word is used in typical language, was shown to have significant effects on various recognition memory accuracy metrics in multiple studies (Glanzer & Bowles, 1976). Concreteness, a measure of a word's abstractness and imageability, was shown to have a roughly 8% effect on correct recognition rate by Fliessbach et al. (2006). Age of acquisition (AOA; Cortese et al., 2010), the age at which words are typically acquired in development, was shown to have a weak-moderate association with recognition confidence ratings. Finally, word length was shown to have a moderate negative relationship with correct recognition rate (Cortese et al., 2010, 2015).

The combination of these word lists into a single participant-facing study/test phase block was to ensure a valid test of the encoding variability hypothesis. If these high and low variability words were split into separate conditions, the item characteristic variability of old and new items within these conditions would have to be equated. This would be to prevent some words in high encoding variability conditions being artefactually more discriminable based on their extreme characteristics. Such an effect would result in simultaneous effects on memory strength, which is undesirable when attempting to devise a "pure" test of the encoding variability hypothesis. However, as σ_o is conceptualised as the ratio of new/old item variance in the UVSD model, equating the item characteristic variability of new and old items might not affect σ_o as intended. Presenting each word list in one study/test phase ensures that the overall variability of old and new word lists will be equal, while allowing word lists for each old-high variability, old-low variability, new-high variability, and new-low variability condition to be separated for analysis. In this way, Experiment

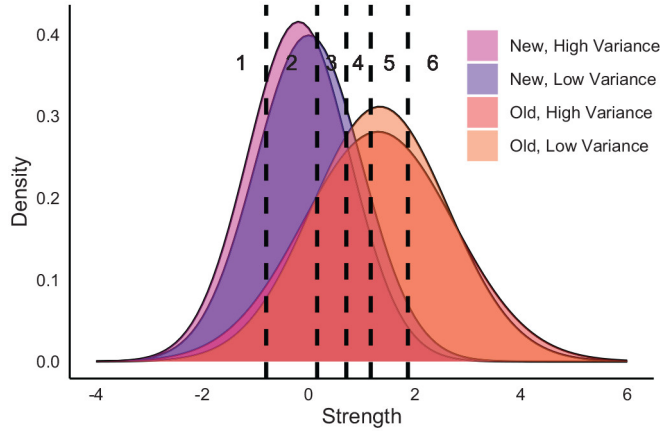


Figure 3.1: A depiction of the extended UVSD model specification, with parameters set to the mean estimates recovered from Experiment 3.

3 provides a principled method to test a key prediction of the encoding variability hypothesis; that σ_o should be greater in high variability conditions than in low variability conditions.

To allow the estimation of the key parameters in this experiment, we must define four distributions in the UVSD model – one for each condition (see Figure 3.1). The mean and standard deviation of the new-low distribution can be fixed so that $\mu_{nl} = 0$ and $\sigma_{nl} = 1$, allowing the means and standard deviations of each other condition to be free and scaled upon these fixed parameters. Since all the conditions appear to the participant in one study-test phase, it follows that the same decision criteria should be used to model judgements for words in every sub-list. Extending the UVSD model to represent this design, therefore, requires the 11 free parameters $\theta = \{\mu_{nh}, \mu_{ol}, \mu_{oh}, \sigma_{nh}, \sigma_{ol}, \sigma_{oh}, C_1, C_2, \dots, C_5\}$.

By contrast, the σ parameters for each distribution in the extended DPSD model are fixed to 1. This reflects the parameterization of familiarity as an equal-variance signal detection process in the conventional DPSD model (Yonelinas, 1994). The DPSD model shares the same free μ parameters as the UVSD, with μ_{nl} being fixed to zero. Two additional parameters are needed to represent recollection for each old item distribution; R_h and R_l . These represent the proportion of old items that are recollected in the high and low conditions, respectively. With this, the extended DPSD model has 10 free parameters $\theta = \{\mu_{nh}, \mu_{ol}, \mu_{oh}, R_h, R_l, C_1, C_2, \dots, C_5\}$. The likelihood functions and parameter recovery simulations for each model can be found in Appendices A and B, respectively.

With these models defined, the behaviour of their parameters can be evaluated alongside the predictions of the encoding variability hypothesis. If the encoding variability hypothesis holds, the UVSD model predicts that σ_{oh} will be greater than σ_{ol} , with no simultaneous effects of overall memory strength. In this case, the DPSD model would show greater estimates of R_h than R_l and

greater estimates of μ_{ol} than μ_{oh} .

3.1.1 Method

Participants

75 undergraduate psychology students from the University of Plymouth (57 females, 16 males, 2 non-binary/other) completed the experiment in exchange for course credits. Three participants were excluded during analyses due to outlying parameter estimates (see Results), leaving an effective sample of 72 participants that allowed for the detection of a minimum effect size $d_z = 0.33$ at 80% power in a paired-samples t -test.

Materials

A total of 240 unique words were used as stimuli (60 in each condition). Chosen words appeared in the SUBTLEX-UK word database (Van Heuven et al., 2014) and databases from Brysbaert et al. (2014) and Kuperman et al. (2012). Names, proper nouns, and hyphenated words were excluded from an aggregate of the above databases before sampling. Word frequency scores for these words were taken from the SUBTLEX-UK database (Van Heuven et al., 2014), concreteness scores were taken from Brysbaert et al. (2014), and AOA scores were taken from Kuperman et al. (2012). In high item characteristic variability conditions, each set of old or new words (four in total) was selected using an algorithm with the following criteria:

1. Words must be 4-10 characters long.
2. Each set of words must have approximately equal mean word frequency (≈ 3), concreteness (≈ 3), and AOA (≈ 10) scores (see Table 1 for exact values).
3. Concreteness and AOA scores must be strongly negatively correlated with word frequency scores within each word list ($r < -.77$ for concreteness and word frequency scores, $r < -.61$ for AOA and word frequency scores, and $r > .26$ for concreteness and AOA scores).
4. The distribution of word frequency, concreteness and AOA scores must not significantly deviate from a normal within each set, according to an Anderson-Darling test ($p > .05$).

The remaining four sets of old/new words in the low item characteristic variability condition were sampled with the following criteria:

1. Words must be 7 characters long.
2. Each set of words must have approximately equal mean word frequency, concreteness, and AOA scores (with the same constraints as the high item characteristic variance condition).
3. Each item characteristic variable must not be highly correlated. Among the word lists generated, word frequency and concreteness had a maximum negative correlation of $r = -.36$. Word frequency and AOA had a maximum negative correlation of $r = -.11$. Concreteness and AOA had a maximum positive correlation of $r = .03$.
4. Word frequency, concreteness and AOA scores must have low variance. For each word, the formula $\sum |(\mu_e - e_i)|$ was used to determine the summed difference between the mean of each item characteristic (e) across all possible words, and its corresponding value in the e^{th} word. The 240 words with the lowest summed difference scores were then randomly sampled without replacement to create the low encoding variability word lists.

The experiment was implemented using the OSWeb functionality of OpenSesame (Mathôt et al., 2012), and hosted on a JATOS (Lange et al., 2015) server. Participants completed the task in a lab, using Lenovo desktop computers running a browser window containing the experiment program.

Procedure

Participants first completed a study phase consisting of 120 trials. In each trial, they viewed a fixation point for 500 ms, a word for 3000 ms, and an inter-trial interval (a blank screen) for 500 ms. The words in the study phase were made up of one set of 60 high-variability words, and one set of 60 low-variability words; these sets were intermixed and presented in a different random order for each participant. The allocation of each high and low variability word list as an old or new item was also randomised across participants. Participants were instructed to pay attention to each word during the study phase, and that they should try to remember as many words as possible for a later memory test. After the study phase, participants had a 60-second break before reading instructions for the test phase. A countdown timer was displayed on screen during the break, showing the number of seconds left until the instructions for the next phase would appear.

The test phase had the same trial level structure as those in Experiments 1 and 2 of Spanton & Berry (2022). A fixation point appeared for 500 ms, followed by a randomly selected word that was either old or new in the centre of the screen. This word was presented until the participant made a recognition confidence judgement based on their degree of certainty that the item was old or new.

Participants made these responses with 1–6 keys at the top of the keyboard, using the response scale “1—Sure New, 2—Probably New, 3—Guess New, 4—Guess Old, 5—Probably Old, 6—Sure Old.” This key, and the prompt “New or Old?” were presented near the bottom of the screen as a static reminder of the response categories throughout each test phase. After each response, a 500 ms ITI (in which no information was displayed in the centre of the screen) was displayed, before the next trial. Participants were instructed to make use of the whole rating scale, and to prioritise the accuracy of their judgements over speed as they completed the task. A total of 240 words were presented (120 old, 120 new), with the new words coming from the remaining high and low item variability lists that were not shown at study. This resulted in a 2 (item characteristic variability level; high, low) \times 2 (item type; old, new) within-subjects design. As in the study phase, each participant completed a different random order of trials. Upon completing the test phase, participants input their age and gender into the experimental program before reading a full debrief.

3.1.2 Results

Three participants were excluded from all analyses for having outlying parameter estimates, in line with the approach taken in Experiments 1 and 2 of Spanton & Berry (2022). Estimates of σ_{oh} , σ_{oh} and σ_{oh} were also log-transformed because, with the value of σ_n fixed to equal 1, σ_o is a ratio and would otherwise violate the assumptions of a 2 \times 2 ANOVA. Bonferroni corrections were applied to all pairwise comparisons.

Item Characteristic Variability Manipulation

Regression analyses were conducted to gauge the effect of each manipulated item characteristic on recognition confidence responses. Each participant’s data was split by item type (old, new) and item characteristic variability level (high, low), and regression models with word frequency, concreteness, AOA, and word length as predictors were fit to each combination of factors. The proportion of significant regression models and mean R^2 values can be found in Table 3.1. A 2 \times 2 within-subjects ANOVA on R^2 was then conducted with item type and item characteristic variability level as factors. This ANOVA revealed a significant main effect of item characteristic variability on R^2 , $F(1, 71) = 18.67$, $p < .001$, $\eta_p^2 = .21$, BF = 1018.28. There was no significant main effect of item type, $F(1, 71) = 0.27$, $p = .61$, $\eta_p^2 < .01$, BF = 0.15, and no significant interaction, $F(1, 71) = 2.33$, $p = .13$, $\eta_p^2 = .03$, BF = 0.50. This indicates that the variance in recognition confidence ratings was largely explained by the item characteristic variability manipulation, rather than the

Table 3.1: The proportion of significant item characteristic regression models and mean R^2 values (standard deviations in brackets) for each condition in Experiments 3 and 4.

Experiment	Conditions	P (significant) regressions	Mean R^2
Experiment 3	Old, High Variability	.19	.09 (.06)
	Old, Low Variability	.04	.05 (.04)
	New, High Variability	.19	.08 (.07)
	New, Low Variability	.04	.06 (.04)
Experiment 4	Old, High Variability	.12	.08 (.05)
	Old, Low Variability	.05	.05 (.04)
	New, High Variability	.21	.09 (.06)
	New, Low Variability	.05	.05 (.04)

Table 3.2: Mean parameter estimates for the extended UVSD and DPSD models in Experiment 3, with standard deviations in parentheses.

Relevant Model	Parameter	UVSD Mean Estimate		DPSD Mean Estimate	
UVSD					
	σ_{nh}	0.96	(1.36)	-	-
	σ_{ol}	1.28	(1.31)	-	-
	σ_{oh}	1.42	(1.40)	-	-
DPSD					
	R_l	-	-	.23	(.20)
	R_h	-	-	.27	(.21)
Shared					
	μ_{nh}	-0.19	(0.28)	-0.22	(0.24)
	μ_{ol}	1.34	(0.69)	0.81	(0.46)
	μ_{oh}	1.31	(0.77)	0.59	(0.45)
	C_1	-0.80	(1.22)	-0.72	(1.10)
	C_2	0.16	(0.60)	0.15	(0.55)
	C_3	0.71	(0.44)	0.67	(0.40)
	C_4	1.17	(0.56)	1.13	(0.55)
	C_5	1.87	(0.70)	2.56	(1.96)

presence of words in the study phase. The high variability words accounted for around 4% more total variance in recognition confidence ratings for old items than low variability words, which is comparable with the results in Experiments 1 and 2 of Spanton & Berry (2022). This confirms that the manipulation of item characteristic variability impacted on recognition responses. With this in mind, I next considered the central issue of whether estimates of σ_o in the UVSD model would be similarly impacted.

UVSD Model Parameter Estimates

The parameter estimates from the UVSD and DPSD models can be found in Table 3.2. A one-factor repeated measures ANOVA with a Greenhouse-Geisser sphericity correction was used to compare the estimates of σ in the old-high, old-low, and new-high conditions. Estimates significantly differed across conditions, $F(1.66, 117.91) = 85.66, p < .001, \eta_p^2 = .55, BF = 1.75 \times 10^{10}$. The

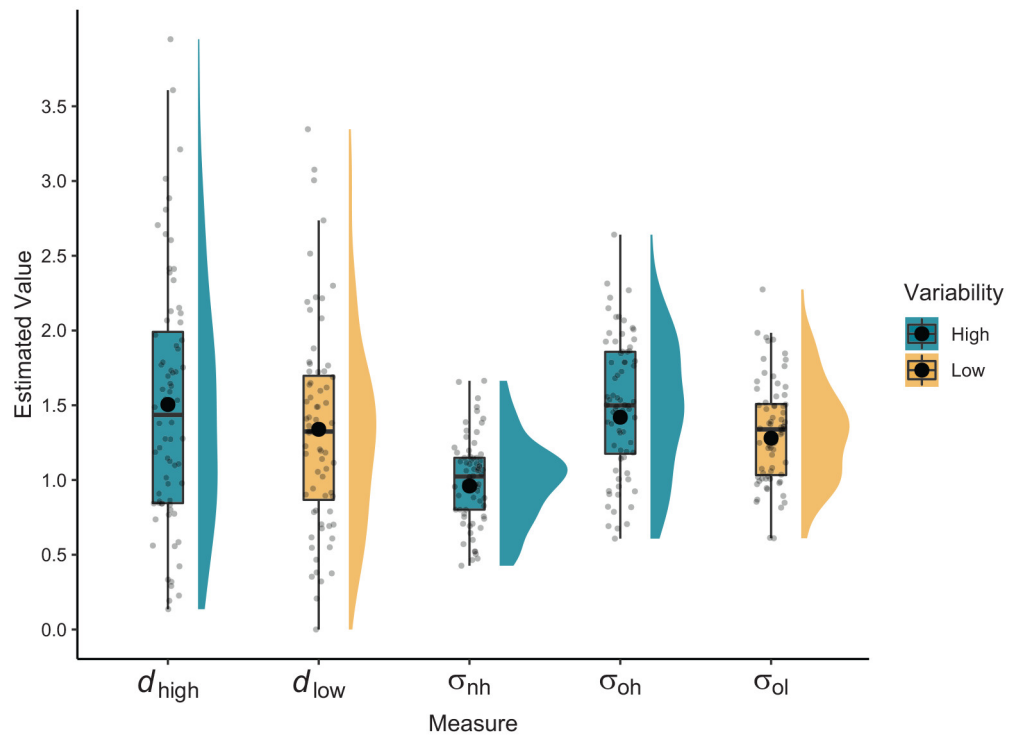


Figure 3.2: Raincloud plots of σ parameter estimates and discriminability measures in the UVSD model, with circular points denoting means. The mean of each log-sigma parameter was calculated and then exponentiated.

ordinal pattern of variance estimates for each distribution can be seen in Figure 3.2. Pairwise comparisons confirmed that estimates of σ_{nh} were reliably lower than those of both σ_{oh} , $t(71) = -11.33$, $p < .001$, $BF = 8.27 \times 10^{14}$, and σ_{ol} , $t(71) = -9.00$, $p < .001$, $BF = 2.42 \times 10^{10}$. Crucially however, estimates of σ_{oh} were significantly greater than estimates of σ_{ol} , $t(71) = 4.06$, $p < .001$, $BF = 300.76$. Further, a one-sample t -test also revealed that σ_{nh} did not significantly differ from 1, the fixed value of σ_{nl} , $t(71) = -1.26$, $p = .21$, 95% CI [-0.11, 0.03], $BF = 0.28$. This means that the manipulation of item characteristic variability only affected old items, in line with the encoding variability hypothesis.

To assess the possibility that differences in old item variance may have been driven by effects of overall memory strength, I calculated discriminability (d) measures for high and low variability conditions. These measures were given by calculating $d_h = \mu_{oh} - \mu_{nh}$ and $d_l = \mu_{ol} - \mu_{nl}$ respectively, on a participant level. Discriminability measures were reliably greater for high variability items than for low variability items, $t(71) = 2.90$, $p = .004$, 95% CI [0.05, 0.28], $BF = 6.05$. This increase in discriminability for high variability items was likely driven by estimates of μ_{nh} being reliably lower than 0, the fixed value of μ_{nl} , $t(71) = -5.72$, $p < .001$, 95% CI [-0.26, -0.12], $BF = 58633.94$. By contrast, no reliable differences were found between estimates of μ_{oh} and μ_{ol} , $t(71) = -0.60$, $p = .55$, 95% CI [-0.11, 0.06], $BF = 0.15$. This means that greater overall memory strength for high

variability items coincided with greater estimates of σ_o for those items. Therefore, as in Experiment 2, it is unclear whether increases in old item variance are due to manipulated encoding variability because these increases were not independent of changes in d .

For comparison with the results of Spanton & Berry (2020) and Spanton & Berry (2022), I examined the relationship between discriminability and σ_o parameters within the high and low variability conditions using linear regression. There was a significant positive relationship between d_h and σ_{oh} , $F(1, 70) = 32.64$, $p < .001$, $R^2 = .32$. There was also a significant positive relationship between d_l and σ_{ol} , $F(1, 70) = 21.07$, $p < .001$, $R^2 = .23$. This indicates that estimates of mean memory strength and variability in memory strength for old items were positively associated.

DPSD Model Parameter Estimates

I also examined the DPSD model's parameter estimates. DPSD model estimates of R_h were significantly greater than estimates of R_l , $t(71) = 2.80$, $p = .007$, 95% CI [0.01, 0.08], BF = 4.71. This indicates that the model estimated that there was greater recollection for items in the high condition than the low. To assess the model's predictions regarding familiarity, measures of d_h and d_l were calculated in the same way as for the UVSD model. There was no reliable difference between these measures, $t(71) = 0.06$, $p = .95$, 95% CI [-0.10, 0.10], BF = 0.13. This lack of a difference can be dissected further by looking at the model's estimated μ parameters. Estimates of μ_{ol} were significantly greater than those of μ_{oh} , $t(71) = -5.49$, $p < .001$, 95% CI [-0.30, -0.14], BF = 24481.46. However, estimates of μ_{nh} were reliably lower than 0, the fixed value of μ_{nl} , $t(71) = -7.93$, $p < .001$, 95% CI [-0.28, -0.16], BF = 4.06×10^8 . When comparing d between high and low item characteristic variability conditions, these effects cancel each other out, resulting in no difference in estimates of familiarity between these conditions.

Comparison of Fits

To compare the goodness of fit of the UVSD and DPSD models, AICs were calculated for each model's fit to each participant's data. Given that the extended UVSD model has 11 free parameters and the extended DPSD model has 10, using the AIC was necessary to account for this difference. The best-fitting model for each participant's data was given by the minimum AIC. The UVSD model fitted best to the data in 38.89% of cases, whereas the DPSD fit best in 61.11% of cases. To determine the extent that each model fit individual participants' data best, AIC differences were calculated in the same way as in Chapter 2. AIC differences for the UVSD model ($M = 2.03$, SD

= 2.70) did not significantly differ from those for the DPSD model ($M = 1.34$, $SD = 2.74$) in a Wilcoxon Signed-Rank test, $V = 1654$, $p = .06$. Only 2.78% of AIC differences for each model were greater than 10, indicating insufficient support for the models in those cases. In the majority of cases, there was substantial empirical support for each model, with $\Delta_{UVSD} \leq 2$ in 59.72% of cases, and $\Delta_{DPSD} \leq 2$ in 80.56% of cases. On balance, this suggests that the UVSD and DPSD models both accounted for most of the data to an equal standard, although the DPSD model gave better fits than the UVSD model in more cases.

3.1.3 Discussion

Parameter estimates from the UVSD model showed a clear increase in estimates of σ_o in high variability conditions. Estimates of new item variance remained constant across high and low variability conditions. However, old and new items were also more discriminable in the high variability conditions than in low variability conditions. Therefore, the increase in old item variance predicted by the UVSD model coincided with an increase in memory strength, and so cannot be taken as clear evidence for the encoding variability hypothesis. Moreover, measures of discriminability and old item variance were positively associated on a participant level in both high and low variability conditions, mirroring the linear relationships found between σ_o and d in previous research (Spanton & Berry, 2020, 2022).

The parameter estimates from the DPSD model showed that R_h tended to be greater than R_l , meaning there was greater recollection in the high variability condition. Estimates of familiarity, given by d_h and d_l , remained unchanged between conditions. It has previously been shown that increases in the R parameter of the DPSD model signify a simultaneous increase in overall memory strength and variability in memory strength for old items at large (Spanton & Berry, 2020). This means that the DPSD model shares the same psychological predictions about the patterns in the present data as the UVSD model. However, the increase in memory strength and variability is driven by the recollection process in the DPSD model, in contrast to a unidimensional strength variable in the UVSD model. Simulations in Chapter 2 showed that increases in the DPSD1 model's recollection parameter decrease the $M - H$ difference, whereas increases in old item variance in the SS model increase this difference. Since the present experimental manipulation prompts these two opposing behaviours simultaneously, the SS and DPSD1 models have the potential to make opposing predictions about identification performance if extended to represent priming data. The following experiment attempts to elicit these predictions.

3.2 Experiment 4

The results of Experiment 3 confirmed the UVSD and DPSD models account for increases in memory strength and variability in memory strength for old items differently. The UVSD model explains this as a simultaneous increase in estimates of σ_o and d , whereas the DPSD model does so through an increase in recollection. These differences in representation have interesting implications. If each model were extended to represent data from an identification task in the same way as the SS and DPSD1 models in Berry et al. (2012), the models would make opposing predictions about identification performance for different types of recognition responses. Unlike Experiments 1 and 2, the present experiment (and the experiments following) used a non-RT measure of identification performance. Instead of a CID-R procedure where a stimulus is repeatedly exposed for an increasing duration against a mask, a self-paced procedure where participants reveal more of a stimulus with a button press was used. The dependent variable in this procedure was the image clarity level at which a stimulus was identified, ranging from 1 (least clear) to 10 (unobscured; see Figure 3.3). This measure was adopted for multiple reasons. Firstly, the self-paced nature of the task reduced the chance of presentation latency in the experiment program affecting the accuracy of the identification measure. Secondly, differences in identification levels are easier to equate to specific differences in stimulus clarity compared with differences in CID RTs. Similar measures have previously been successfully implemented in conjoint studies of explicit memory and perceptual identification (Mazancieux et al., 2020).

Aside from the identification measure, the predictions of the models take the same form as in Chapter 2. The SS model predicts that as estimates of old item variance increase, either alone or alongside discriminability measures, so does the $M - H$ difference. The DPSD1 model predicts that as the probability of recollection increases, $M - H$ decreases. Since these opposing behaviours may coincide in the same condition, as Experiment 3 showed, this provides a critical test of the SS and DPSD1 models. I now proceed to verify the predictions of each model through simulation and extend the design of the previous experiment to test these predictions.

3.2.1 Expected Value Simulations

I first extended the SS and DPSD1 models to fit recognition and priming data from four item characteristic conditions, as in Experiment 3. The extended SS model has 14 free parameters: the means of the two old item strength distributions, μ_{o1}, μ_{o2} , the mean of the new-high condition

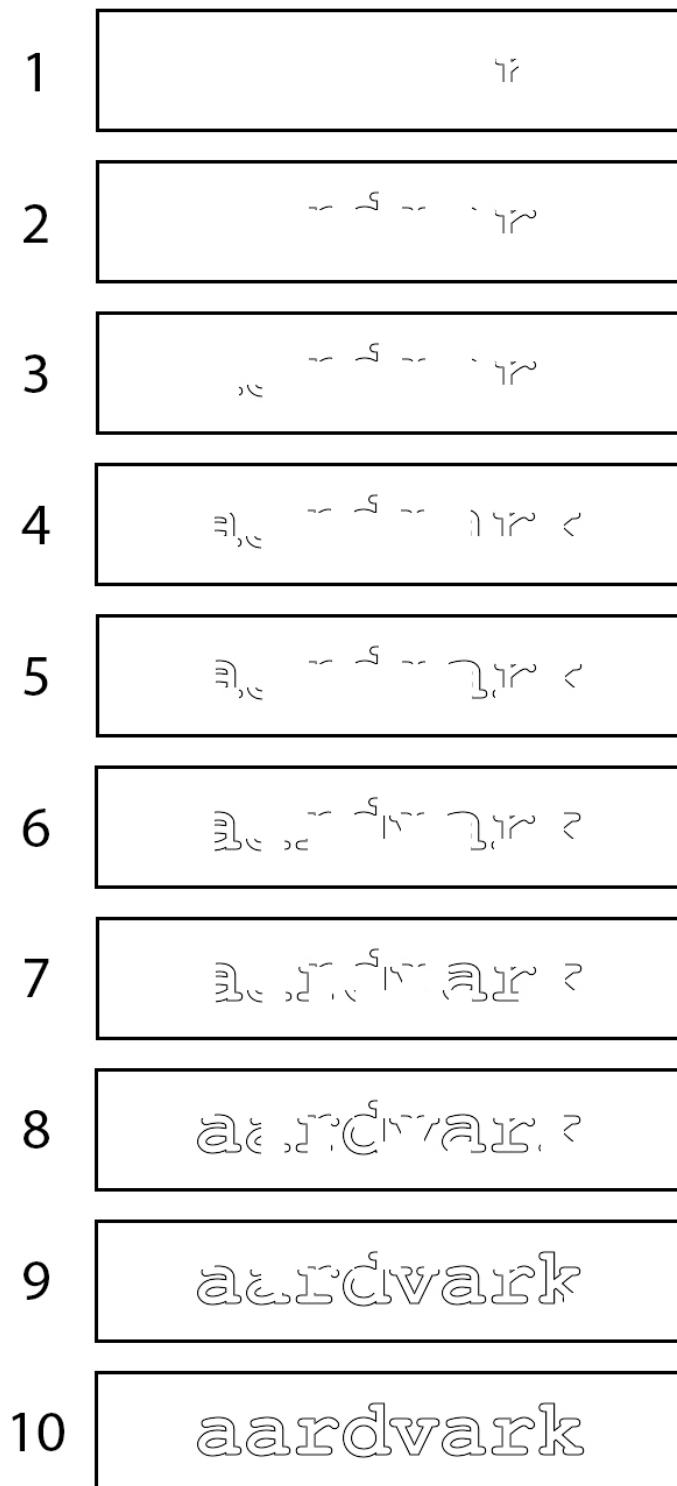


Figure 3.3: An example of a fragmented word image stimulus from Experiments 4, 7, 8, and 9.

Table 3.3: The lower and upper bounds on the uniform distributions used to generate simulated data from the SS and DPSD1 models. Values of μ_{nh} were fixed to 0 in the SS model simulation, and values of each σ parameter was fixed to $\sqrt{0.5}$ in the DPSD1 model.

Parameter	SS		DPSD1	
	a	b	a	b
μ_{nh}	-	-	-1	1
μ_{oh}	0	4	0	4
μ_{ol}	0	4	0	4
σ_{nh}	$\sqrt{0.5}$	3	-	-
σ_{oh}	$\sqrt{0.5}$	4	-	-
σ_{ol}	$\sqrt{0.5}$	4	-	-
b	8	13	8	13
s	0.1	2	0.1	2
σ_p	1	3	1	2
C_3	0.5	1.5	0	2

distribution, μ_{nh} , the variance parameters $\sigma_{nh}, \sigma_{ol}, \sigma_{oh}$, priming parameters b, s, σ_p , and criteria C_1, \dots, C_5 . The extended DPSD1 model has 13 free parameters: three strength distribution means, $\mu_{nh}, \mu_{ol}, \mu_{oh}$, parameters that represent the probability of recollection in each old item condition, R_{ol}, R_{oh} , and the same priming parameters and criteria as the SS model. The likelihood equations for these extended models can be found in Appendix A. Parameter recovery simulations confirmed that both models recovered parameter estimates that were close to true generative parameters of simulated data (see Appendix B).

This modelling exercise followed the same procedure as the expected value simulation in Chapter 2. First, 100,000 sets of generative parameters for both the extended SS and DPSD1 models were sampled from uniform distributions. The bounds of these uniform distributions were set to cover a theoretically plausible range of parameter values that could be feasibly observed in standard experimental data (see Table 3.3). Additionally, another set of generative parameters for the SS model was also sampled to reflect the outcome that estimates of old item variance and mean memory strength tend to scale with one another in Gaussian signal-detection models (Spanton & Berry, 2020, 2022). In this set, the σ_{ol} and σ_{oh} parameters were generated to have a strong linear relationship with the μ_{ol} and μ_{oh} parameters respectively. This relationship was specified using the equation

$$\sigma_{o,K} = \sqrt{0.5} + \mu_{o,K} \times 0.25 + \mathcal{N}(0, 1) \quad (3.1)$$

where K represents the variability condition (high, low). These generative parameters were then used to derive expected values of perceptual identification levels for hit and miss responses. For

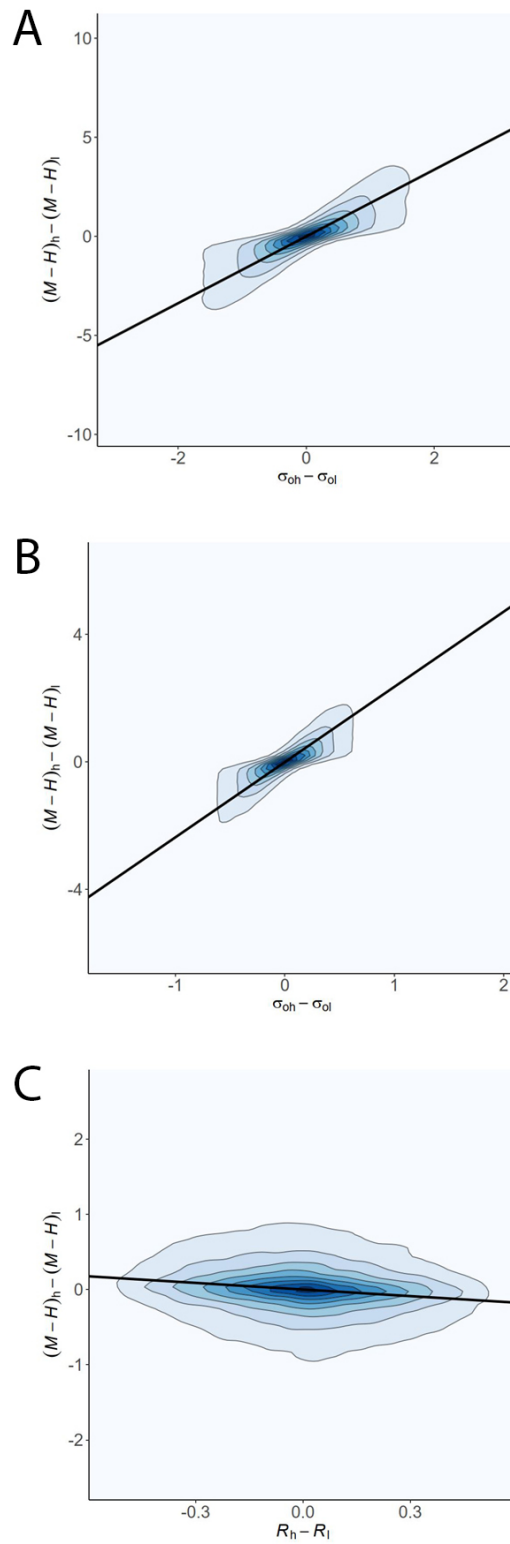


Figure 3.4: Density plots showing the extended SS and DPSD1 model predictions about the M-H difference between variability conditions as estimates of σ and R parameters change, respectively. Panel A shows data from the SS model expected value simulation when μ and σ parameters were independently generated. Panel B shows data from this simulation when σ parameters were related according to Equation 3.1. Panel C shows data from the DPSD1 model expected value simulation.

the SS model, Equation 2.1 was used to calculate identification reaction times conditional on a recognition judgement for each variability condition. This function was also used for misses in the DPSD1 model. For hits in the DPSD1 model, Equation 2.2 was used.

The identification level difference $(M - H)_h - (M - H)_l$ was calculated for each participant and analysed as a function of $\sigma_{oh} - \sigma_{ol}$ in the SS model, and $R_{oh} - R_{ol}$ in the DPSD1 model. These relationships can be seen in Figure 3.4. Linear regression models showed a strong positive relationship between $\sigma_{oh} - \sigma_{ol}$ and $(M - H)_h - (M - H)_l$ in the SS model where strength scaling was assumed, $R^2 = .74$, $F(1, 99998) = 286288.40$, $p < .001$. An even stronger positive relationship between these two indices was found in the SS model where strength scaling was not assumed, $R^2 = .77$, $F(1, 99998) = 330519.10$, $p < .001$. A weak but significant negative relationship between $R_{oh} - R_{ol}$ and $(M - H)_h - (M - H)_l$ was observed in the DPSD1 model, $R^2 = .01$, $F(1, 99998) = 1394$, $p < .001$. These results verify the SS and DPSD1 model's unique predictions. When the difference between the σ_o parameters in the SS model increases, so does the $M - H$ difference between high and low variability conditions. In the DPSD model, increasing the difference between R_h and R_l does the opposite, slightly decreasing the difference between $M - H$ in the high and low conditions. These contrasting predictions about the $M - H$ difference can now be tested in an experiment that extends the method of Experiment 3 to measure identification performance.

3.2.2 Method

Participants

75 undergraduate psychology students from the University of Plymouth completed the experiment. They did so in exchange for participation credits needed to pass a module on their course. Three participants were excluded from analyses (see Results section), leaving an effective sample of 73 participants (61 females, 11 males, 1 non-binary/other) with a mean age of 19.7 years ($SD = 1.96$). This enabled the detection of a minimum effect size $d_z = 0.33$ at 80% power in a paired-samples t -test.

Materials

The word lists from the previous experiment were used. The words were presented in each phase as images that were generated using the *magick* R package (Ooms, 2021). Each word was generated in white "mono" font with a black outline, against a white 600×100 px background. For the fragment

identification task, each word was obscured at 10 levels of clarity (1 being the most obscured, and 10 being the wholly revealed word). These fragmented images were generated using an R script that divided each image into a grid of 20×20 px cells, discounted any cells that were wholly part of the image background, and then specified the number of the remaining cells to randomly obscure with white squares at each fragmentation level. The number of cells to obscure at each fragmentation level (x) was calculated as $\lceil k \times (1 - 0.75^{(N-x)}) \rceil, x \in [1, N]$, where k equals the total number of non-background cells for a given image, and N equals the total number of fragmentation levels. As in Berry et al. (2014), this power function was used to ensure progressively larger proportions of the stimulus were revealed as each trial progressed. The experimental program was hosted on a JATOS server (Lange et al., 2015) and displayed through web browsers on Lenovo computers using monitors with 1920×1080 px resolution.

Procedure

There were two-word lists for each item characteristic variability condition (high and low). For each condition, one of the word lists was assigned to be old, and the other new. This assignment was random, meaning that all four combinations of word list-condition pairings were used across the sample. Each participant saw a different random order of trials in the study and test phases.

After giving informed consent, participants studied words from the old-high and old-low conditions. These word lists were intermixed and presented in sequential trials in a single study phase. In each study phase trial, a fixation point appeared for 500 ms, followed by an old word image for 3000 ms, and then an inter-trial interval of 500 ms. Participants were instructed to keep their attention focused on the words being presented and try to remember as many as possible for a later memory test. After the study phase, participants completed a sixty-second retention interval, where they were instructed to take a break before the next phase began. A countdown timer appeared on the screen, which displayed the number of seconds remaining until the end of the retention interval.

In the test phase, participants saw the old-high, old-low, new-high, and new-low items in a random order of trials. In each trial, the participants gave a fragment identification response, and then a recognition response. In the fragment identification task, participants first saw a studied item obscured at the lowest level of image clarity. If they could not identify the item, they would press the 'R' key to see the next most obscured version of the item, revealing more of the object with each response. They were instructed to identify the image at the lowest level of clarity possible (therefore pressing 'R' as little as possible). When they were able to identify the item, they pressed

the 'N' key to name it. A static reminder of these keys was presented at the bottom of the screen until the participant pressed 'N', at which point a text box appeared in its place. The prompt "Name the word: Please type your response in the text box below. Press Space to submit your response." was presented as a static prompt when the text box appeared.

After naming the word and submitting their response, participants then saw the fully unobscured image of the word. They were instructed to give a "New or Old?" recognition judgement using the same 1-6 rating scale as the previous experiments. After pressing the relevant number key to indicate their response, the next test trial began. After completing 80 trials and 160 test trials, the participant was given the option on a separate screen to take a sixty-second break. They pressed 'B' to take a break, and 'C' to continue with the task. Upon pressing 'B', the participant saw a screen with a countdown timer displaying the seconds remaining until the end of their break. After completing all 240 test trials, participants gave their age and gender and were debriefed.

3.2.3 Results

Before further analyses, all trials where participants made incorrect identification responses were excluded. Excluded trials made up 4.59% of all trials. σ parameter estimates from the SS model were also log-transformed as in the previous experiment. Three participants were also excluded list-wise from all analyses because their data prompted outlying parameter estimates from the DPSD1 and SS models. Two participants were excluded for large outlying estimates of C_5 in the DPSD1 model, and the other for a negative outlying estimate of σ_{nh} resulting from the log transformation.

Item Characteristic Variability Manipulation

As in the previous experiment, regression analyses were used to gauge the effect of the item characteristic manipulation on recognition confidence ratings. The proportion of significant regression models and mean R^2 values from this analysis can be found in Table 3.1. A 2×2 within-subjects ANOVA on R^2 with item type (old, new) and item characteristic variability condition (high, low) as factors found a main effect of item characteristic variability condition, $F(1, 71) = 32.68$, $p < .001$, $\eta_p^2 = .32$, $BF = 160050.50$. There was no effect of item type on R^2 , $F(1, 71) = 1.14$, $p = .29$, $\eta_p^2 = .02$, $BF = 0.23$, nor an interaction, $F(1, 71) = 2.24$, $p = .14$, $\eta_p^2 = .03$, $BF = 0.55$. This indicates that, as in the previous experiment, the manipulation of item characteristics explained a substantial proportion of variability in recognition confidence ratings, whereas the presence of items in the

Table 3.4: Mean parameter estimates for the extended SS and DPSD1 models in Experiment 4, with standard deviations in parentheses. Dashes indicate that a given parameter does not belong to the model in question.

Parameter	SS-extended Mean Estimate		DPSD1-extended Mean Estimate	
σ_{nh}	0.81	(0.36)	-	-
σ_{ol}	1.16	(0.47)	-	-
σ_{oh}	1.35	(0.42)	-	-
R_l	-	-	0.25	(0.17)
R_h	-	-	0.28	(0.17)
μ_{nh}	-0.24	(0.28)	-0.15	(0.21)
μ_{ol}	1.18	(0.70)	0.58	(0.30)
μ_{oh}	1.22	(0.79)	0.48	(0.27)
b	8.63	(0.47)	8.63	(0.48)
s	0.18	(0.13)	0.33	(0.17)
σ_p	0.76	(0.12)	0.75	(0.12)
C_1	-0.99	(0.91)	-0.84	(0.81)
C_2	-0.11	(0.55)	-0.08	(0.47)
C_3	0.43	(0.46)	0.40	(0.38)
C_4	1.01	(0.50)	0.94	(0.41)
C_5	1.74	(0.60)	2.20	(1.78)

study phase did not.

Task Performance

A two-way within-subjects ANOVA with factors item state (old, new) and item characteristic variability level (high, low) found a main effect of item state on fragmentation level, $F(1, 71) = 137.29$, $p < .001$, $\eta_p^2 = .66$, $BF = 5.67 \times 10^{32}$. No main effect of item characteristic variability level was found, $F(1, 71) = 0.59$, $p = .44$, $\eta_p^2 = .01$, $BF = 0.14$. No reliable evidence for an interaction was found, $F(1, 71) = 1.13$, $p = .29$, $\eta_p^2 = .02$, $BF = 0.24$. The main effect of item state showed a priming effect; fragmentation performance was better for old items than new items, as expected. However, there was no identification benefit for either high or low-variability items and no interaction. d' was calculated from each participant's data to assess overall recognition memory performance. These measures of d' ($M = 0.92$, $SD = 0.45$) were reliably greater than 0, $t(71) = 17.33$, $p < .001$, 95% CI [0.81, 1.02], $BF = 1.06 \times 10^{24}$, which shows that participants had above-chance recognition memory performance.

SS Extended Model Fits

The means and standard deviations of parameter estimates for the SS and DPSD1 extended models can be found in Table 3.4. A one-way repeated measures ANOVA with a Greenhouse-Geisser sphericity correction was used to compare estimates of the σ parameter in the new-high, old-low, and old-high conditions. There was a significant main effect of condition, $F(1.74, 123.61) = 51.84$, $p < .001$, $\eta_p^2 = .42$, $BF = 5.08 \times 10^{14}$. The ordinal pattern of these parameter estimates is shown in Figure 3.5. Bonferroni-corrected pairwise comparisons confirmed that estimates of σ_{nh} were significantly lower than estimates of σ_{oh} , $t(71) = -9.50$, $p < .001$, $BF = 2.62 \times 10^{11}$, and σ_{ol} , $t(71) = -6.18$, $p < .001$, $BF = 346231.60$. There was also a significant difference between estimates of σ_{oh} and σ_{ol} , $t(71) = 3.66$, $p = .001$, $BF = 49.26$. This indicates that both old item distributions in the SS model had greater strength variance than the new-low distribution. The old-high distribution also had greater variance than the old-low distribution.

To assess overall memory strength in the extended SS model, measures of d_h and d_l were calculated the same way as in the previous experiment. A paired samples t -test revealed that estimates of d_h ($M = 1.45$, $SD = 0.90$) were significantly greater than estimates of d_l ($M = 1.18$, $SD = 0.70$), $t(71) = 4.84$, $p < .001$, 95% CI [0.16, 0.39], $BF = 2352.05$. This difference resulted from the same pattern of parameter estimates as in the previous experiment. There was no reliable difference between estimates of μ_{oh} and μ_{ol} , $t(71) = 0.84$, $p = .85$, 95% CI [-0.05, 0.13], $BF = 0.18$. However, values of μ_{nh} tended to be less than 0, the fixed value of μ_{nl} , $t(71) = -7.16$, $p < .001$, 95% CI [-0.30, -0.17], $BF = 17485068$. This means that high-variability items had greater discriminability than low-variability items. This effect was driven by new-high variability items having a lower mean strength than new-low variability items.

DPSD1 Extended Model Fits

In the extended DPSD1 model, estimates of R_h were no greater than R_l , $t(71) = 1.44$, $p = .15$, 95% CI [-0.01, 0.05], $BF = 0.35$. Calculated measures of d_h ($M = 0.63$, $SD = 0.34$) did not differ significantly from values of d_l ($M = 0.58$, $SD = 0.30$), $t(71) = 1.37$, $p = .17$, 95% CI [-0.02, 0.11], $BF = 0.32$. However, underlying this null effect were two results of interest. Estimates of μ_{oh} were reliably lower than estimates of μ_{ol} , in a pattern opposing the SS model, $t(71) = -4.28$, $p < .001$, 95% CI [-0.15, -0.05], $BF = 340.36$. Secondly, estimates of μ_{nh} were reliably less than 0, the fixed value of μ_{nl} , $t(71) = -5.93$, $p < .001$, 95% CI [-0.20, -0.10], $BF = 133363.50$. Interpreting these results, we can see that overall, the DPSD1 model predicted no difference in overall memory

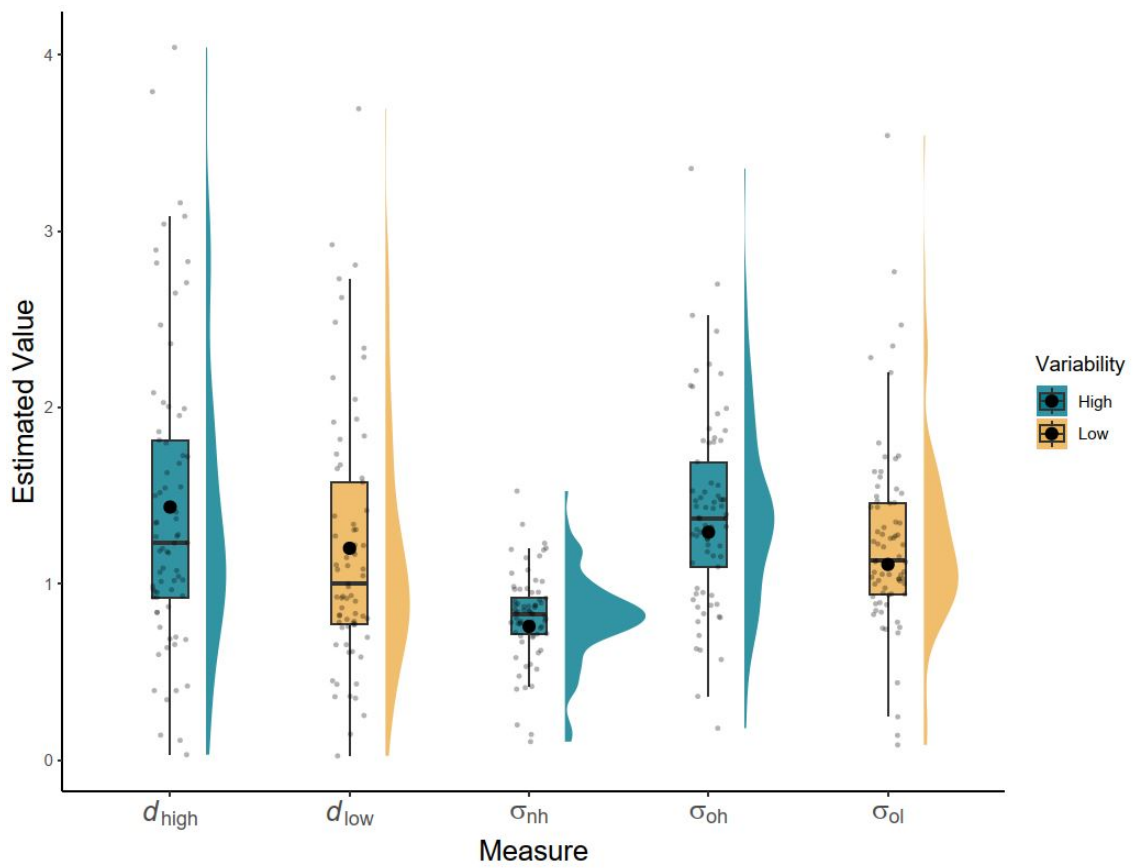


Figure 3.5: Raincloud plots of σ parameter estimates and discriminability measures in the SS model, with circular points denoting means. The mean of each log-sigma parameter was calculated and then exponentiated.

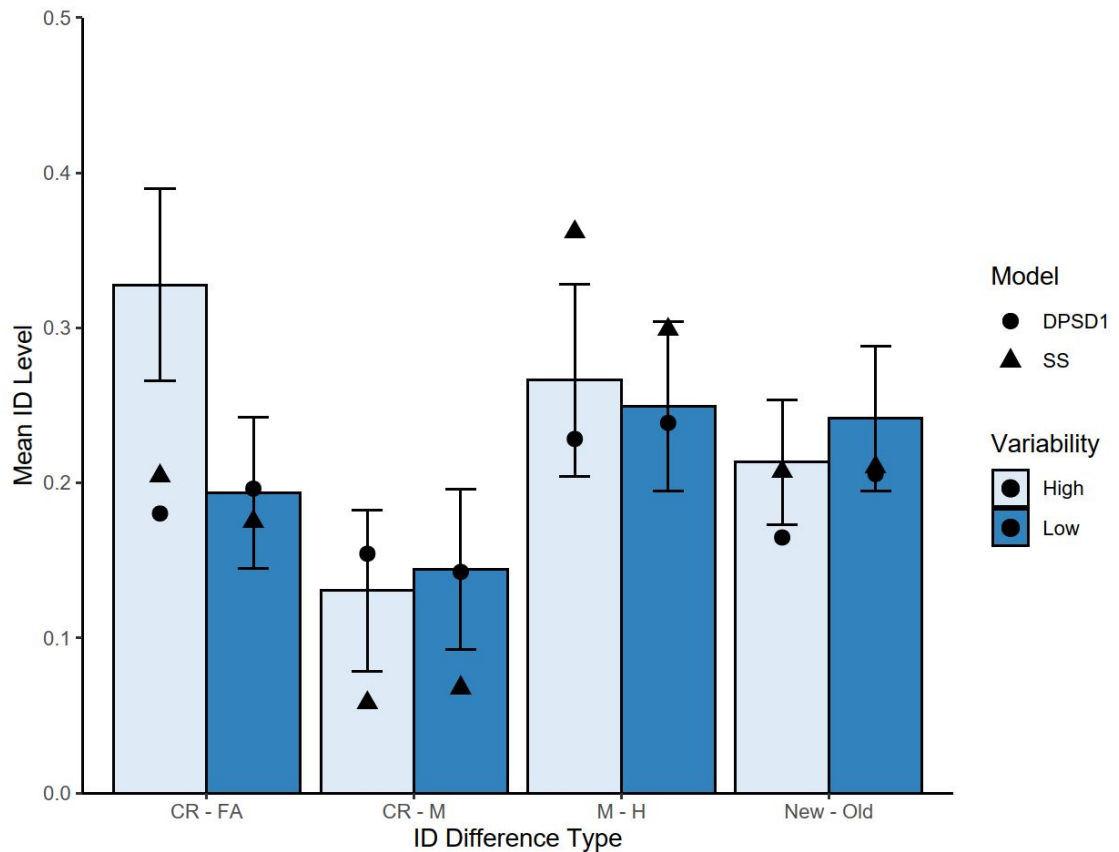


Figure 3.6: Bar plots of differences between identification performance for items receiving different types of recognition responses across high and low conditions. H = hit, M = miss, FA = false alarm, CR = correct rejection. New - Old refers to the priming effect. Mean expected values from the SS and the DPSD1 models are denoted by the circular and triangular points for each of these measures. The error bars represent the 95% confidence intervals of each mean. Both models predicted above-zero differences in each comparison.

strength between high and low variability conditions; estimates of both R parameters did not reliably differ, and neither did measures of d . Because the DPSD1 model represents variability in old item recognition memory strength with its R and μ parameters, it is apparent that this did not differ between variability conditions. This puts the DPSD1 model's psychological predictions about memory strength in opposition to those of the SS model, which showed an increase in both overall strength and variability in strength for old-high items.

Identification Performance Predictions

Within the old-high stimulus condition, miss responses were identified at a significantly lower identification level than hit responses, $t(71) = -8.49$, $p < .001$, 95% CI [-0.33, -0.21], $BF = 4.11 \times 10^9$. This difference was also observed within the old-low condition, $t(71) = -8.76$, $p < .001$, 95% CI [-0.30, -0.19], $BF = 1.24 \times 10^{10}$. To assess each model's key predictions about identification performance, the identification metric $M - H$ was calculated within old-low and old-high conditions

for each participant (see Figure 3.6). Values of $M - H$ in high variability conditions ($M = 0.27$, $SD = 0.27$) did not differ from those in low variability conditions ($M = 0.24$, $SD = 0.24$), $t(71) = 0.56$, $p = .58$, 95% CI [-0.06, 0.10], $BF = 0.15$. Inputting fitted values into each model's expected value functions allowed for each model's prediction of the $M - H$ difference to be calculated analytically (see Figure 3.6). The SS model predicted that values of $M - H$ should be greater in the high condition ($M = 0.34$, $SD = 0.19$) than the low condition ($M = 0.28$, $SD = 0.18$), $t(71) = 3.88$, $p < .001$, 95% CI [0.03, 0.09], $BF = 94.99$. The DPSD1 model predicted the opposite; that values of $M - H$ should be greater in the low condition ($M = 0.23$, $SD = 0.13$) than the high condition ($M = 0.22$, $SD = 0.13$), $t(71) = -3.74$, $p < .001$, 95% CI [-0.02, -0.01], $BF = 62.40$, though the difference in means was very small in real terms. This analysis shows that neither model predicted the null effect observed in the data, based upon fitted parameter estimates.

As an exploratory analysis, further identification predictions made by the SS model were tested in high and low conditions, as done in the speeded and unspeeded conditions of the experiments in Chapter 2 (see Figure 3.6). The mean identification level for correct rejections was reliably greater than that for false alarms in the high variability condition, $t(71) = 9.84$, $p < .001$, 95% CI [0.26, 0.39], $BF = 1.02 \times 10^{12}$. This difference also held in the low variability condition, $t(71) = 7.70$, $p < .001$, 95% CI [0.14, 0.24], $BF = 1.57 \times 10^8$. The mean identification level for correct rejections was also greater than that for misses in the high condition, $t(71) = 4.82$, $p < .001$, 95% CI [0.08, 0.18], $BF = 2141.40$, and in the low condition, $t(71) = 5.68$, $p < .001$, 95% CI [0.09, 0.19], $BF = 51934.77$. Finally, the new-old priming effect was significantly greater than the difference in mean identification for correct rejections and misses in the high condition, $t(71) = 3.94$, $p < .001$, 95% CI [0.04, 0.12], $BF = 116.11$. This difference was also found in the low variability condition, $t(71) = 4.53$, $p < .001$, 95% CI [0.05, 0.13], $BF = 793.59$. All of these results are predicted by the SS model (Berry et al., 2012). The DPSD1 model also makes similar predictions for these comparisons. Although these comparisons do not distinguish between the models, they serve as a confirmation of previous results pertaining to each model's validity.

Comparison of Fits

AICs were used to assess each model's goodness of fit. The SS model fitted best to the data in 44.4% of cases, and the DPSD1 model fitted best in the remaining 55.6%. The AIC differences for the extended SS model ($M = 2.49$, $SD = 3.55$) did not differ significantly from those DPSD1 model ($M = 1.88$, $SD = 3.72$), $V = 1542$, $p = .20$. The AIC difference for each model was greater than 10 in

a small proportion of cases; specifically, in 5.56% of fits for the SS model and 5.56% for the DPSD1 model. More often than not, both models had AIC differences under two: $\Delta_{SS} \leq 2$ in 62.5% of cases, and $\Delta_{DPSD1} \leq 2$ in 72.2% of cases, indicating that both models had substantial empirical support and neither gave a meaningfully lower quality of fit to the data than the other. Taken together, these results indicate that both models accounted for most participants' data comparably well, and there was not sufficient evidence that either model fitted reliably better than the other.

3.2.4 Discussion

The results of the present experiment showed no difference in the key identification metric $M - H$ between high and low item characteristic variability conditions. The SS model's recognition parameter estimates showed the same pattern of estimates as the UVSD model in the previous experiment; variability in old item strength increased in high variability conditions, but so did discriminability between old and new items. This shows further evidence for the strength scaling trend first identified by Spanton & Berry (2020). By contrast, the DPSD1 model gave a different ordinal pattern of parameter estimates than the DPSD model in the previous experiment. Contrary to previous model fits, estimates of R_h and R_l did not differ in the present study, and nor did measures of d_h and d_l . Based on fitted values of these key parameters, both models did not predict the lack of difference in values of $M - H$ between variability conditions. The SS model predicted that $M - H$ would be greater in the high variability condition, and the DPSD1 model in the low variability condition, although the magnitude of these predicted mean differences was small. Regardless, neither model's qualitative prediction was observed in the data.

Both models gave roughly comparable quantitative fits to the data, although the DPSD1 model fitted best to a greater proportion of individual datasets than the SS model (55.6% of datasets, vs. the SS model's 44.4%). Both models had a small percentage of very poor fits to the data (5.56%) compared with the other, according to analyses of AIC differences. This meant that although both models gave comparable quantitative fits to the data, the DPSD1 model fitted best to a greater proportion of datasets. Taken with previous results, this puts the quantitative fit of both models on an even standing. Although both models gave an equivalent quality of fit in Experiment 1 of the previous chapter, Experiment 2 found evidence for the SS model's superiority. Berry et al. (2012) also found that when SS and DPSD1 model fits were compared in two experiments, each model fitted best to data from one experiment each. These varying results show the need to test competing predictions from each model in order to discriminate them.

Despite not predicting an absence of a difference in $M - H$ between conditions, the SS model was supported by the results of exploratory identification level comparisons. The model predicts that the mean identification level for correct rejection responses should be greater than that for false alarms and misses in both high and low-variability conditions. The new-old priming effect was reliably greater than the difference between the mean identification level for correct rejections and misses, which the SS model also predicted. These results can also be accounted for by the DPSD1 model specification, and so do not differentiate the models in the present experiment. However, they align with previous results from Berry et al. (2012) as evidence in favour of the SS model over a strict multiple-systems (MS1) model, and so contribute to theory development in the memory-systems debate.

It is possible that the strength of the item characteristic variability manipulations in the present experiment prevented the SS and DPSD1 models from making opposing predictions. Despite manipulating multiple item characteristics at once to achieve a meaningful effect on recognition confidence ratings, this only resulted in an R^2 difference between conditions corresponding to a small to medium effect size (Cohen, 1988). This effect size may have been too small to provoke the models to make opposing predictions that were detectable at the sample size attained in the present experiment. This difficulty to find a strong item characteristic manipulation has also pervaded previous attempts to manipulate encoding variability (Spanton & Berry, 2020, 2022). Future efforts to test SS and DPSD1 model predictions about encoding variability should seek to find a manipulation that has a stronger effect on recognition confidence ratings in order to mitigate this issue.

3.3 General Discussion

The experiments in this chapter showed that single-system and dual-process models of recognition memory and priming could once again not be distinguished based on their unique predictions about identification performance. Extended UVSD and DPSD models of recognition gave promising patterns of parameter estimates in Experiment 3, with each model predicting increases in overall memory strength and old item variance for words with high variability in item characteristics. This justified the extension of the SS and DPSD1 models of recognition and priming, as simulations showed that doing so could reveal their competing predictions about identification performance. However, parameter estimates in Experiment 4 showed that the SS and DPSD1 models could not be distinguished on the basis of these predictions. While the SS model gave a pattern of estimates

consistent with those in Experiment 3, the DPSD1 model did not, revealing no strong evidence in support of either model. Indeed, using fitted values in each model's expected value functions showed that neither model predicted the absence of a difference between $M - H$ between variability conditions that was observed in the data. This outcome, like the results of Chapter 2, shows that although single-system and dual-process models of recognition and priming make competing predictions about the relationship between explicit and implicit memory, using these predictions to make strong inferences about their validity is challenging in practice.

Although the results of Experiments 3 and 4 did not prompt newly identified predictions from the models in question, they provide insight into the encoding variability hypothesis. Although it has been suggested that encoding variability causes the old item variance effect (Wixted, 2007), previous research has suggested that it cannot solely account for the UVSD model's unequal variance assumption (Koen et al., 2013; Spanton & Berry, 2022). Experiment 3 showed that increasing item characteristic variability led to greater old item variance while new item variance remained constant. However, this coincided with increased discriminability measures in high variability conditions. In Experiment 4, the SS model's recognition parameters reproduced this pattern of results. These findings show that encoding variability does not solely explain increases in old item variance, and simultaneous increases in overall strength (discriminability) necessarily accompanied these changes. Taken with previous results (Spanton & Berry, 2020, 2022), this gives evidence that overall strength and old item variance scale with one another in unequal variance signal detection models of recognition memory.

The implications for the dual-process explanation of the old item variance effect are less clear. In Experiment 3, the DPSD model predicted an increase in recollection with greater levels of item characteristic variability. Since recollection affects both overall strength and the distribution of old item strength values in the DPSD model (Spanton & Berry, 2020), this behaviour was consistent with a strength scaling account. In Experiment 4, the DPSD1 model did not give the same pattern of recognition parameter estimates. There was no difference in recollection or discriminability between high and low conditions, although μ parameters were lower in high variability conditions. The DPSD1 model therefore did not predict that strength scales with the old item variance, in contrast with the results of Experiment 3 and the SS model's pattern of parameters in Experiment 4. The cause of this discrepancy could be investigated in further research to understand how the psychological predictions made by the DPSD1 model can change in response to different item characteristic manipulations. Regardless, the present results do not refute a dual-process view of recognition and priming.

While the present results did not give strong support for either the SS or DPSD1 model, their competing predictions about identification performance for hits and misses still stand, and it is possible that a different experimental manipulation could test these more successfully. For instance, it is possible that manipulating other facets of the study phase might have a greater effect on old item variance, compared to item characteristics. Wixted (2007) defined old items in the context of the encoding variability hypothesis as "lures that have had memory strength added to them by virtue of their appearance on the study list". This description implies that any variable that affects memory strength during the study phase could be used to manipulate encoding variability. But, other popular theoretical frameworks might not assume that item characteristics affect memory strength in this way.

For instance, the REM model (Shiffrin & Steyvers, 1997) assumes that characteristics like word frequency affect memory strength, and successfully predicts mirror effects on hits and false alarms as a result of lowering word frequency (Malmberg et al., 2002). However, REM does not assume this effect is caused by word frequency influencing the formation of memory traces during encoding. Rather, item characteristics like word frequency only influence the probability of matching feature values during retrieval (Shiffrin & Steyvers, 1997) and not initial feature storage. If this is true, it could explain why encoding variability was not found to solely influence estimates of old item variance in Experiments 3 and 4. Other manipulations that directly affect feature storage in REM are possible but require further consideration. For instance, REM predicts that increased study time allows more chances for features to be encoded (Shiffrin & Steyvers, 1997). Other endogenous variables such as attention could feasibly allow for this, or increase the probability of features being stored without error. However, previous attempts to add variability to these factors by Spanton & Berry (2020) resulted in even weaker effects on recognition confidence ratings than the present manipulations.

Despite the issues that arise when testing models of recognition memory and priming, there are also many other ways to approach the broader question of whether explicit and implicit memory are driven by multiple systems. For instance, instead of focusing on recognition as a form of explicit memory, one could compare performance in priming and a recall task. Recall is often thought to rely almost solely upon explicit memory (Yonelinas, 2002; Quamme et al., 2004), rather than including implicit memory processes (but see Mandler, 1980; Ozubko et al., 2021). This means that recall is closer to being a pure expression of explicit memory than recognition, and therefore would more strictly reflect the assumptions of single-system and dual-process accounts. It would also circumvent the difficulties of having to selectively manipulate recollection and familiarity in

the same recognition task, as attempted in the present and previous chapters. For these reasons, the following chapters focus on the relationship between implicit memory and performance in recall tasks.

To conclude, I conducted two experiments that manipulated the variability in recognition memory strength for old items, by way of item characteristics. Experiment 3 did so in a design where only recognition memory was tested, and extended versions of the UVSD and DPSD models were fitted to these data. Both models showed an increase in both old item variance and discriminability in high item characteristic variability conditions, albeit using different parameters. Simulations using the SS and DPSD1 models confirmed that these patterns of estimates would prompt the models to make opposing predictions about identification performance for items that receive hit and miss responses. To assess these predictions, Experiment 4 extended the previous method to include identification judgements alongside recognition memory tests. However, the results of this experiment were not able to distinguish the SS and DPSD1 models. Following from this, alternative methods for testing the single-system and dual-process accounts of explicit and implicit memory should be explored, including those that test recall, rather than recognition memory.

Chapter 4

Continuity Between Cued Recall and Priming

Although the SS model was developed to explain recognition and priming data, its predictions have also influenced studies on the relationship between recognition, priming, and cued recall. In their first experiment, Mazancieux et al. (2020) had participants study compound stimuli of famous faces against background images. They were then tested on their perceptual identification and recognition memory of the famous faces, and their recall of the background paired with each face they judged "old" in the recognition task. Identification performance was better for items that were correctly recalled than for those that were not, giving evidence for a positive association between explicit and implicit memory performance. In a second experiment, Mazancieux et al. (2020) used a similar procedure to demonstrate an association between identification performance for a stimulus and the number of semantic properties a participant was able to recall about that stimulus. These results are compatible with a single-system account that assumes cued recall and implicit memory for a given item arise from the same memory strength source. In this account, greater strength for an item leads to an increased likelihood of it being correctly recalled and a greater repetition priming effect.

The rationale given by Mazancieux et al. (2020) for extending the single-system account to recall is that it gives a purer measure of explicit memory than recognition. While recognition memory judgements can be based on recollection, they can also be influenced by more implicit feelings of familiarity toward a stimulus in the absence of explicit contextual details (Mandler, 1980). This is thought to be especially reflected in low-confidence recognition judgements, where participants use familiarity to make their recognition judgements because they cannot remember strong episodic details from the study phase. The distinction between these recollection and familiarity processes has in itself been a subject of intense debate in the recognition memory literature (Yonelinas, 2002; Yonelinas et al., 2022). By contrast, it is less likely that participants can use familiarity to reproduce

contextual information associated with a cue or open-ended prompt in recall (Mazancieux et al., 2020). This is because the target item itself is absent in a cued recall test, making it harder to recall from feelings of familiarity with the cue item. Any resultant association between measures of cued recall and implicit memory is therefore unlikely to be a consequence of implicit memory contaminating recall performance. Testing implicit memory and recall for the same items is, therefore, an ideal way to assess the predictions of the single-system account (Mazancieux et al., 2020).

However, there is evidence that implicit memory processes can influence recall in certain tasks. Ozubko et al. (2021) found that a notable proportion of words that are generated in semantic cued recall tasks are not recognised immediately afterwards. They concluded based on ERP analysis that these recallable but not recognisable words were driven by semantic priming rather than an explicit memory process, suggesting a role of implicit memory in some cued recall decisions. Other research has shown that in remember/know tasks, a large percentage of items receive "know" responses in cued recall (Uner & Roediger, 2018; Hamilton & Rajaram, 2003). As these responses indicate feelings of familiarity for a stimulus without concrete contextual information, it is possible that memory for items in cued recall paradigms is not purely explicit. Taken together, these results show that, like recognition, cued recall may be influenced by implicit memory processes.

Although these results arguably support the influence of implicit memory processes in recall, they do not necessarily detract from Mazancieux et al. (2020)'s conclusions. Evidence from remember/know experiments places weight upon subjective reports from participants which are ultimately unverifiable and may not cleanly correspond with latent explicit and implicit memory processes. Some of the influence of implicit memory upon cued recall also appears to depend upon the semantic relationships between cue and target items. For instance, Ozubko et al. (2021) tested cued recall for word pairs with strong semantic links like "right-left" and "princess-king". It is therefore plausible that participants may rely partly on guessing recall targets using these semantic associations rather than retrieving memory traces from the study phase. Ozubko et al. (2021) found no reliable evidence of recallable but not recognisable words in their free recall task, adding credence to the idea that non-explicit influences upon recall depend upon the presence of a semantically associated cue. In their first experiment, Mazancieux et al. (2020) tested episodic cued recall for stimulus pairs without strong semantic associations. It is therefore likely that their cued recall measure was not contaminated by substantial implicit influence. At the very least, this measure likely provided a purer measure of explicit memory than recognition; a task more generally considered susceptible to implicit influences.

With these considerations in mind, Mazancieux et al. (2020)'s demonstration of an association between episodic cued recall and identification leaves possibilities for further study. Their use of famous face and background scene pairings raises the question of whether identification for cues and recall for paired items are associated when testing participants on common pairings of the same stimulus type, such as objects or words. Mazancieux et al. (2020) also did not randomise the face-background pairings in their experiments; each famous face was paired with the same unique background across all participants. It is plausible that some face-background pairs could have had stronger semantic associations than others, and that randomising these pairs would have controlled for any influence of non-explicit memory processes better¹. Testing for continuity between cued recall and implicit memory performance in a new method will provide a conceptual replication of the results of the first experiment reported by Mazancieux et al. (2020). If their main result is replicated, this will add validity to a single-system account of explicit and implicit memory over and above a strict multiple-systems alternative. This can guide further model and theory development. The following experiments address these points, providing a conceptual replication and extension of Mazancieux et al.'s (2020) finding that episodic cued recall and identification performance are positively associated. Experiment 5 investigates the relationship between priming, recognition, and cued recall for randomised object pairings, following a method similar to Mazancieux et al.'s (2020) first experiment. Experiment 6 simplified this method, testing whether the relationship between cued recall and identification performance for studied items holds when recognition is not tested and new items are not presented. To preface the results, both experiments showed better identification performance in trials with correct cued recall, versus those with no attempted recall.

4.1 Experiment 5

The aim of Experiment 5 was to conceptually replicate the key findings of Mazancieux et al. (2020). In their first experiment, participants studied face-background pairs and then gave three responses in the test phase to a mix of studied and new faces. In each test trial, participants gradually unobscured and attempted to identify a blurred face at the lowest level of clarity possible in a perceptual identification task. They then made old/new recognition judgements about the face, before attempting to recall its paired background if they responded "old" in the recognition test. While the present method adopted this same approach, some details differed. Firstly, we used pairs of object pictures rather than famous face-background compound stimuli. This ensured that the

¹However, it is likely that this effect was accounted for to some degree because Mazancieux et al. (2020) included item-level random effects in their analyses.

stimuli being identified and recalled were of the same type and that our results reflected memory for everyday items rather than more artificial scenarios. Object pairings were also randomised across participants, in contrast to the static pairings used by Mazancieux et al. (2020). This was intended to minimise the chance of particular pairs being either consistently semantically related or easier to recall than others across participants. Finally, stimuli in the present identification task were obscured by white squares, not Gaussian blur. This is because Gaussian blur would have preserved the outline of the object stimuli, which could have enabled early identification. This was not an issue for Mazancieux et al. (2020), whose cue items were face stimuli with the same oval outline shape.

With this method, we were able to investigate whether identification performance differed for trials with correct recall, incorrect recall, and no recall. A single-system account of explicit and implicit memory assumes that cued recall and identification in a priming task share a common memory strength signal. This leads to the prediction that participants will identify cue items better — at a lower level of image clarity — when they can correctly recall the item paired with the cue, compared with trials where recall is incorrect or absent. This account also predicts greater identification performance in trials with incorrect recall than in trials with no recall, under the assumption that making an incorrect recall response requires more memory strength than producing no response.

4.1.1 Method

Participants

80 participants (34 males, 45 females, 1 other/non-binary), with a mean age of 37.01 ($SD = 11.98$) took part in this experiment in return for a £7.50 cash payment. Participants were members of the public residing in the United Kingdom, recruited through the Prolific research participation platform. Assuming 80% power, the sample size allowed for the detection of an effect size $d_z = 0.32$ in a paired-sample t -test; an approximation of the pairwise comparison in our linear mixed effects model.

Materials

The stimuli were 90 desaturated images of objects and animals from the BOSS Phase II stimulus bank (Brodeur et al., 2014), each measuring 400×400 pixels. For the fragment identification task, each image was partially obscured by white squares at 15 levels of fragmentation (see Figure 4.1).

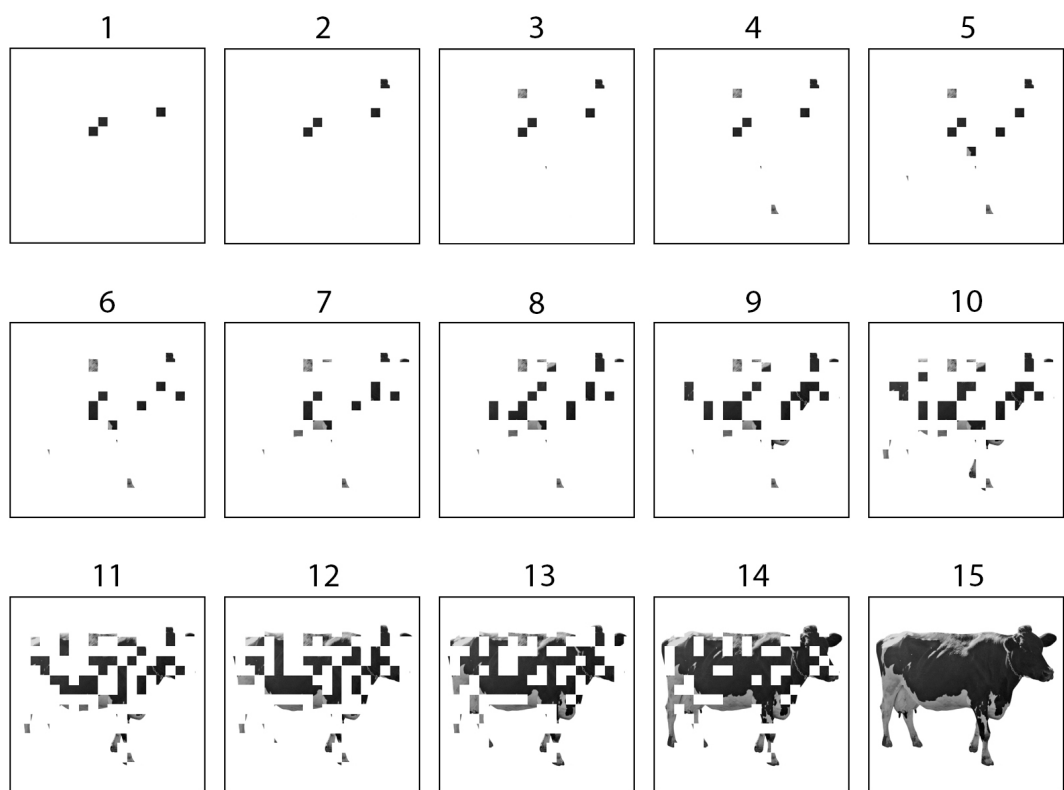


Figure 4.1: An example of a fragmented stimulus used in an identification trial from Experiments 5 and 6. Participants were instructed to identify each stimulus at the lowest level of image clarity possible (denoted by labels 1-15). If they could not identify the stimulus, they could press a button to reveal more of the image until it was completely visible at level 15.

These fragmented images were generated using the same R script as in Experiment 4. The stimulus used in the retention interval was a composite of two nearly identical images of a beach scene, with each image presented side by side. Ten details of the scene on the right differed from their counterparts on the left. Due to the experiment being conducted online, each participant used either their own laptop or desktop computer to complete the procedure. The program for this experiment and all others following was created using the OSWeb functionality of the OpenSesame experiment builder (Mathôt et al., 2012) and deployed on a JATOS server (Lange et al., 2015). The program ran within a browser window at a resolution of 1024×720 pixels.

Procedure

After giving informed consent, participants completed a study phase lasting 30 trials. In each study trial, participants saw a fixation point (“+”) in the centre of the screen for 500 ms, then two objects presented side-by-side for 5000 ms, followed by a blank screen for 500 ms. The objects that appeared in the study phase were randomly selected from the pool of all 90 possible stimuli and randomly assigned to either the left or right side of the screen. This randomisation was unique for each participant. Participants were instructed to pay attention to each object pairing and remember which objects were presented together.

After the study phase, participants completed a short interval task in which they had to spot differences between two nearly identical scenes. Participants were instructed to click on any differences they observed on the rightmost scene. Upon correctly identifying a difference between the two scenes, a white square would appear around the difference. The static prompt “Click on the right side to spot the differences (*N* differences remaining)” appeared at the top of the screen throughout the task, with *N* being replaced with an updating count of the number of differences left to identify. This retention task timed out after 60 seconds. Participants received instructions to wait until the end of the task if they identified all 10 possible differences between the scenes before the time limit.

Participants then started the test phase, which consisted of 60 trials. All 30 studied cue items were presented across these trials, intermixed with 30 new items. Each test phase trial included a fragment identification task, followed by a recognition decision, and then a paired recall task. In the fragment identification task, participants first saw a studied or new item obscured at the lowest level of image clarity (see Figure 4.1). If they could not identify the item, they would press the ‘R’ key to see the next clearest version of the item, revealing more of the object with each response.

They were instructed to identify the image at the lowest level of clarity possible, therefore pressing 'R' as little as possible. When they were able to identify the item, they pressed the 'N' key to name it. A static reminder of these keys was presented at the bottom of the screen until the participant pressed 'N', at which point a text box appeared in its place. The prompt "What is the name of this item? Please type your response in the text box below. Press Enter to submit your response." was presented near the top of the screen when the text box appeared.

Upon typing their response and pressing Enter, the participant was shown the clarified stimulus at the centre of the screen and asked to judge whether they had seen it before in the study phase. Participants made this recognition response by pressing one of 6 number keys according to the following response scale: "1 = High Confidence New, 2 = Medium Confidence New, 3 = Low Confidence New, 4 = Low Confidence Old, 5 = Medium Confidence Old, 6 = High Confidence Old". This scale was presented as a static prompt at the bottom of the screen. The phrase "Is this item New or Old? Please respond by pressing one of the keys below:" was displayed as a prompt at the top of the screen. Participants had an unlimited amount of time to make their recognition response. After typing their response, participants progressed to the cued recall judgement. The unobscured image of the item they had just identified was presented in the middle of the screen, with a text box beneath. The prompt "Which item was paired with the picture below? Please type your response in the text box below. Press Enter to submit your response." was displayed near the top of the screen. The test phase instructions encouraged participants to try their best to guess the paired object, but that it was permissible to give no response if they were certain that the cue item was "new". Once participants typed their response into the text box and pressed enter, the next test phase trial would display. After completing every test phase trial, participants reported their age and gender before being debriefed.

4.1.2 Results and Discussion

The *lme4* package (Bates et al., 2015b) was used for linear mixed model analyses. The means and standard errors reported in these mixed model analyses were marginal estimates computed using the *emmeans* package (Lenth, 2021). We applied Bonferroni corrections to pairwise comparisons in all experiments unless otherwise noted.

In each experiment in Chapters 4 and 5, linear mixed models were fitted to the data. There are several approaches to fitting linear mixed models in the literature. One procedure detailed by Singmann & Kellen (2019) involves specifying the maximal complexity model structure warranted

by the data, then simplifying the model until it converges without errors or statistical concerns. Another procedure uses principle components analysis to determine the random effects structure of the final model (Bates et al., 2015a). This was the approach taken by Mazancieux et al. (2020). The present analyses used a "bottom-up" approach to determine the final random-effects structures of each mixed model. Starting with a fixed-effects-only model, I iteratively increased the complexity of the random effects structure, evaluating each improvement in model fit with likelihood-ratio (LR) tests. The final model specification was determined when additional random effects components either did not converge or did not significantly improve the model fit.

This approach is justified for several reasons. Evaluating increases in model complexity using LR tests gave an explicit statistical justification for including each random-effects component. With a view to parsimony, this approach ensured that each random effects component played a substantive role in explaining variance in the data. This helped to avoid situations where extra random effects components did not significantly increase the goodness of fit but still allowed the model to converge. One drawback of this procedure is that conducting many LR tests on successive models inflates the Type-I error rate for model selection. To address this, each LR test was evaluated against a more stringent alpha level of $\alpha = .001$ to reduce the chance of Type-I errors affecting model selection.

Task Performance

We excluded trials in which items were not identified correctly; these made up 6.40% of all trials. To assess the presence of a priming effect, the difference between the mean identification level for new and old items was calculated for each participant. Priming ($M = 0.53$, $SE = 0.05$) was significantly greater than 0, $t(79) = 10.47$, $p < .001$, 95% CI [0.43, 0.63], $BF = 4.54 \times 10^{13}$, $d_z = 1.17$, indicating that old items were identified at a lower level of image clarity than new items.

We calculated the proportion of each type of recall response (correct, incorrect, no recall) per participant. A recall response was coded as correct if it matched with the relevant paired item from the study phase. Minor misspellings where it was clear that the participant had identified the target item (for example, responding "bookshelv" where the correct pair was "bookshelf") were allowed. An incorrect recall response was defined as when a participant attempted to recall a given item, but their response did not clearly match the identity of the relevant paired item (for example, responding "bucket" where the correct pair was "bookshelf"). A "no recall" response was coded when the participant either did not attempt to recall the relevant stimulus or responded with nonsense letters. The mean proportion of correctly recalled items per participant was .31 ($SE = .03$). The mean

proportion of incorrectly recalled items was .38 ($SE = .03$). The mean proportion of items that were not recalled was .31 ($SE = .04$). Recall performance was, therefore, higher than in Mazancieux et al. (2020), where the mean proportion of correct recall responses was .13 in their Experiment 1. To measure recognition memory performance, d' was calculated for each participant using the formula $d' = \Phi(H) - \Phi(FA)$ where Φ is the inverse cumulative normal distribution function and H and FA are the corrected hit and false alarm rates using the correction given by Snodgrass & Corwin (1988). Values of d' ($M = 2.47$, $SE = 0.13$) were significantly greater than 0, $t(79) = 19.76$, $p < .001$, 95% CI [2.22, 2.72], $BF = 1.09 \times 10^{29}$, $d_z = 2.21$, confirming that recognition memory performance was above chance.

Recognition Performance and Identification

We tested whether participant recognition responses (old or new) and the actual state of items (studied or non-studied) predicted identification level using a linear mixed model with participant and cue item level random slopes and intercepts for item state. This analysis revealed a significant effect of item state, $t(164.42) = 7.01$, $p < .001$, showing that studied items ($M = 12.01$, $SE = 0.17$) were identified at a lower level of image clarity than non-studied items ($M = 12.46$, $SE = 0.16$). This indicates that there was a priming effect even when accounting for participant and cue item-level variability. A significant effect of recognition response was also found, meaning that items judged old ($M = 12.15$, $SE = 0.16$) were identified at a lower level of image clarity than items judged new ($M = 12.33$, $SE = 0.16$), $t(2432) = 3.19$, $p = .001$. There was no significant interaction between these two factors, $t(3276.67) = 0.40$, $p = .69$. This meant that the identification advantage for items judged old was equivalent within studied and non-studied items.

Recall Performance and Identification

Turning next to the main issue of whether there was a greater priming effect for recalled versus non-recalled items, we compared identification for each type of recall response in a linear mixed model with random intercepts for participants and cue items. The mean fragment identification levels for each participant per recall response can be found in Figure 4.2. Pairwise comparisons revealed that cue items were identified at lower levels of image clarity when their paired item was correctly recalled ($M = 11.84$, $SE = 0.17$), compared to not recalled ($M = 12.12$, $SE = 0.17$), $t(2186) = -3.096$, $p = .006$. Cue items were also identified at lower levels of image clarity when their paired item was incorrectly recalled ($M = 11.91$, $SE = 0.17$), rather than not recalled, $t(2180) = -2.54$, $p =$

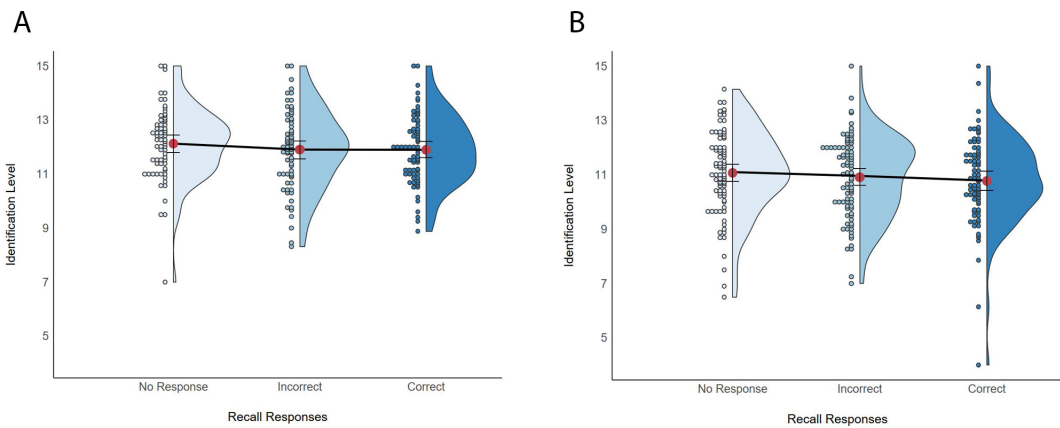


Figure 4.2: Mean fragment identification level per participant for each recall response type in Experiments 5 and 6. The red dots denote the mean of means, and the whiskers the 95% confidence interval.

.03. There was no significant difference in identification performance when comparing correctly recalled and incorrectly recalled items, however, $t(2156) = -0.82, p \approx 1$.

Like Mazancieux et al. (2020), we found that identification was better in trials where correct recall responses were made, compared to those where no recall response was attempted. We also found an improvement in identification in trials where an incorrect recall attempt was made, as opposed to no attempt. Both of these findings were also observed by Mazancieux et al. (2020). These results provide evidence of a relationship between cued recall and identification performance, indicating that these measures of explicit and implicit memory are associated. We also reproduced the finding that identification is better for items that were judged old in a recognition task versus those judged new. This result is consistent with previous literature that supports the SS model of recognition and priming (Berry et al., 2012). These results show that explicit memory responses are not independent of the repetition priming effect, which is consistent with a single-system account of explicit and implicit memory.

4.2 Experiment 6

Following the previous results, Experiment 6 investigated whether the relationship between cued recall and identification performance held when recognition memory was not also tested. When testing identification before recognition memory, it has been hypothesised that early identification can act as a cue for participants to assume they have seen an item before (Jacoby & Dallas, 1981). This can lead to participants giving higher recognition confidence ratings to certain items because of fluency (Lange et al., 2019). If a participant is confident that they have seen an item before, they

may alter their behaviour when attempting to recall its paired item. For instance, they could try harder to remember the paired item because they are certain that its cue was old. In this case, the association between identification and cued recall performance could be bolstered by the inclusion of a recognition test. To investigate this possibility, I tested whether correctly recalled items are better identified than items that are not recalled in an experiment without new items or a recognition memory test. I reasoned that, since all items were studied, participants would be less likely to engage in the process of recognition for a given item, reducing the potential for recognition to impact recall.

4.2.1 Method

Participants

96 participants (14 male, 82 female) with a mean age of 22.01 ($SD = 6.75$) completed the experiment in return for participation credit that counted towards a pass/fail component of their course. Participants were undergraduate psychology students at the University of Plymouth. The sample size was justified as the number of participants needed to detect an effect size $d_z = 0.29$ in a paired-sample t -test, assuming 80% power.

Materials and Procedure

The stimuli were 60 desaturated images of objects or animals. These were taken from the BOSS Phase II stimulus bank (Brodeur et al., 2014), each measuring 400×400 px. The fragmented versions of the stimuli were generated in the same way as in Experiment 5. Each participant used either their own laptop or desktop computer to complete the procedure. The experiment was displayed within a browser window at a resolution of 1024×720 pixels.

The procedure for this experiment was very similar to that of Experiment 1. Like the previous experiment, there were 30 study trials in total. The test phase was also 30 trials long, with participants attempting to identify a studied cue item and then recall its paired item in each trial. The identification procedure was identical to that in Experiment 5. Immediately after submitting an identification response, participants then made their recall judgement in the same way as in Experiment 5. Once the participants completed the test phase, they input their age and gender into the experimental program before viewing a full debrief.

4.2.2 Results

Task Performance

Before performing other key analyses, trials with incorrect identification judgements were excluded across the sample; these comprised 3.82% of all trials, meaning that 96.18% of all identification trials were correct. The proportion of each type of recall response (no response, incorrect recall, correct recall) was calculated. The mean proportion of correct recall responses across participants was .33 ($SE = 0.03$); the mean proportion of incorrect recall responses was .35 ($SE = 0.03$), and the mean proportion of no recall responses was .32 ($SE = 0.03$).

Recall and Identification Performance

As in the previous experiment, we assessed the effect of recall response in a mixed model with random intercepts on the level of participants and cue items. The participant-level means for each recall type are displayed in Figure 4.2. This analysis revealed that correct recall ($M = 10.8$, $SE = 0.21$) was associated with an earlier fragment identification than no recall ($M = 11.1$, $SE = 0.21$), $t(2685) = -3.97$, $p < .001$. However, there was no significant difference between fragment identification for items that received a correct recall with an incorrect recall ($M = 11.0$, $SE = 0.21$) response, $t(2712) = -1.96$, $p = .15$. There was also no significant difference between incorrect and no recall items in terms of fragment identification, $t(2705) = -1.71$, $p = .26$.

4.2.3 Discussion

Like Experiment 1, these results suggest identification was better for correctly recalled items than for non-recalled items. However, further differences in identification for incorrectly recalled and non-recalled items were not found. This comes in contrast with Mazancieux et al. (2020) and the results of Experiment 5. It is therefore possible that incorrectly recalling an item only benefits identification when recognition memory is also tested. The specific mechanism behind this is unclear, however. The removal of the recognition test did not alter the proportion of incorrectly recalled items in Experiment 6 when compared with Experiment 5. A two-sample t -test revealed no difference between these proportions, $t(173.69) = 0.70$, $p = .49$, 95% CI [-0.06, 0.12], $BF = 0.20$, meaning that the presence of the recognition test in Experiment 5 did not prompt participants to incorrectly guess more during recall. Regardless, the key finding that identification was better in trials with correct recall versus no recall is still of theoretical significance, providing evidence in

favour of a single-system account, and weighing against a strict multiple-systems account.

4.3 General Discussion

Experiments 5 and 6 tested whether identification performance could be predicted by cued recall performance on an item level, as in Mazancieux et al. (2020). Experiment 5 followed Mazancieux et al.'s (2020) method closely by testing identification, recognition, and cued recall for the same item pairings. Experiment 6 tested cued recall and identification for studied items, removing the recognition test and new items from the method. Both experiments found better identification for cue items in trials where pairs were correctly recalled, versus not recalled. Experiment 5 also found an identification benefit in incorrect recall trials compared to those where recall was not attempted, though there was no reliable evidence for this effect in Experiment 6. This replicated the key pattern of results in Mazancieux et al.'s (2020) Experiment 1. The recognition memory results in Experiment 5 also followed theirs, showing evidence that items judged old were better identified than those judged new. This shows that, on balance, there is a consistent increase in perceptual identification performance where cued recall is correct, compared with where it is not attempted. This weighs against a strict multiple-systems interpretation of cued recall and implicit memory, which would predict no relationship in performance between the two tasks. Instead, the results can be more parsimoniously explained by a single-system theory of explicit and implicit memory where cued recall and implicit memory share a common memory strength source.

Despite there being evidence that identification improved in correct recall trials versus no recall trials, there was no indication of an implicit memory benefit in correct versus incorrect recall trials. It seems intuitive that producing a correct recall response requires a greater degree of memory strength for the cue-target pairing than producing an incorrect recall response. If this is the case, then the lack of an identification benefit for correct versus incorrect recall trials may be at odds with a single-system account. However, some incorrect recall responses may arise not from a lack of memory strength, but from other processes relating to the retrieval of the recall response, such as attention or language production. This would allow for recall to share a common strength source with implicit memory, but for the association between performance in these tasks to be diluted by variables that are not memory related. This would also explain why the identification benefit for correctly recalled versus non-recalled items was numerically small, despite being reliably observed. Since the present experiments do not evaluate model-based representations of single-system and dual-process theories, it is not possible to conclusively validate this explanation's compatibility with

either account. Future conjoint models of explicit and implicit memory could formally represent the role of non-mnemonic factors to measure the extent that they affect the relationship between cued recall and implicit memory compared with shared memory sources. This has proved useful in quantifying the role of measurement sensitivity in driving dissociations between recognition and priming (Berry et al., 2008a,b), and so has promise in further model development.

While the present experiments validate Mazancieux et al.'s (2020) key finding that there is a relationship between episodic cued recall and identification performance, there is potential for their results on semantic cued recall to be extended by further research. In their second experiment, Mazancieux et al. (2020) had participants study famous faces, before identifying them and making a recognition judgement as in their Experiment 1. After the recognition judgement, participants were asked to recall semantic details about the famous face, such as the first and last name of the individual, their nationality, and their profession. Mazancieux et al. (2020) found that participants identified faces better when they could recall more semantic features, consistent with a single-system view and previous research showing implicit influences in semantic cued recall (Ozubko et al., 2021). However, this semantic recall measure may have been susceptible to implicit influence, since participants could feasibly guess one or more of the semantic features from the stimulus or use familiarity as a cue to retrieve those features. For instance, a participant might guess the nationality of a famous face from its appearance, despite not having that knowledge before the experiment and therefore not retrieving any semantic information specific to the identity of the stimulus from memory. The association between implicit and explicit memory in Mazancieux et al. (2020)'s Experiment 2 may therefore have been driven by guessing or implicit contamination in the semantic recall measure, and not as a result of recall and priming sharing a common memory source.

Further research could address this possibility by investigating the role of guessing in driving implicit contamination in cued recall. In one condition, participants could study stimuli with strong semantic links, and then be tested on their recognition of the cue items, and their recall of the paired items. In another condition, the participants could complete the same method, but where a mask is presented too quickly for conscious perception during study, in place of any items. This would give participants the awareness that they have seen a stimulus in each study trial, but, as in Conroy et al. (2005), it would prevent them from encoding information relating to the test phase in memory. This would result in chance-level recognition memory and recall responses driven wholly by semantic association in the absence of memory from the study phase. Ozubko et al. (2021) found that when semantic priming drives cued recall, items can be recalled but not recognised. However, if this

phenomenon is simply due to semantic association, the rate of recognition failures in correct recall should not differ between conditions. If this is the case, any association between semantic cued recall and identification performance is less likely to be a result of a common memory process. Assessing this possibility is important for evaluating the degree to which results like those from Mazancieux et al. (2020)'s Experiment 2 support a single-system view of explicit and implicit memory.

Alternatively, the possibility of guessing a paired item based on semantic association or implicit contamination could be avoided by using a different recall task. In free recall, participants do not attempt to retrieve stimuli based on item-specific cues, but rather a general prompt to recall as many items as possible from a prior context. This mitigates the possibility that participants might use familiarity or guesswork to generate targets from a cue because the prompt is generic and has no relationship with any given stimulus. Generating a correct free recall response is therefore assumed to require specific episodic information about a stimulus from the study context and can therefore be seen as the purest expression of explicit memory. The experiments in the following chapter investigate the association between free recall and implicit memory as a strong test of the single-system and dual-process accounts.

Chapter 5

Free Recall and Priming

Many tasks thought to involve explicit memory have been studied in conjunction with implicit memory, including cued recall (Mazancieux et al., 2020), recognition memory (Berry et al., 2012), and source memory (Lange & Berry, 2021; Lange et al., 2019; Huang & Shanks, 2021). However, performance in each of these tasks is thought to be susceptible to the influence of implicit memory. Compared with these examples, free recall is an explicit memory task that is relatively uncontaminated by implicit memory processes. Free recall requires producing a word from memory in response to an open-ended prompt, in the absence of any information about specific stimuli. From a dual-process theoretical perspective, it is difficult to ascertain how familiarity or fluency could facilitate its reproduction in recall (Mazancieux et al., 2020). Without any item-specific cues, participants must presumably access the study context to recall an item. A consistent association between free recall and implicit memory would therefore align with a single-system view that a common memory system drives performance in both tasks.

Dual-process theorists often state that recollection alone is responsible for free recall performance (Yonelinas, 2002; Quamme et al., 2004). Dissociations between free recall and implicit memory performance (Paller, 1990; Hunt & Toth, 1990; Roediger & Challis, 1992) also imply some degree of separation between the two tasks. However, there is some evidence that free recall may be influenced by implicit memory. As in cued recall, previous research has shown that freely recalled items often receive "know" responses in remember/know recognition tests (Tulving, 1985; McDermott, 2006; Mickes et al., 2013). This implies that despite recalling these stimuli from memory, participants lack a feeling of recollection of their presentation in a study phase context. McCabe et al. (2011) also found evidence that attentional manipulations have selective effects on remember and know judgements in free recall, concluding that free recall is not only driven by pure

recollection. Despite this evidence, the use of free recall to test the association between explicit and implicit memory can still be justified. As with cued recall, a participant's subjective judgements for recalled items might not cleanly correspond with the memory processes underlying their responses. Furthermore, the contribution of implicit memory to free recall is the least well-documented of all recall paradigms (McCabe et al., 2011), warranting further study. Finally, even if it is not an entirely "process-pure" measure, free recall still gives participants fewer item-specific cues that could lead to contamination by means of non-explicit memory influences than other tasks like recognition or cued recall. Investigating the relationship between free recall and perceptual identification performance for the same items is therefore a comparatively strict test of the single-system account of explicit and implicit memory, and can give unique theoretical insights as a result.

The following experiments study the dependencies between free recall and perceptual identification performance. Experiments 7 and 8 tested whether identification is better for correctly recalled items versus items that were not recalled. Experiment 9 also investigated this question while controlling for stimulus repetition. Like in the cued recall experiments in Chapter 4, a single-system account predicts identification performance should be better for items that were correctly recalled, versus those that are not recalled. Assuming free recall is driven solely by recollection, a strict dual-process account predicts no consistent association between free recall and identification performance.

5.1 Experiment 7

Experiment 7 tested the association between free recall and identification performance. As in Experiments 5 and 6, I asked whether old items that are correctly recalled are identified at a lower level of image clarity than those that are not recalled. However, this question requires additional methodological considerations due to the nature of the free recall task. Unlike cued recall and recognition memory, free recall and identification for a given item cannot be assessed within the same test trial. This is because the participant is not given any item-specific cues to aid their memory in free recall. Instead, they receive an open-ended prompt to retrieve items from a study context. As a result, tests of perceptual identification and free recall must take place successively in separate phases.

To assess priming in this experiment, new items were required to appear in the identification phase. If identification precedes recall, this may lead participants to mistake new items introduced in the identification phase for studied items, resulting in a high intrusion rate of new items being falsely recalled. Reporting a small number of intrusions could even be a good strategy for achieving

a high correct recall percentage in this task. If a participant has reported all of the items they are sure appeared in the study phase, they may also report items that they think appeared during identification. This is because such items are immediately available to the participant and could seem more likely to be correctly recalled compared to any other items the participant would generate. However, if a participant's intrusion rate is comparable to their correct recall rate, this implies that they did not attempt to retrieve items only from the study phase. Instead, they generated as many items from outside of a deliberate study context, and so did not follow the task instructions. As free recall intrusions are already common in many variants of the task (Zaromb et al., 2006), it is important to address this possibility.

To do so, I included two types of new items in the identification phase of this experiment. In each block of the experiment, 15 old items were studied and then identified. A set of 15 new items were also presented twice in the identification phase. This equates the number of presentations of old and new items before recall, as old items are also presented twice before recall (once at study, once during identification). Therefore, comparing recall for twice-presented new items and old items gives a fair test of whether participants are following instructions to recall from the study phase. This design also allows for the comparison of identification performance between old, first presentation new items, and second presentation new items. If identification is better for old items than twice-presented new items and participants have a low intrusion rate, this would indicate an identification benefit for items being intentionally remembered. This would demonstrate a link between implicit memory and items studied as part of an explicit memory task, and so would be accounted for by a single-system theory. Each of these considerations enabled the key comparison of identification for correctly recalled old items versus non-recalled old items.

5.1.1 Method

Participants

90 undergraduates from the School of Psychology at the University of Plymouth (28 male, 60 female, 2 non-binary/other) with a mean age of 21.20 ($SD = 5.34$) took part in this experiment. They participated in return for partial fulfilment of a mandatory research participation requirement on their course. The sample size was chosen to enable detection of a minimum effect size $d_z = 0.30$ at 80% power in a paired-samples t -test, as an approximation of our key pairwise comparison.

Materials

The stimuli used were 60 concrete nouns from Banks & Connell (2022). Thirty of these nouns described animals, and the remaining thirty described tools or utensils. This semantic distinction was intended to minimise any interference between blocks, such as a participant falsely recalling words presented in the previous block. Each word was between 4 and 9 characters in length, with a production frequency of under 0.1. That is, fewer than 10% of participants in Banks & Connell (2022) reported each word when prompted to freely produce examples from its given category. The words were sampled with this requirement to increase the chance that correct recall responses were a product of memory and not generated by guessing from a set of typical animals or tools and utensils. The words were presented in each phase as images that were generated using the *magick* R package (Ooms, 2021). Each word was generated in white monospaced font with a black outline, against a white 600×100 px background (see Figure 3.3). For the fragment identification task, each word was obscured at 10 levels of clarity (1 being the most obscured, and 10 being the wholly revealed word). These fragmented images were generated according to the same process as in the previous experiments. The same stimulus as the previous experiments was also used for the “spot the difference” interval task. The experimental program was displayed through web browsers on Lenovo computers using monitors with 1920×1080 px resolution.

Procedure

The experiment consisted of two blocks, each with a study phase, a retention interval, a fragment identification phase, and a recall phase. The experiment was split into blocks to ensure that the list length in each study phase was short enough to enable participants to make enough correct recall responses for our analysis. One block contained words pertaining to tools and utensils and the other to animals. The order of experimental blocks was randomised for each participant. Within each block, old and new words were assigned at random from the pool of 30 total words in that phase. This randomisation occurred on a participant level; each participant also saw a different random order of trials in each phase.

After being briefed and giving informed consent, participants completed ten practice trials. In each trial, a fixation point (“+”) was presented in the centre of the screen for 500 ms, followed by a word image for 3500 ms. During this presentation window, participants were instructed to judge whether the object described by each word could fit inside a shoebox. To respond yes, they pressed “Z”, and to respond no, they pressed “M”. A static prompt (“Could this thing fit inside a shoebox?”) was

presented above each word, along with a prompt featuring the response keys (“Z = Yes, M = No”) below the word during each presentation window. Participants then viewed a 500 ms inter-trial interval – a blank screen – before proceeding to the next trial. After completing the practice trials, the participants completed the two main blocks of the experiment. The trial-level procedure of the study phase was identical to that of the practice phase, with participants completing 15 study trials per block. After the study phase, participants completed a retention interval with the same procedure as in Experiments 1 and 2.

Participants then completed a fragment identification phase. Each trial contained either an old word from the study phase in that block or a new word that had not appeared during the previous study phase. Each new word was presented twice in this phase, with studied words appearing once, for a total of 45 trials. Each participant saw a different pseudo-random order of trials that ensured that at least two intervening words came between the original and repeat presentations of each new word. In each fragment identification trial, a word was initially presented at the lowest level of image clarity. Participants were instructed to press “R” to reveal more of the word, or “N” to name it. These responses appeared as static prompts on screen throughout the identification phase. Upon pressing “R”, the next most obscured fragmented image was presented, revealing more of the stimulus to the participant. Upon pressing “N”, a textbox replaced the fragmented stimulus and participants typed their identification response, pressing the Enter key to submit it. Participants were then prompted on a separate screen to press Space to continue to the next trial.

Finally, participants completed a recall phase, in which they were instructed to recall as many words as possible from the previous study phase. To recall a word, participants typed it into a textbox. Each word would then appear inside a large white rectangle underneath the textbox after being submitted with a “Space” keypress. Participants had to make at least 10 recall responses to end the recall phase but were instructed to attempt to recall as many old words as possible. This variant of the free recall task was used to ensure participants gave enough recall responses to enable our planned analyses. If they could not recall any more words before the 10-word threshold, they were instructed to make guesses. After surpassing 10 responses and being able to make no more reasonable recall attempts, participants were instructed to press the “Up” arrow key to end the recall phase, and thus, the block. After both blocks were completed, participants submitted their age and gender before being debriefed.

5.1.2 Results

Task Performance

Four participants were excluded from this experiment for making at least 10% more recall intrusions than correct recall responses. This was a predefined criterion made explicit in our preregistration. Before other analyses were performed, trials in which stimuli were not correctly identified were excluded. These comprised of 3.85% of all trials.

To compare the fragment identification level for each item type in the identification phase, a linear mixed model analysis with random intercepts for each participant was conducted on fragment identification level with item type (old, new first presentation, new second presentation) as a factor. Pairwise comparisons on the model output revealed that participants identified new items on their first presentation ($M = 7.69$, $SE = 0.07$) at higher levels of image clarity than for both old items ($M = 7.22$, $SE = 0.07$), $t(178) = 11.87$, $p < .001$, and new items on their second presentation ($M = 7.15$, $SE = 0.07$), $t(178) = 13.51$, $p < .001$. Fragment identification level did not reliably differ between new items on their second presentation and old items, $t(178) = -1.64$, $p = .31$. This implies that participants had better identification performance for items presented twice before the point of identification, regardless of whether those items were old or new.

To test whether participants understood the instruction to attempt to recall only items that were present in the study phase, the proportion of old and new items recalled was compared. Participants made a greater proportion of correct recall responses ($M = .48$, $SE = .01$) than new item intrusions ($M = .16$, $SE = .01$), $t(89) = 15.93$, $p < .001$, 95% CI [0.28, 0.36], $BF = 4.14 \times 10^{24}$. This indicated that, as instructed, participants reliably recalled more items from the study phase than items that were introduced in the identification phase. This suggests that they were source monitoring effectively.

Recall and Identification Performance

To investigate the effects of recall and item type on identification performance, a linear mixed effects model on identification level with recall state (recalled, not recalled) and item type (old, new) as fixed was fit to the data. The random effects structure in this model featured participant-level random intercepts and slopes for item type, and item-level random intercepts and slopes for recall state and item type. Trials with new items that were presented twice were excluded from this analysis. The model revealed a significant effect of item type, $t(81.14) = 11.61$, $p < .001$, meaning

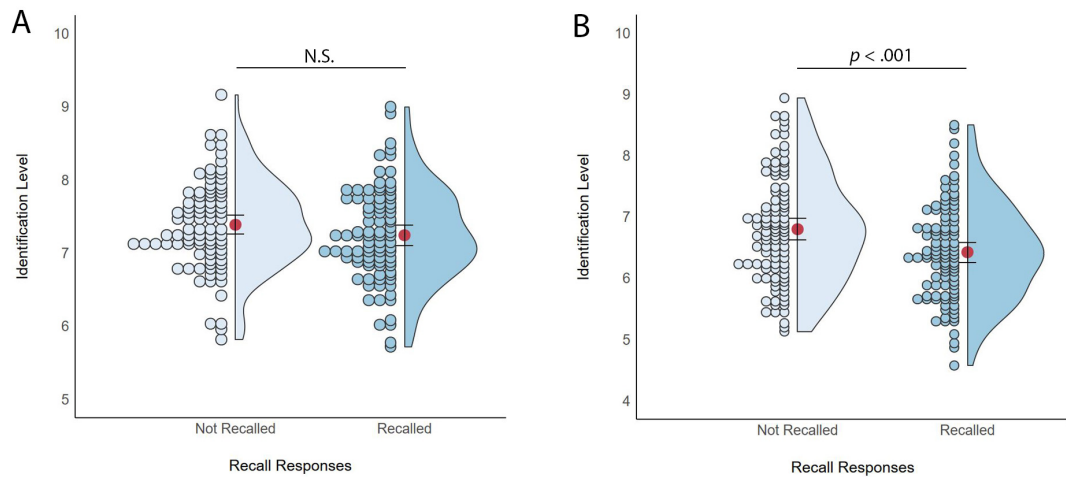


Figure 5.1: Mean fragment identification level per participant for recalled and non-recalled items. Panel A shows results from Experiment 7, and panel B shows results from Experiment 8. The red dots denote the grand mean, and the whiskers the 95% confidence interval.

that identification performance was reliably better for old items ($M = 7.28$, $SE = 0.11$) than for new items ($M = 7.83$, $SE = 0.11$). There was no significant effect of recall state, $t(62.30) = -0.92$, $p = .36$, nor an interaction, $t(4315.33) = 0.05$, $p = .87$. Crucially, pairwise comparisons with a Bonferroni correction showed no significant difference in fragment identification level when comparing old items that were not recalled ($M = 7.30$, $SE = 0.11$) and old items that were correctly recalled ($M = 7.26$, $SE = 0.11$), $t(114.6) = 0.94$, $p = 1$. The mean fragment identification level per participant by recall state is shown in Figure 5.1. In sum, these results show that free recall performance is not associated with identification performance in this experiment.

5.1.3 Discussion

This experiment investigated whether old items that were correctly recalled in a free recall task were identified at a lower level of image clarity than those that were not recalled. Although old items were better identified than new items on their first presentation, showing a repetition priming effect, the same was not true for correctly recalled items over non-recalled items. This contrasts the results of Experiments 5 and 6, where correctly recalled items were better identified than those not recalled. This difference is of theoretical significance. If cued recall, but not free recall, is associated with identification performance in a priming task, this suggests a limit to the single-system account of explicit and implicit memory. It is therefore worth considering whether task differences between the present and previous experiments might explain this pattern of results.

One possibility is that there were differences in the sensitivity of the identification measurements in

the cued and free recall experiments. Identification stimuli in the present free recall experiment only had 10 levels of image clarity, whereas those in the previous cued recall experiments had 15. Therefore, the former experiments could have provided a more sensitive measure of the exact point of identification in each trial, leading to better detection of a subtle association between recall and identification. However, the repetition priming effect sizes in Experiments 5 and 7 (Experiment 6 had no old items and therefore no repetition priming measure) were comparable. Cohen's d_z was 1.17 for this effect in Experiment 5, and 1.25 in Experiment 7. This suggests that, if anything, the present experiment detected a slightly larger priming effect, despite having fewer image clarity levels. As such, it is unlikely that differences in measurement sensitivity were the reason why no association between free recall and identification performance was observed.

Other design differences could have contributed to the divergent pattern of results in Experiment 7. In Experiments 5 and 6, participants identified stimuli and then gave explicit memory judgements immediately after, within the same test trials. Experiment 7 departed from this format by making participants give their identification and free recall responses in separate, blocked phases. This was necessary due to the nature of free recall. Since participants do not receive any item-specific cues to aid their memory in free recall, this task could not have been implemented on a trial level as cued recall had in the previous experiments. Experiments 5 and 6 do not rule out the possibility that the relationship between recall and implicit memory could be caused by a common, non-mnemonic variable. For instance, trial-to-trial variability in attention at test could have driven the relationship between recall and identification performance on a trial level. If a variable like this was dependent upon the testing configuration in the previous experiments, then blocking the recall and identification tests in the present experiment would have minimised its effect. However, since free recall and identification trials must be blocked, any further experiments using these tasks that find a relationship between explicit and implicit memory performance can refute this alternative explanation.

It is also possible that the design of the present experiment added an additional layer of difficulty when compared with Experiments 5 and 6. In Experiment 7, participants identified stimuli before making their explicit memory judgements. This was to avoid a potential confound. If participants recalled items before identifying them, their identification of correctly recalled items might have been artificially improved by simply seeing them on screen directly beforehand. Such a result would have been caused by cross-modal priming and not explicit memory linked with recall. Yet, another consequence of having identification precede recall was that participants had two presentation contexts to choose from whenever they recalled items; the study phase, and the identification phase.

This meant that when participants generated an item for recall, they had to second-guess the context it came from. Since they were instructed to only recall stimuli that appeared in the study phase, the additional source decision introduced by this design could have complicated their recall judgement. It is therefore important that the relationship between free recall and identification performance is tested in a context where it is certain that participants do not have source confusion influencing their recall decision.

5.2 Experiment 8

Experiment 8 simplified the design of Experiment 7 in order to eliminate any additional source memory demands in the free recall task. The new items from the previous design were removed, meaning all items in the identification phase were studied. Participants also recalled items after each study phase and retention interval; the identification phase came after each recall test. This removed the need for participants to second-guess whether each item they recalled appeared in a study phase or a subsequent identification phase. However, since participants attempted to recall items before identifying them, it was important to ensure any association between recall and identification was not facilitated by participants simply seeing some items during the recall phase. To mitigate this risk, participants did not see items on the screen as they entered their recall responses. They only saw asterisks representing each character they typed when recalling a word, as is common in a password entry field. They also did not see any representation of previously recalled words once their responses were submitted. These choices reduced the chance that participants would have better identification of recalled words due to having seen them on screen in the previous phase, rather than as a consequence of the recall process itself. With this design in place, identification performance for recalled than not recalled items can be compared. If items are identified at a lower level of image clarity for recalled items, this is consistent with a single-system view of free recall and implicit memory.

5.2.1 Method

Participants

105 undergraduate participants from the School of Psychology, University of Plymouth, took part in this experiment in exchange for course credit. Five participants were excluded for making correct judgements in less than 60% of study phase trials. This exclusion criterion was included in our

preregistration and was therefore set before data collection began. This left a final sample of 100 participants (84 females, 12 males, 4 other/non-binary). Participants had a mean age of 19.43 ($SD = 1.57$), spoke English as a first language, and had not taken part in Experiment 3. The sample size in this experiment was chosen to allow detection of a minimum effect size $d_z = 0.28$ at 80% power in a paired samples t -test, as an approximation of our key comparison.

Materials

Forty concrete nouns from Banks & Connell (2022) were used in this experiment. Twenty of these described animals, with the remainder describing tools and utensils. Each word was between four and nine characters long and had a production frequency of .05 or below. The words were made into images and then partially obscured in the same way as in Experiment 3. Like the previous experiment, the program was displayed on web browsers using Lenovo computers and 1920×1080 px resolution monitors.

Procedure

As in the previous experiment, there were two blocks. In each block, participants completed a study phase, a retention interval, a free recall phase, a second retention interval, and an identification phase. Words from either category (animals or tools and utensils) were shown in each block. The order of blocks was randomised across the sample, as was the order of trials in each applicable phase.

After viewing a brief and giving informed consent, participants began the first block. They studied twenty word stimuli in the same study phase procedure as in Experiment 8. After the study phase, they completed a "spot the difference" retention interval with the same procedure as the previous experiments. Participants then began a recall phase, where they were asked to freely recall as many words from the study phase as possible. Participants recalled words in this phase by typing them into a text box in the centre of the screen, and pressing the "Space" key to submit each response. Participants saw an asterisk in place of each letter they typed inside the textbox, as in a conventional password entry form. The asterisks disappeared after they submitted each response, and each word was not displayed on screen at any time during the procedure. After participants were certain that they could not recall any more words, they were instructed to press the "Up" arrow key to end the recall phase.

Participants then completed a second interval phase. In each trial, participants were presented a

randomly generated equation of the form $A \pm B \pm C = ?$. There were three unique response options below the equation, with one correct answer and two incorrect answers. The participants were instructed to press the button response (either "A", "S", or "D") that corresponded to the correct answer to the equation. Once a response was made, the next trial began. This procedure continued for 60 seconds, after which participants moved on to the identification phase. The identification trials had the same structure as in Experiment 3. The second block contained the same order of phases. After completing both blocks, participants gave their age and gender and were debriefed.

5.2.2 Results and Discussion

Task Performance

Before any other analyses were performed, trials where the identification response was not correct were excluded. These excluded trials made up 3.36% of all trials. Before these exclusions, the mean proportion of correctly identified items per participant was .97 ($SE = .003$). Five participants were also excluded for poor performance in the study phase. This exclusion criterion was defined in our preregistration as a less than 60% accuracy rate in the study phase size judgement. After exclusions, the mean proportion of correct responses in the study phase was .85 ($SE = 0.01$). Study phase performance did not differ between experiment blocks, $t(99) = -0.90$, $p = .37$, 95% CI [-0.04, 0.02], $BF = 0.16$.

The mean proportion of items recalled across participants was .50 ($SE = 0.01$), and the mean proportion of items not recalled was also .50 ($SE = 0.01$). We also tested whether recall performance differed between blocks, finding no significant difference between blocks and inconclusive Bayesian evidence, $t(99) = -1.84$, $p = .07$, 95% CI [-0.05, 0], $BF = 0.56$. This suggests that participants did not reliably recall a greater proportion of words representing tools and utensils ($M = .51$, $SE = .02$) than animals ($M = .48$, $SE = .01$).

Recall and Identification Performance

We assessed the effect of recall state (recalled, non-recalled) in a linear mixed model with participant-level random intercepts, and item-level correlated random intercepts and slopes. We found a significant effect of recall state, $t(40.52) = -6.33$, $p < .001$, indicating that identification was better for recalled items ($M = 6.42$, $SE = 0.14$) than non-recalled items ($M = 6.82$, $SE = 0.15$). The mean fragment identification level per participant by recall state is shown in Figure 5.1.

Contrary to Experiment 3, these results suggest an association between free recall and identification performance, in that identification performance was better for items that were able to be recalled.

The results of the present experiment showed that participants identified recalled words at a lower level of image clarity than those that were not recalled. This stands in contrast to the results of Experiment 7, where no such difference was observed. Given the similarity of the stimuli and tasks in both experiments, it is worth investigating which factors may have led to these different results. Whereas Experiment 7 had participants complete the identification phase before the recall phase in each block, this order was reversed in Experiment 8 with methodological considerations in mind. However, it opens the possibility that participants identified recalled words more easily as a result of rehearsing them in the recall phase, and not because recall and identification share a common memory source. It is therefore worth testing whether the act of recalling studied information improves subsequent identification to the same degree as repeating studied stimuli in a way that does not involve memory.

5.3 Experiment 9

Experiment 9 tested whether there is a relationship between recall state and identification when non-recalled items are re-encountered again after recall is attempted. After the participants finished their free recall attempt, each of the studied words they did not recall was presented to them auditorily. This auditory modality was chosen as an approximation of the experience of hearing a word internally while retrieving it from memory. This was also appropriate since participants did not see their recall responses on screen as they made them, bringing our repetition manipulation closer to the experience of recall than if we presented non-recalled words visually. We also included a subset of new words in the identification phase of this experiment. This allowed us to establish whether participants showed repetition priming for recalled words, non-recalled words, or both.

If subsequent identification for recalled words is still greater than that of non-recalled words in this design, this would suggest that the specific act of retrieval in free recall leads to identification benefits. If there is no difference in identification for recalled and non-recalled words, then the act of recalling words has no benefit on identification over merely being exposed to words again. This is an important test of the single-system idea that even deliberate, conscious expressions of memory (like recall retrieval) and indirect, implicit expressions of memory (like perceptual identification) are driven by a common source at least to some degree.

5.3.1 Method

Participants

85 undergraduate participants from the University of Plymouth School of Psychology took part in this experiment in exchange for course credit. Nine participants were excluded for low study phase performance with the same criteria as the previous experiment, leaving a final sample of 76 participants (60 female, 15 male, 1 non-binary/other). The mean age of the sample was 20.20 ($SD = 2.80$). This sample size ensures a minimum detectable effect size of $d_z = .46$, assuming 80% power.

Materials

The same 60 images of words from Experiment 3 were used in the current experiment. WAV files of each word being spoken aloud were generated using the text2speech (Muschelli, 2020) and the tuneR (Ligges et al., 2023) R packages, interfacing with the Google Cloud Text to Speech API. Each word was spoken by the "en-US-Standard-C" voice. The experiment was presented by an OpenSesame program, hosted on a JATOS server and displayed through a web browser in the same way as the previous experiments.

Procedure

There were two experimental blocks, each featuring twenty old words and ten new words that represented either animals or tools and utensils. The assignment of old and new stimuli was randomised within each block for each participant. The order of trials in each phase was also randomised on a participant level, as was the order of blocks. In each block, participants first studied twenty word images while making comparative size judgements using the same procedure as in Experiment 8. The participants then completed the same "spot-the-difference" interval task as in the previous experiments. Participants then completed the same recall phase as in Experiment 8. Immediately after the recall task, the participants were played audio clips of any studied words that they did not correctly recall through headphones. There was a 3000 ms gap between the start of each word clip. Following this, participants completed the same post-recall interval phase and identification test phase as in Experiment 8, before moving onto the next block. After completing both blocks, participants gave their age and gender before being debriefed.

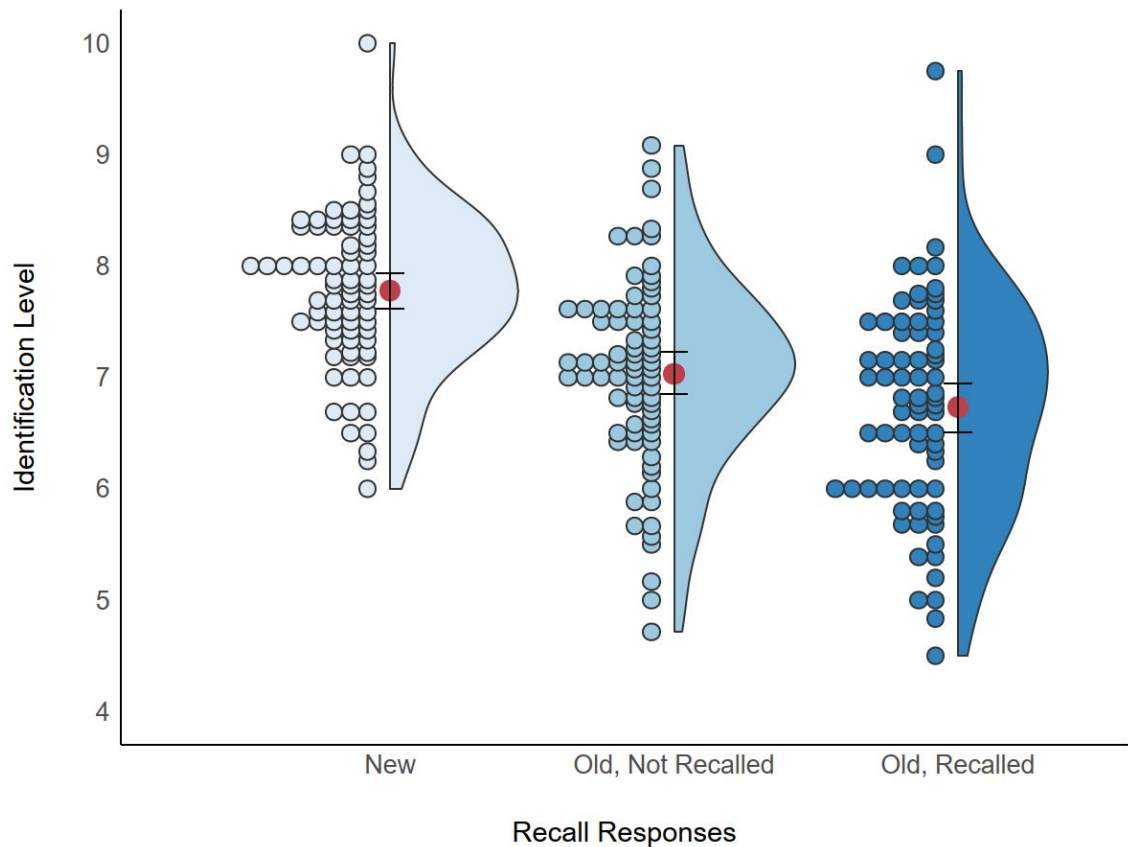


Figure 5.2: Mean fragment identification level per participant for recalled and non-recalled items from Experiment 9. The red dots denote the mean across participants, and the whiskers the 95% confidence interval.

5.3.2 Results and Discussion

Task Performance

Before other analyses, 3.82% of trials were excluded for having incorrect identification responses. Before these exclusions, the mean correct identification rate was .96 ($SE = .01$). The mean proportion of correct study phase responses was .82 ($SE = .02$). Study phase performance significantly differed between blocks, $t(75) = 2.61$, $p = .01$, 95% CI [.02, .11], $BF = 2.97$. This meant that the mean proportion of correct size classifications for animals ($M = .89$, $SE = .01$) tended to be greater than that for tools and utensils ($M = .83$, $SE = .02$) in the study phase. The mean proportion of studied items recalled across participants was .39 ($SE = .02$), leaving the mean proportion of non-recalled items as .61 ($SE = .02$). There was no reliable difference between recall performance in the animals block ($M = .40$, $SE = .03$) and the tools and utensils block ($M = .36$, $SE = .03$), $t(75) = 1.45$, $p = .14$, 95% CI [-.02, .11], $BF = 0.34$.

Recall and Identification Performance

The difference between identification levels for recalled, non-recalled, and new items was assessed using a linear mixed model with participant and item-level random intercepts. The participant-level differences in means between these item states are shown in Figure 5.2. Bonferroni-corrected pairwise comparisons from this model revealed that new items ($M = 7.80$, $SE = 0.13$) were identified at a higher level of image clarity than both non-recalled items ($M = 7.03$, $SE = 0.12$), $t(1334) = 12.01$, $p < .001$, and recalled items ($M = 6.85$, $SE = 0.13$), $t(1336) = 13.10$, $p < .001$. Most importantly, recalled items were identified at a lower level of image clarity than non-recalled items $t(1352) = 2.47$, $p = .04$. This suggests that there was a slight identification advantage for items that were recalled, compared with items that were not recalled but auditorily reinstated before identification.

These results show that recalled items are identified at a reliably lower level of image clarity than items that are not recalled, despite these non-recalled items being auditorily reinstated after recall was attempted. This indicates that the act of successful free recall retrieval confers an implicit memory advantage compared to simply being presented with the stimulus again. This result is additionally meaningful given that both recalled and non-recalled words were better identified than new words. This shows that even though stimulus repetition leads to a significant identification benefit, implicit memory for a stimulus is even better when that stimulus is recalled from memory. This result is consistent with a single-system account in which explicit and implicit memory are driven by a common memory system and casts doubt on a multiple-systems view that free recall is entirely independent of implicit memory.

5.4 General Discussion

The experiments in the present chapter investigated the relationship between free recall and perceptual identification performance. Experiment 7 found that correctly recalled items were no better identified than those that were not recalled, contrasting the results of Experiments 5 and 6 which found evidence for such an effect in cued recall. However, in a simplified design, correctly recalled items were reliably identified better than those that were not recalled in Experiment 8. This suggested that there may be an association between free recall and implicit memory performance when additional task demands present in Experiment 7 were removed. Experiment 9 reinforced this conclusion, showing that correctly recalled items are better identified than non-recalled items

even when non-recalled items were reinstated auditorily after recall was attempted. These effects were numerically small; identification was 0.40 fragmentation levels earlier for recalled items in Experiment 8, and 0.17 in Experiment 9. However, the benefit of recalling items on perceptual identification was consistent in two of three experiments (one of which was preregistered), reliable even after conservative corrections for pairwise comparisons were applied, and detectable over and above the effect of simple stimulus repetition in Experiment 9. These results therefore show associations between free recall performance and repetition priming which refute a strict multiple-systems account of explicit and implicit memory. Instead, the present results may be better explained by a single-system account.

Although the present results could be accounted for by a shared memory source, there is still the possibility that, like the results in the previous chapter, the association between free recall and implicit memory could be caused by a non-mnemonic variable. For instance, participants might pay attention to some stimuli more than others during the study phase, resulting in stronger encoding of those stimuli. Even if separate systems are assumed to store or process the memory traces that are probed in free recall and perceptual identification, this common attention variable may still give rise to an association between performance in the two tasks. By this logic, one cannot assume a causal link between a simple association in performance between two tasks, such as that found in Experiment 8.

However, the results of Experiment 9 give a stronger indication that the association between recall and identification is memory-driven. In this experiment, recalled items were better identified than non-recalled items, despite the latter being re-presented auditorily after recall was attempted. This second presentation allowed non-recalled items to be affected by further variables that might affect memory. This is important; for instance, if recalled items received greater attention or better encoding in the study phase, non-recalled items would also be subject to the same variability during their second auditory presentation. This ensures that recalled items are not differentiated from non-recalled items by the quality of their encoding during study, leaving their retrieval state in the recall phase as the only major difference between them. Therefore, if a non-mnemonic variable at encoding was solely responsible for the association between priming and free recall, no difference in identification for recalled and non-recalled groups would be expected in this design. On this basis, Experiment 9 shows that memory processes contribute to the association between free recall and perceptual identification performance, adding plausibility to a single-system account.

The present results have interesting implications for theoretical explanations of free recall. Although

free recall has often been considered a task that relies wholly upon explicit memory (Yonelinas, 2002; Quamme et al., 2004), there is also evidence to suggest that it can be subject to implicit influences (Tulving, 1985; McDermott, 2006; Mickes et al., 2013; McCabe et al., 2011). The present results expand on this work by providing the first evidence for an association between free recall and implicit memory for the same stimulus set. The consistency of this association also weighs against the idea that free recall is governed by a distinct explicit memory process, such as recollection in a dual-process account. This information can inform future efforts to model implicit and explicit memory, necessitating some degree of commonality between the source of free recall and implicit memory. Such models may also represent non-mnemonic factors that influence the link between recall and implicit memory, providing further insight into variables that may contribute to the association in performance between the two tasks.

Future empirical tests could also consider the relationship between free recall and implicit memory in studies of serial recall. Serial recall effects are well documented and are considered benchmark phenomena for computational models of free recall (Kahana, 2020). Yet, the literature on recall and implicit memory has mostly concerned item recall and has given less consideration to how single-system or dual-process accounts might explain how serial position effects in recall might relate to performance in perceptual identification tasks. There is also theoretical motivation for focusing on serial recall, as it adds an additional demand upon the participant to recall contextual information — namely, the serial position of each recalled item. Although researchers have argued that item free recall might be influenced by implicit memory (Tulving, 1985; McDermott, 2006; Mickes et al., 2013; McCabe et al., 2011), it is harder to see how an implicit memory or familiarity process could enable serial recall. By definition, such processes do not allow the participant access to concrete contextual details from which to base their memory judgement. Since an item's serial position forms part of the context in which it was studied, serial recall performance is a relatively pure reflection of explicit memory. If there is continuity between serial recall and implicit memory, this would therefore provide further evidence against a strict multiple-systems or dual-process view of the two tasks.

This research objective could also tie in with efforts to establish the boundary conditions of the association between free recall and implicit memory. Although Experiments 8 and 9 found evidence for this association, Experiment 7 did not, finding no reliable evidence for a difference in identification between recalled and non-recalled old items. The additional task demands in this experiment likely diluted the association between recall and identification performance; participants had to determine whether generated items were old or new before reporting them in the recall phase.

It is possible that adding the requirement to recall items in a serial position order would result in the same null effect on identification. One could also establish whether free recall performance relates to different expressions of implicit memory, such as conceptual priming. There is evidence that suggests free recall and conceptual priming are dissociable (Paller, 1990), which in turn places a limit upon the association between recall and implicit memory. Such a result could be verified in a method using serial recall. As discussed in Chapter 4, priming measures that rely on semantic or conceptual links risk allowing participants to guess the identity of correct recall targets, rather than engaging in a memory search process to retrieve them. However, since serial free recall requires contextual reinstatement to complete successfully, any semantic influences would be unlikely to contaminate this free recall measure. An experiment measuring the conjoint association between these tasks may therefore provide another test of the link between explicit and implicit memory, with implications for the single-system and dual-process accounts. Such experiments should be conducted to build upon the present results and inform future theory and model development in the domain of the memory systems debate.

Chapter 6

Discussion

This thesis aimed to provide new tests of single-system and dual-process accounts of explicit and implicit memory. Chapter 2 established new tests of single-system and dual-process models of recognition memory and priming, although the experimental manipulations that attempted to elicit each model's competing predictions were unsuccessful. Chapter 3 used different manipulations to test similar predictions in extended single-system and dual-process models, leveraging recent work on the encoding variability hypothesis (Spanton & Berry, 2020, 2022). However, despite yielding new results of theoretical significance for the encoding variability hypothesis, these experimental manipulations failed to elicit competing predictions in the models. As such, the single-system and dual-process models tested were not able to be evaluated on the basis of their qualitative predictions or even their quantitative fit to the data, which was roughly equivalent. The latter half of the thesis concerned the relationship between perceptual identification and forms of recall. Chapter 4 showed a relationship between cued recall and perceptual identification performance, confirming the results of previous work on the topic (Mazancieux et al., 2020). Chapter 5 then tested the relationship between free recall performance and implicit memory, finding an association in two of three experiments. I now consider the practical and theoretical implications of these results for the memory systems debate.

6.1 Single-System and Dual-Process Models of Recognition Memory and Priming

The SS model predicts a continuity between recognition and priming performance. Specifically, if memory strength is greater for an item, then that item should have greater recognition strength and

be identified at a lower level of clarity in a priming task. Predictions that arise from this continuity have been tested extensively (Berry et al., 2006a, 2008a,b, 2012). By contrast, the DPSD1 model assumes that priming is driven wholly by familiarity, an equal variance signal detection process, and that any changes in recognition memory resulting from the model's recollection process are not expected to affect priming. This distinction allows the models to make opposing predictions about measures of priming when certain features of recognition memory are manipulated. Experiments 1 and 2 investigated whether a recognition response speeding manipulation impacts the difference in RTs for items that receive miss and hit responses. Building on a method validated by Experiment 3, Experiment 4 tested the effect of adding encoding variability to recognition memory on the $M - H$ identification difference. However, none of the present methods elicited the models' contrasting predictions, with key parameter estimates remaining unchanged.

Although the SS model has been extensively tested against a variety of memory phenomena (Ward et al., 2013b,a; Berry et al., 2014; Rothen et al., 2020), Experiments 1, 2, and 4 mark its only direct comparisons with the DPSD1 model since both models were specified by Berry et al. (2012). Although the SS model has gained support from more experimental results and model fitting exercises than the DPSD1 model, both models are on a relatively even standing when compared with one another. Berry et al. (2012) compared the quantitative fit of the DPSD1 and SS models in two experiments, finding each model fitted best to data from one experiment. However, the result that identification RTs associated with "remember" responses were shorter than those for "know" responses in their Experiment 3 was not consistent with the DPSD1 model. The SS model predicted this result by assuming a common memory strength source for identification RTs and "remember" responses. Because the present experiments were unable to discriminate the models based on their qualitative predictions, Berry et al. (2012)'s Experiment 3 remains the only instance that the models were distinguishable on this basis.

However, adjustments to the DPSD1 model's specification could have accounted for this result while remaining in line with a dual-process view of recognition and priming. Berry et al. (2012) highlight that the DPSD1 model could be modified to assume recollection can affect identification in select cases. For example, if a partially-obscured stimulus showed enough visible features to allow the participant to recollect it during identification, this could act as a cue to enable faster identification. There is also evidence that short-term priming can speed the onset of recollection (Park & Donaldson, 2016) and that familiarity might enable a faster search of episodic memory that leads to a greater likelihood of recollection (Woollams et al., 2008). One may reflect this in a DPSD1-like model by introducing dependencies between the familiarity signal f and the

recollection parameter R . This would allow the model to account for certain associations between recognition and familiarity, including faster identification RTs for "remember" responses.

Some dual-process models of recognition memory assume the memory strength distribution for recollected items is continuous (Wixted & Mickes, 2010; Moran & Goshen-Gottstein, 2015). Because recollection is graded in these models, it is possible for a recollected item to have lower memory strength than an item with high familiarity. By contrast, the DPSD1 model represents recollection as a threshold process and assumes recollected items have an undetermined strength value that necessitates the highest confidence "old" judgement possible. If implemented in models of recognition and priming, each of these assumptions could lead to substantially different predictions from the current DPSD1 model. As a result, there is scope for different dual-process models to be compared against one another, and against the SS model in future research. If a particular dual-process model was found to be favourable over other specifications, this could constrain the assumptions that are possible under a dual-process view and give insight into how the hypothesised familiarity and recollection processes might interact in conjoint tests of recognition and priming.

A single-system model of recognition and priming could also be specified in many ways. While the SS model shares many assumptions with the UVSD model of recognition memory, there are other ways to model recognition memory strength along a continuum. For instance, DeCarlo (2002) specified a mixture SDT model in which old items are split across two distributions, A and A' , that correspond to different levels of encoding quality. Items that are encoded strongly are represented in the A distribution, which has a greater mean than the A' distribution of item strengths that are encoded partially or not attended to at all. Such a model could be extended to represent priming from the same memory source as recognition, forming a model in which a single memory system is assumed to drive the two memory tasks. This model could make many of the same predictions as Berry et al. (2012)'s SS model, most notably that items that have high recognition memory strength should be identified more quickly in perceptual priming tasks. Both single-system and dual-process models of recognition and priming could also be adapted to represent priming measures more accurately. Although it is well known that RTs in a perceptual identification task are positively skewed, the modelling framework specified by Berry et al. (2012) assumes a Gaussian RT distribution for the purpose of computational simplicity. Lange et al. (2019) circumvented this issue in their SS model of recognition, priming, and source memory by modelling log-transformed identification RTs which were not positively skewed. However, by specifying non-Gaussian RT distributions, future models may provide more principled accounts of priming data in CID-R tasks without relying on statistical transformations. In the case of each of these

changes, qualitative predictions and quantitative fits could be compared with existing models to assess the most principled ways of representing a single-system account.

When testing theoretical accounts of recognition and priming, it is also worth considering the commonalities between explicit and implicit memory processes at different time points in an experiment. This has been a focus of recent study. For instance Kim (2019) found in a meta-analysis that effects indicative of explicit and implicit memory share extensive neural correlates during encoding, but few at retrieval. However, while SDT provides useful metrics about memory as measured in the test phase, it is agnostic as to how encoding and retrieval processes influence later memory performance. While one can measure the results of encoding, retention, or retrieval manipulations by fitting an SDT model to experimental data, the model itself does not explicitly represent the timescale upon which these manipulations affect memory; only their outcomes. With this limitation, it can be difficult to differentiate flexible multiple-systems SDT models from single-system alternatives. For instance, the SS model is mathematically nested underneath the MS2 model, and so is hard to distinguish on the basis of unique qualitative predictions (Berry et al., 2012). Although dual-process SDT models do not share this issue, the results of Chapters 2 and 3 show that it is also difficult to test their competing qualitative predictions against the SS model using experimental manipulations. These models would have to be substantially extended to provide explicit representations of shared or separate processes at encoding or retrieval. Alternatively, models from other frameworks could be used to represent these constructs and provide further information that may validate single-system or dual-process accounts of recognition and priming.

The Retrieving Effectively from Memory: Implicit (REMI; Schooler et al., 2001) model is one such model that extends upon the popular REM model of recognition memory (Shiffrin & Steyvers, 1997). Unlike SDT, REMI includes explicit representations of encoded information, representing memory traces as vectors of feature values. It also formalises a process by which these traces decay and change as a factor of time and repeated exposure. To make a recognition judgement, the similarity between a probe vector and all remembered vectors is compared in a Bayesian inference procedure. In an identification task, the feature value vector of a partially obscured stimulus is compared in a similar way with all those in memory, with an item being named if the number of matches between a memory trace and the current probe exceeds a criterion parameter. While recognition and identification judgements share the same memory trace and have similar retrieval processes, the factors that affect these processes differ based on the requirements of the task. In explicit tasks, for instance, contextual information plays a greater role in determining the response than in implicit tasks (Schooler et al., 2001). The model also states that priming works by

changing the lexical-semantic feature representation of a stimulus in memory; a process which also affects recognition. While not specifically developed to represent a single-system account, REMI's predictions about the factors that affect explicit and implicit memory have clear implications for the memory systems debate and provide testable, mechanistic explanations of recognition and priming that SDT models cannot. Given the difficulty in discriminating between single-system and dual-process SDT models of recognition and priming, focusing on such predictions is important for the broader investigation of the relationship between explicit and implicit memory.

6.2 Continuity Between Recall and Implicit Memory

Chapters 4 and 5 investigated the association between cued and free recall and implicit memory performance in perceptual identification tasks. Experiments 5 and 6 showed that cue items were identified at a lower level of image clarity when their paired target was correctly recalled, compared with when no recall attempt was made. This finding replicated previous research by Mazancieux et al. (2020), who found the same association between recall and implicit memory performance in both episodic and semantic recall tasks. In Experiment 7, there was no reliable association between free recall performance and identification. Yet, Experiments 8 and 9 found evidence for this association, removing additional task demands that were present in Experiment 7, and showing an improvement in identification for freely recalled items over non-recalled, auditorily reinstated items. These results refute a strict dual-process account of explicit and implicit memory which states that recall is driven by recollection, a purely explicit process completely independent of familiarity. The association between recall and implicit memory may better be explained by a single-system theory, complementing existing results that support this account of explicit and implicit memory in the context of other tasks (Berry et al., 2012; Lange et al., 2019).

An association between perceptual identification and recall might be expected if explicit memory contaminated identification. Cases of participants completing implicit memory tasks by using explicit memory strategies are well documented and can lead to artefactual associations between explicit and implicit memory (MacLeod, 2008). For example, a participant could reveal part of a stimulus in a perceptual identification task and then conduct an explicit memory search to assist their identification response. However, the appearance of new items in Experiments 5, 7, and 9 would have made this strategy ineffectual, as participants would have had no explicit memory of these items in the context of the experiment. Furthermore, there is evidence that knowing whether an item is old or new does not affect priming (Ward et al., 2013b). This shows that participants

do not reliably use explicit memory cues to aid their implicit memory even when they know an item is studied. As such, it is likely that participants did not use explicit strategies in the present identification tasks, as these would have generally led to a greater expenditure of effort and lower performance than simply performing the task as instructed.

Another potential alternative interpretation of the results of Chapters 4 and 5 is that the relationship between recall and identification is stimulus-specific, rather than the result of common memory processes. That is, if an item can be recalled, it is likely salient or memorable in itself, and therefore more likely to be identified at a lower level of clarity. This could explain the association between recall and identification without the need for a common memory strength source. However, the linear mixed model analyses in Chapters 4 and 5 included item-level random effects. This means that even after variability in identification between stimuli was accounted for, identification was reliably better for recalled items. The five experiments in Chapters 4 and 5 also demonstrate this effect using both words and object images as stimuli, showing that it is not dependent upon a certain class of stimulus. Mazancieux et al. (2020) also found this effect using faces and naturalistic scenes while accounting for item-level variability in their analyses. As such, it is likely that the relationship between recall and identification performance is relatively stimulus invariant, and is, therefore, a meaningful trend of psychological interest.

It is worth acknowledging that the differences in identification performance for recalled and non-recalled items in each experiment were relatively small. In Experiments 5 and 6, the marginal mean identification level for recalled items was only 0.3 identification levels lower than the mean identification level for non-recalled items on a 15-point scale. In Experiment 8, recalled items were identified an average of 0.4 identification levels lower than non-recalled items on a 10-point scale. In Experiment 9, the difference between mean identification levels was 0.18. This was half the magnitude of that seen in the previous experiment, likely because non-recalled items were better identified due to being presented twice before identification, rather than only once. In each experiment, the identification differences likely only translated to minor differences in stimulus clarity. Regardless of size, these effects are notable because of their consistency. In four out of five experiments in Chapters 4 and 5, there were reliable differences in identification performance for recalled and non-recalled stimuli. Further, these effects were detectable even after accounting for participant and stimulus-level random effects and applying conservative statistical corrections to key pairwise comparisons. This supports the existence of a reliable association between performance in perceptual identification and recall tasks that should be accounted for by conjoint psychological theories of these tasks.

Single-system models have previously explained similar associations between performance in explicit memory tasks and repetition priming (Berry et al., 2012; Lange & Berry, 2021; Lange et al., 2019). The present results demonstrate an association between recall and priming and discredit any strict multiple-systems or dual-process view that the memory systems governing these tasks are completely independent. But, they cannot rule out a flexible multiple-systems account in which recall and implicit memory arise from separate but partially correlated memory strength sources. Despite this, there are conceptual reasons to prefer a single-system account, or at least investigate it further. Firstly, the predictions of a single-system theory are more constrained than a flexible multiple-systems or dual-process theory and are therefore easier to test and disambiguate from other models. Secondly, by positing one memory strength source rather than multiple, a single-system account would give a more parsimonious explanation of our results. These points led Berry et al. (2012) to prefer their SS model of recognition and priming over their flexible multiple-systems (MS2) model. Although these models could not be disambiguated based on their theoretical predictions, only the SS model could specify key predictions prior to data collection, because of its constraints. The SS model also fitted better than the MS2 model to experimental data according to the AIC, a goodness of fit metric that penalises models on their complexity. This helped to establish the SS model as not only a strong account of recognition and priming but one that can drive further theoretical development by making well-defined, a priori predictions. It is likely that similar characteristics would also aid a single-system model of recall and priming in comparison to a flexible multiple-systems alternative.

6.3 Considerations For Modelling Recall and Implicit Memory

Although the present results show a link between recall and implicit memory performance that ties into established psychological theory, they do not present a model-based account of this association. Formal models are invaluable tools for theory development (Guest & Martin, 2021), and have done much to inform the memory systems debate in regard to other explicit and implicit memory tasks. Now that Chapters 4 and 5 have established a relationship between recall and implicit memory performance, a logical next step would be to develop formal, conjoint models of recall and priming that represent this association. The SDT framework has been used extensively to model recognition memory (for review, see Rotello, 2017), and has been extended to form conjoint models of recognition and priming (Berry et al., 2012) and recognition, priming, and source memory (Lange et al., 2019; Lange & Berry, 2021). SDT has also been used to represent recall, both in

models of cued recall and source attribution (Marsh & Bower, 1993; Marsh & Landau, 1995), and cued recall and metacognition (Jang et al., 2012b). However, these models have limitations that decrease their validity and hinder their extension to represent recall, recognition, and priming in one framework.

As a modelling framework built to represent classification judgements, SDT does not naturally provide a clean description of cued recall, which is not a classification task. In tasks like recognition memory, the participant must distinguish between old and new items by making "old" or "new" classification judgements, resulting in a 2×2 error matrix. The response types in each cell of this matrix are countable and can be used to calculate further metrics of task performance, such as d' . As previously established, a signal detection model represents the memory strength underlying these judgements in Gaussian distributions corresponding to each item type (old, new). Each type of possible response in the error matrix is therefore represented by the model's strength distributions in relation to the criterion. Cued and free recall tasks are less amenable to this kind of representation. In recall, participants produce the identity of each stimulus from memory and are not restricted to a set of predefined responses. Participants can therefore make many types of errors. In cued recall, they can erroneously recall a target that was paired with a different cue, or mistakenly recall a cue item as a target. In any recall task, participants can recall items that were not presented in the experiment at all, or make no coherent response whatsoever. By contrast, there is only one way to make a correct cued recall response; to retrieve a target from the correct study context, be it a target paired with a cue, or a studied item in a free recall task. Because of this imbalance, response types in recall tasks do not conform to an $N \times N$ error matrix. This means that to represent cued or free recall, the traditional SDT model must be adapted.

Jang et al. (2012b) presented a multinomial SDT-like model that provided insight into judgements of learning preceding cued recall. However, the model was limited in other ways. Firstly, it cannot represent recall intrusions; items that did not appear in the experiment that were erroneously recalled. Although these could be represented with another distribution in the model, this distribution would also have to represent the entire set of items that did not appear in the experiment that were not erroneously recalled. As this set is infinite, specifying a corresponding memory strength distribution for it would present a challenge. This model, along with those presented by Marsh & Bower (1993) and Marsh & Landau (1995), also assumed that cued recall decisions are based upon memory strength for the recall target. Yet it is much more likely that cued recall strength is not a property of the target alone, but rather the cue-target match during retrieval (Hollins et al., 2016). Again, the model specification could be adjusted to account for this, with the final cued

recall strength distribution being the output of some function of the cue and target strength variables. However, specifying this in a theoretically principled way in a conjoint model would take careful consideration, and is a task beyond the scope of the present work.

Finally, SDT models assume that the latent strength variables in each test trial are independent. Although noise can be included in an SDT model to represent shifting decision criteria (Benjamin et al., 2009), the strength values in each trial cannot be influenced by previous trials, or affect those following. Although this assumption has been validated when applied to recognition memory (Kellen et al., 2021), it is likely to be problematic in regards to modelling any recall task. There are many well-known sequential recall effects in the serial recall literature, where the probability of recalling an item is influenced by factors in the preceding trials (Kahana, 2020). For example, items that are close in presentation order during study are more likely to be recalled together, as are semantically similar items. The presence of errors such as prior-list intrusions also depends upon contiguity, semantics, and recency effects. To give an accurate account of cued or free recall, a model would have to extend the typical signal detection theory framework to represent these inter-trial dependencies. Together, these considerations make developing a signal detection model of recall and implicit memory a substantial task.

With this in mind, it is worth considering whether other modelling frameworks would be better suited to representing conjoint models of recall and implicit memory. There are several popular computational models of recall, including the Search for Associative Memory (SAM; Raaijmakers & Shiffrin, 1980) model and models relating to Retrieved Context Theory (RCT; Howard & Kahana, 1999, 2002; Kahana, 2020). These can account for many established recall phenomena that involve temporal and semantic relationships between items (Kahana, 2020). However, these models do not have implicit memory components, and so would require extension to represent recall and priming. A related computational model that includes implicit memory is REMI (Schooler et al., 2001). As previously established, REMI makes explicit the process by which memory traces form and change, and identifies task-specific factors that may differentially affect explicit and implicit memory retrieval. Although REMI was developed with recognition in mind, it was inspired by models that have been used to represent recall such as SAM and REM, and so could feasibly be extended for the purpose of investigating the relationship between recall and priming. This could enable the model to make predictions about the effects of specific manipulations on recall and priming, driving further theory and model development. In all, these models give viable accounts of recall because they make explicit assumptions about the representation of items in memory from encoding to retrieval, and the relationship between these representations. Future conjoint models

should strive to do the same in order to present plausible accounts of recall and priming.

6.4 Alternative Classifications of Memory

Theorists have long subdivided human memory into explicit or implicit categories. This subdivision originated from studies of amnesic patients that found dissociations between expressions of memory that are consciously accessible, and those argued to be inaccessible (Squire & Zola-Morgan, 1991). Viewing memory according to this dichotomy has led to fruitful theoretical advances in our understanding of how various expressions of memory work, and the extent to which they require conscious awareness. However, explicit and implicit memory are theoretical constructs. Although some evidence suggests that each memory system has independent neural correlates (Woollams et al., 2008; Schott et al., 2005, 2006), there is also evidence to suggest that expressions of explicit and implicit memory share some neural activation (Kim, 2019; Addante, 2015) or are behaviourally associated (Mazancieux et al., 2020). Mathematical modelling evidence shows that single-system models of recognition and priming (Berry et al., 2012) and source memory and priming (Lange et al., 2019; Lange & Berry, 2021) can provide strong accounts of observed data, even explaining results initially thought to support multiple-systems accounts (Berry et al., 2008a). Indeed, the present results build upon this evidence, confirming results that were predicted by the SS model of recognition and priming, and finding continuities between both cued and free recall and identification performance.

There is also evidence suggesting that various memory tasks can be completed by explicit or implicit means. Recognition memory tasks are often thought to be influenced by mnemonic information that is accompanied by varying levels of conscious awareness (Yonelinas, 2002; Yonelinas et al., 2022). There is also evidence that cued and free recall measures may rely upon implicit memory, despite being traditionally considered pure measures of explicit memory (Ozubko et al., 2021; McCabe et al., 2011; Uner & Roediger, 2018; Hamilton & Rajaram, 2003). The results of Chapters 4 and 5 also show relationships between recall and implicit memory for the same items, weighing against the idea that separate recollection and familiarity processes are responsible for each task, respectively. These points raise important questions. If expressions of explicit and implicit memory overlap, and there is consistent evidence against strict multiple-systems and dual-process accounts, is it still worthwhile to conceptualise memory according to this divide? Are there different ways to categorise expressions of human memory that could lead to further theoretical advances?

Some studies have provided information about the structure of memory by investigating the effect of endogenous and exogenous variables. Exogenous variables are those which originate outside of

the participant, such as the time of day or the characteristics of stimuli being remembered, whereas endogenous variables are the result of the cognitive processes or attributes of the participant. Kahana et al. (2018) investigated the effect of these variables on variability in recall using linear mixed modelling in a multi-session experiment. The results showed that prior list recall predicted performance in successive lists to a greater extent than exogenous variables such as word memorability. Kahana et al. (2018) concluded on this basis that endogenous factors that affect encoding and retrieval are the primary sources of variability in recall performance. This conclusion was reinforced by findings showing EEG signals associated with endogenous neural activity predicted successful encoding in recall better than external variables (Weidemann & Kahana, 2021). These results have clear implications for formal models of recall, despite not making assumptions about the conscious accessibility of the variables involved in the task. They also open possibilities for the study of implicit memory. It is understood that implicit memory tasks are more susceptible to the influence of non-mnemonic factors (Buchner & Wippich, 2000), yet it is unclear whether this means that exogenous variables have a greater influence on variability in priming than recall. Future methods could establish the extent to which exogenous and endogenous variables influence priming, and compare these results with those from explicit memory tasks such as recall. This may provide further insight into the mechanisms underpinning both tasks, with the potential to add to the conclusions of Chapters 4 and 5.

This is only one other possible subdivision of human memory. However, it is unlikely that any current dichotomous theoretical account can explain the full complexity of human memory and its interrelation with other cognitive functions. Indeed, when considering the innumerable factors that can influence memory, reducing its study to simple dichotomies such as explicit and implicit systems and endogenous and exogenous variables may seem reductionist. Yet, studying theoretical accounts linked to these dichotomies has led to substantial advances in understanding memory and the cognitive processes underlying it. The present thesis contributes to this knowledge and identifies many possible tests and developments of single-system and dual-process theories for future research. Such research should pursue these avenues in order to constrain theory in the domain of the memory systems debate. These results may then be integrated with those from other theoretical viewpoints to inform the development of future models that represent the conscious accessibility of human memory.

6.5 Conclusion

To conclude, this thesis investigated the relationship between explicit and implicit memory using recognition, cued recall, free recall, and perceptual identification tasks. Chapters 2 and 3 presented novel qualitative predictions made by the SS and DPSD1 models of recognition memory and priming in response to experimental manipulations. However, these manipulations were ultimately unable to elicit these predictions, and the models could not be discriminated on this basis, nor by comparisons of quantitative fit. Chapter 4 focused on cued recall, finding correctly recalled items were identified better than those that were not recalled. Chapter 5 found the same pattern of results in two out of three experiments using free recall. Taken together, both sets of recall results provided evidence against strict multiple-systems and dual-process accounts of repetition priming and recall, and are better explained by a single-system account. Going forward, these results can constrain theory in the domain of the memory systems debate, and inform the development of new conjoint models of explicit and implicit memory. Further research may also investigate the relationship between recall and priming using other types of implicit memory task to gauge the generalisability of this association.

References

- Addante, R. J. (2015). A critical role of the human hippocampus in an electrophysiological measure of implicit memory. *NeuroImage, 109*, 515–528.
<https://doi.org/10.1016/j.neuroimage.2014.12.069>
- Banks, B., & Connell, L. (2022). Category production norms for 117 concrete and abstract categories. *Behavior Research Methods*, (pp. 1–22).
<https://doi.org/10.31234/osf.io/7peqx>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). Parsimonious mixed models. *arXiv preprint*.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.
<https://doi.org/10.18637/jss.v067.i01>
- Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: applications to recognition memory. *Psychological review, 116*(1), 84.
<https://doi.org/10.1037/a0014351>
- Berry, C. J., Henson, R. N., & Shanks, D. R. (2006a). On the relationship between repetition priming and recognition memory: Insights from a computational model. *Journal of Memory and Language, 55*(4), 515–533.
<https://doi.org/10.1016/j.jml.2006.08.008>
- Berry, C. J., Kessels, R. P., Wester, A. J., & Shanks, D. R. (2014). A single-system model predicts recognition memory and repetition priming in amnesia. *Journal of Neuroscience, 34*(33), 10963–10974.
<https://doi.org/10.1523/JNEUROSCI.0764-14.2014>
- Berry, C. J., Shanks, D. R., & Henson, R. N. (2006b). On the status of unconscious memory: Merikle and reingold (1991) revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 925–934.
<https://doi.org/10.1037/0278-7393.32.4.925>
- Berry, C. J., Shanks, D. R., & Henson, R. N. (2008a). A single-system account of the relationship between priming, recognition, and fluency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 97–111.
<https://doi.org/10.1037/0278-7393.34.1.97>
- Berry, C. J., Shanks, D. R., & Henson, R. N. (2008b). A unitary signal-detection model of implicit and explicit memory. *Trends in cognitive sciences, 12*(10), 367–373.
<https://doi.org/10.1016/j.tics.2008.06.005>
- Berry, C. J., Shanks, D. R., Li, S., Rains, L. S., & Henson, R. N. (2010). Can “pure” implicit memory be isolated? a test of a single-system model of recognition and repetition priming. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 64*(4), 241–255.
<https://doi.org/10.1037/a0021525>

- Berry, C. J., Shanks, D. R., Speekenbrink, M., & Henson, R. N. (2012). Models of recognition, repetition priming, and fluency: exploring a new framework. *Psychological review*, *119*(1), 40–79.
<https://doi.org/10.1037/a0025464>
- Berry, C. J., Ward, E. V., & Shanks, D. R. (2017). Does study duration have opposite effects on recognition and repetition priming? *Journal of Memory and Language*, *97*, 154–174.
<https://doi.org/10.1016/j.jml.2017.07.004>
- Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of standardized stimuli (boss) phase ii: 930 new normative photos. *PloS one*, *9*(9), e106953.
<https://doi.org/10.1371/journal.pone.0106953>
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, *46*(3), 904–911.
<https://doi.org/10.3758/s13428-013-0403-5>
- Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive psychology*, *40*(3), 227–259.
<https://doi.org/10.1006/cogp.1999.0731>
- Burnham, K. P., & Anderson, D. R. (1998). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York, NY: Springer.
<https://doi.org/10.1007/978-1-4757-2917-7>
- Cermak, L. S., Talbot, N., Chandler, K., & Wolbarst, L. R. (1985). The perceptual priming phenomenon in amnesia. *Neuropsychologia*, *23*(5), 615–622.
[https://doi.org/10.1016/0028-3932\(85\)90063-6](https://doi.org/10.1016/0028-3932(85)90063-6)
- Cha, J., & Dobbins, I. G. (2021). Critical tests of the continuous dual-process model of recognition. *Cognition*, *215*, 104827.
<https://doi.org/10.1016/j.cognition.2021.104827>
- Challis, B. H., Velichkovsky, B. M., & Craik, F. I. (1996). Levels-of-processing effects on a variety of memory tasks: New findings and theoretical implications. *Consciousness and cognition*, *5*(1-2), 142–164.
<https://doi.org/10.1006/ccog.1996.0009>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.
- Cohen, J. (1994). The earth is round ($p < .05$). *American psychologist*, *49*(12), 997–1003.
<https://doi.org/10.1037/0003-066X.49.12.997>
- Conroy, M., Hopkins, R., & Squire, L. (2005). On the contribution of perceptual fluency and priming to recognition memory. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 14–20.
<https://doi.org/10.3758/CABN.5.1.14>
- Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*(6), 595–609.
<https://doi.org/10.1080/09658211.2010.493892>
- Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology*, *68*(8), 1489–1501.
<https://doi.org/10.1080/17470218.2014.945096>
- Craik, F. I., Moscovitch, M., & McDowd, J. M. (1994). Contributions of surface and conceptual information to performance on implicit and explicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 864–875.
<https://doi.org/10.1037/0278-7393.20.4.864>

- DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: theoretical developments with applications to recognition memory. *Psychological review*, *109*(4), 710–721.
- Denny, E. B., & Hunt, R. R. (1992). Affective valence and memory in depression: dissociation of recall and fragment completion. *Journal of Abnormal Psychology*, *101*(3), 575–580.
<https://doi.org/10.1037/0021-843X.101.3.575>
- Dew, I. T., & Mulligan, N. W. (2008). The effects of generation on auditory implicit memory. *Memory & cognition*, *36*, 1157–1167.
<https://doi.org/10.3758/MC.36.6.1157>
- Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2008). The effects of unitization on familiarity-based source memory: testing a behavioral prediction derived from neuroimaging data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 730–740.
<https://doi.org/10.1037/0278-7393.34.4.730>
- Dopkins, S., Varner, K., & Hoyer, D. (2017). Variation in the standard deviation of the lure rating distribution: Implications for estimates of recollection probability. *Psychonomic bulletin & review*, *24*, 1658–1664.
<https://doi.org/10.3758/s13423-017-1232-9>
- Dunn, J. C. (2010). How to fit models of recognition memory data using maximum likelihood. *International Journal of Psychological Research*, *3*(1), 140–149.
<https://doi.org/10.21500/20112084.859>
- Egan, J. P. (1958). Recognition memory and the operating characteristic. *USAF Operational Applications Laboratory Technical Note*.
- Ekstrom, A. D., & Bookheimer, S. Y. (2007). Spatial and temporal episodic memory retrieval recruit dissociable functional networks in the human brain. *Learning & memory*, *14*(10), 645–654.
<https://doi.org/10.1101/lm.575107>
- Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge University Press.
<https://doi.org/10.1017/CB09781316272503>
- Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General*, *112*(3), 309–346.
<https://doi.org/10.1037/0096-3445.112.3.309>
- Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage*, *32*(3), 1413–1421.
<https://doi.org/10.1016/j.neuroimage.2006.06.007>
- Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annual review of psychology*, *49*(1), 87–115.
<https://doi.org/10.1146/annurev.psych.49.1.87>
- Gabrieli, J. D., Fleischman, D. A., Keane, M. M., Reminger, S. L., & Morrell, F. (1995). Double dissociation between memory systems underlying explicit and implicit memory in the human brain. *Psychological Science*, *6*(2), 76–82.
<https://doi.org/10.1111/j.1467-9280.1995.tb00310.x>
- Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 5–16.
<https://doi.org/10.1037/0278-7393.16.1.5>

- Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 2(1), 21–31.
<https://doi.org/10.1037/0278-7393.2.1.21>
- Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 500–513.
<https://doi.org/10.1037/0278-7393.25.2.500>
- Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior*, 23(5), 553–568.
[https://doi.org/10.1016/S0022-5371\(84\)90346-3](https://doi.org/10.1016/S0022-5371(84)90346-3)
- Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 164–178.
<https://doi.org/10.1037/0278-7393.10.1.164>
- Green, D. M., Swets, J. A., et al. (1966). *Signal detection theory and psychophysics*, vol. 1. New York, NY: Wiley.
- Grider, R. C., & Malmberg, K. J. (2008). Discriminating between changes in bias and changes in accuracy for recognition memory of emotional stimuli. *Memory & cognition*, 36, 933–946.
<https://doi.org/10.3758/MC.36.5.933>
- Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, 16(4), 789–802.
<https://doi.org/10.1177/1745691620970585>
- Hamilton, M., & Rajaram, S. (2003). States of awareness across multiple memory tasks: Obtaining a “pure” measure of conscious recollection. *Acta Psychologica*, 112(1), 43–69.
[https://doi.org/10.1016/S0001-6918\(02\)00100-2](https://doi.org/10.1016/S0001-6918(02)00100-2)
- Hannula, D. E., & Greene, A. J. (2012). The hippocampus reevaluated in unconscious learning and memory: at a tipping point? *Frontiers in human neuroscience*, 6, 80.
<https://doi.org/10.3389/fnhum.2012.00080>
- Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1210–1230.
<https://doi.org/10.1037/0278-7393.29.6.1210>
- Hirshman, E., & Hostetter, M. (2000). Using roc curves to test models of recognition memory: The relationship between presentation duration and slope. *Memory & Cognition*, 28(2), 161–166.
<https://doi.org/10.3758/BF03213795>
- Hollins, T. J., Lange, N., Berry, C. J., & Dennis, I. (2016). Giving and stealing ideas in memory: Source errors in recall are influenced by both early-selection and late-correction retrieval processes. *Journal of Memory and Language*, 88, 87–103.
<https://doi.org/10.1016/j.jml.2016.01.004>
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 923–941.
<https://doi.org/10.1037/0278-7393.25.4.923>
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of mathematical psychology*, 46(3), 269–299.
<https://doi.org/10.1006/jmps.2001.1388>

- Huang, T. S.-T., & Shanks, D. R. (2021). Examining the relationship between processing fluency and memory for source information. *Royal Society Open Science*, 8(4), 190430.
<https://doi.org/10.1098/rsos.190430>
- Hunt, R. R., & Toth, J. P. (1990). Perceptual identification, fragment completion, and free recall: concepts and data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(2), 282–290.
<https://doi.org/10.1037/0278-7393.16.2.282>
- Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of verbal learning and verbal behavior*, 22(5), 485–508.
[https://doi.org/10.1016/S0022-5371\(83\)90301-8](https://doi.org/10.1016/S0022-5371(83)90301-8)
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110(3), 306–340.
<https://doi.org/10.1037/0096-3445.110.3.306>
- Jang, Y., Mickes, L., & Wixted, J. T. (2012a). Three tests and three corrections: comment on koen and yonelinas (2010). *Journal of Experimental Psychology; Learning, Memory and Cognition*, 32(8), 513–523.
<https://doi.org/10.1037/a0025880>
- Jang, Y., Wallsten, T. S., & Huber, D. E. (2012b). A stochastic detection and retrieval model for the study of metacognition. *Psychological Review*, 119(1), 186–200.
<https://doi.org/10.1037/a0025960>
- Jernigan, T. L., & Ostergaard, A. L. (1993). Word priming and recognition memory are both affected by mesial temporal lobe damage. *Neuropsychology*, 7(1), 14–26.
<https://doi.org/10.1037/0894-4105.7.1.14>
- Johnston, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(1), 3–11.
<https://doi.org/10.1037/0278-7393.11.1.3>
- Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology*, 71, 107–138.
<https://doi.org/10.1146/annurev-psych-010418-103358>
- Kahana, M. J., Aggarwal, E. V., & Phan, T. D. (2018). The variability puzzle in human memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(12), 1857–1863.
- Kellen, D., Winiger, S., Dunn, J. C., & Singmann, H. (2021). Testing the foundations of signal detection theory in recognition memory. *Psychological Review*, 128(6), 1022–1050.
<https://doi.org/10.1037/rev0000288>
- Kim, H. (2019). Neural correlates of explicit and implicit memory at encoding and retrieval: A unified framework and meta-analysis of functional neuroimaging studies. *Biological Psychology*, 145, 96–111.
<https://doi.org/10.1016/j.biopsycho.2019.04.006>
- Kinder, A., & Shanks, D. R. (2003). Neuropsychological dissociations between priming and recognition: a single-system connectionist account. *Psychological Review*, 110(4), 728–744.
<https://doi.org/10.1037/0033-295X.110.4.728>
- Kirwan, C. B., Wixted, J. T., & Squire, L. R. (2008). Activity in the medial temporal lobe predicts memory strength, whereas activity in the prefrontal cortex predicts recollection. *Journal of Neuroscience*, 28(42), 10541–10548.
<https://doi.org/10.1523/JNEUROSCI.3456-08.2008>

- Koen, J. D., Aly, M., Wang, W.-C., & Yonelinas, A. P. (2013). Examining the causes of memory strength variability: Recollection, attention failure, or encoding variability? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1726–1741.
<https://doi.org/10.1037/a0033671>
- Koen, J. D., & Yonelinas, A. P. (2010). Memory variability is due to the contribution of recollection and familiarity, not to encoding variability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1536–1542.
<https://doi.org/10.1037/a0020448>
- Koen, J. D., & Yonelinas, A. P. (2011). From humans to rats and back again: Bridging the divide between human and animal studies of recognition memory with receiver operating characteristics. *Learning & Memory*, *18*(8), 519–522.
<https://doi.org/10.1101/lm.2214511>
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 english words. *Behavior research methods*, *44*(4), 978–990.
<https://doi.org/10.3758/s13428-012-0210-4>
- Lange, K., Kühn, S., & Filevich, E. (2015). "just another tool for online studies"(jatos): An easy solution for setup and management of web servers supporting online studies. *PloS one*, *10*(6), e0130834.
<https://doi.org/10.1371/journal.pone.0130834>
- Lange, N., & Berry, C. J. (2021). Explaining the association between repetition priming and source memory: No evidence for a contribution of recognition or fluency. *Quarterly Journal of Experimental Psychology*, *74*(10), 1806–1817.
<https://doi.org/10.1177/17470218211008406>
- Lange, N., Berry, C. J., & Hollins, T. J. (2019). Linking repetition priming, recognition, and source memory: A single-system signal-detection account. *Journal of Memory and Language*, *109*, 104039.
<https://doi.org/10.1016/j.jml.2019.104039>
- Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.6.2-1.
<https://CRAN.R-project.org/package=emmeans>
- Ligges, U., Krey, S., Mersmann, O., & Schnackenberg, S. (2023). *tuneR: Analysis of Music and Speech*.
<https://CRAN.R-project.org/package=tuneR>
- MacLeod, C. M. (2008). Implicit memory tests: Techniques for reducing conscious intrusion. In J. Dunlosky, & R. A. Bjork (Eds.) *Handbook of Metamemory and Memory*, (pp. 245–263). Psychology Press.
<https://doi.org/10.4324/9780203805503>
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.
<https://doi.org/10.4324/9781410611147>
- Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition*, *30*, 607–613.
<https://doi.org/10.3758/BF03194962>
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review*, *87*(3), 252–271.
<https://doi.org/10.1037/0033-295X.87.3.252>

- Marsh, R. L., & Bower, G. H. (1993). Eliciting cryptomnesia: unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 673–688. <https://doi.org/10.1037/0278-7393.19.3.673>
- Marsh, R. L., & Landau, J. D. (1995). Item availability in cryptomnesia: Assessing its role in two paradigms of unconscious plagiarism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1568–1582. <https://doi.org/10.1037/0278-7393.21.6.1568>
- Masson, M. E., & MacLeod, C. M. (1992). Reenacting the route to interpretation: Enhanced perceptual identification without prior perception. *Journal of Experimental Psychology: General*, *121*(2), 145–176. <https://doi.org/10.1037/0096-3445.121.2.145>
- Masson, M. E., & MacLeod, C. M. (2002). Covert operations: Orthographic recoding as a basis for repetition priming in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 858–871. <https://doi.org/10.1037/0278-7393.28.5.858>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, *44*(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- Mazancieux, A., Pandiani, T., & Moulin, C. J. (2020). Perceptual identification task points to continuity between implicit memory and recall. *Cognition*, *197*, 104168. <https://doi.org/10.1016/j.cognition.2019.104168>
- McCabe, D. P., Roediger, H. L., & Karpicke, J. D. (2011). Automatic processing influences free recall: Converging evidence from the process dissociation procedure and remember-know judgments. *Memory & Cognition*, *39*(3), 389–402. <https://doi.org/10.3758/s13421-010-0040-5>
- McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition*, *34*(2), 261–267. <https://doi.org/10.3758/BF03193404>
- Mickes, L., Seale-Carlisle, T. M., & Wixted, J. T. (2013). Rethinking familiarity: Remember/know judgments in free recall. *Journal of memory and language*, *68*(4), 333–349. <https://doi.org/10.1016/j.jml.2013.01.001>
- Moran, R., & Goshen-Gottstein, Y. (2015). Old processes, new perspectives: Familiarity is correlated with (not independent of) recollection and is more (not equally) variable for targets than for lures. *Cognitive Psychology*, *79*, 40–67. <https://doi.org/10.1016/j.cogpsych.2015.01.005>
- Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological methods*, *16*(4), 406–419. <https://doi.org/10.1037/a0024377>
- Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs*. R package version 0.9.12-4.2. <https://CRAN.R-project.org/package=BayesFactor>
- Moscovitch, M., & Bentin, S. (1993). The fate of repetition effects when recognition approaches chance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 148–158. <https://doi.org/10.1037/0278-7393.19.1.148>

- Mulligan, N. W., & Dew, I. T. (2009). Generation and perceptual implicit memory: Different generation tasks produce different effects on perceptual priming. *Journal of experimental psychology: Learning, memory, and cognition*, 35(6), 1522–1538.
<https://doi.org/10.1037/a0017398>
- Mulligan, N. W., & Osborn, K. (2009). The modality-match effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 564–571.
<https://doi.org/10.1037/a0014524>
- Muschelli, J. (2020). *text2speech: Text to Speech*. R package version 0.2.13.
<https://CRAN.R-project.org/package=text2speech>
- Ooms, J. (2021). *magick: Advanced Graphics and Image-Processing in R*. R package version 2.7.3.
<https://CRAN.R-project.org/package=magick>
- Ostergaard, A. L. (1999). Priming deficits in amnesia: now you see them, now you don't. *Journal of the International Neuropsychological Society*, 5(3), 175–190.
<https://doi.org/10.1017/S1355617799533018>
- Ostergaard, A. L., & Jernigan, T. L. (1993). Are word priming and explicit memory mediated by different brain structures. In P. Graf, & M. E. J. Masson (Eds.) *Implicit memory: New directions in cognitive, development, and neuropsychology*. Hillsdale, NJ: Erlbaum.
- Ozubko, J. D., Sirianni, L. A., Ahmad, F. N., MacLeod, C. M., & Addante, R. J. (2021). Recallable but not recognizable: The influence of semantic priming in recall paradigms. *Cognitive, Affective, & Behavioral Neuroscience*, 21(1), 119–143.
<https://doi.org/10.3758/s13415-020-00854-w>
- Paller, K. A. (1990). Recall and stem-completion priming have different electrophysiological correlates and are modified differentially by directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6), 1021–1032.
<https://doi.org/10.1037/0278-7393.16.6.1021>
- Park, J. L., & Donaldson, D. I. (2016). Investigating the relationship between implicit and explicit memory: Evidence that masked repetition priming speeds the onset of recollection. *NeuroImage*, 139, 8–16.
<https://doi.org/10.1016/j.neuroimage.2016.06.013>
- Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review*, 114(1), 118–201.
<https://doi.org/10.1037/0033-295X.114.1.188>
- Platt, J. R. (1964). Strong inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*, 146(3642), 347–353.
<https://doi.org/10.1126/science.146.3642.347>
- Quamme, J. R., Yonelinas, A. P., Widaman, K. F., Kroll, N. E., & Sauvé, M. J. (2004). Recall and recognition in mild hypoxia: using covariance structural modeling to test competing theories of explicit memory. *Neuropsychologia*, 42(5), 672–691.
<https://doi.org/10.1016/j.neuropsychologia.2003.09.008>
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
<https://www.R-project.org/>
- Raaijmakers, J. G., & Shiffrin, R. M. (1980). Sam: A theory of probabilistic search of associative memory. In *Psychology of learning and motivation*, vol. 14, (pp. 207–262). Elsevier.
[https://doi.org/10.1016/S0079-7421\(08\)60162-0](https://doi.org/10.1016/S0079-7421(08)60162-0)

- Rabe, M. M., Stephen Lindsay, D., & Kliegl, R. (2021). Roc asymmetry is not diagnostic of unequal residual variance in gaussian signal detection theory. *PsyArXiv*.
<https://doi.org/10.31234/osf.io/erzvp>
- Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 763–785.
<https://doi.org/10.1037/0278-7393.20.4.763>
- Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using roc curves. *Psychological review*, 99(3), 518–535.
<https://doi.org/10.1037/0033-295X.99.3.518>
- Richardson-Klavehn, A., Clarke, A. B., & Gardiner, J. M. (1999). Conjoint dissociations reveal involuntary “perceptual” priming from generating at study. *Consciousness and Cognition*, 8(3), 271–284.
<https://doi.org/10.1006/ccog.1999.0382>
- Roediger, H. L., & Challis, B. H. (1992). Effects of exact repetition and conceptual repetition on free recall and primed word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(1), 3–14.
<https://doi.org/10.1037/0278-7393.18.1.3>
- Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal subjects. *Handbook of Neuropsychology*, 8, 63–131.
- Rotello, C. M. (2017). Signal detection theories of recognition memory. In J. H. Byrne (Ed.) *Learning and Memory: A Comprehensive Reference*, (pp. 201–225). Oxford: Academic Press, second edition ed.
<https://doi.org/10.1016/B978-0-12-809324-5.21044-4>
- Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic bulletin & review*, 12(5), 865–873.
<https://doi.org/10.3758/BF03196778>
- Rothen, N., Berry, C. J., Seth, A. K., Oligschläger, S., & Ward, J. (2020). A single system account of enhanced recognition memory in synaesthesia. *Memory & cognition*, 48(2), 188–199.
<https://doi.org/10.3758/s13421-019-01001-8>
- Schooler, L. J., Shiffrin, R. M., & Raaijmakers, J. G. (2001). A bayesian model for implicit effects in perceptual identification. *Psychological review*, 108(1), 257–272.
<https://doi.org/10.1037/0033-295X.108.1.257>
- Schott, B. H., Henson, R. N., Richardson-Klavehn, A., Becker, C., Thoma, V., Heinze, H.-J., & Düzel, E. (2005). Redefining implicit and explicit memory: The functional neuroanatomy of priming, remembering, and control of retrieval. *Proceedings of the National Academy of Sciences*, 102(4), 1257–1262.
<https://doi.org/10.1073/pnas.0409070102>
- Schott, B. H., Richardson-Klavehn, A., Henson, R. N., Becker, C., Heinze, H.-J., & Düzel, E. (2006). Neuroanatomical dissociation of encoding processes related to priming and explicit memory. *Journal of Neuroscience*, 26(3), 792–800.
<https://doi.org/10.1523/JNEUROSCI.2402-05.2006>
- Shanks, D. R., & Perruchet, P. (2002). Dissociation between priming and recognition in the expression of sequential knowledge. *Psychonomic Bulletin & Review*, 9(2), 362–367.
<https://doi.org/10.3758/BF03196294>

- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic bulletin & review*, *4*, 145–166.
<https://doi.org/10.3758/BF03209391>
- Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In *New methods in cognitive psychology*, (pp. 4–31). Routledge.
<https://doi.org/10.4324/9780429318405-2>
- Slotnick, S. D., & Thakral, P. P. (2013). The hippocampus operates in a threshold manner during spatial source memory. *NeuroReport*, *24*(5), 265–269.
<https://doi.org/10.1097/WNR.0b013e32835f282d>
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of experimental psychology: General*, *117*(1), 34–50.
<https://doi.org/10.1037/0096-3445.117.1.34>
- Spanton, R. W., & Berry, C. J. (2020). The unequal variance signal-detection model of recognition memory: Investigating the encoding variability hypothesis. *Quarterly Journal of Experimental Psychology*, *73*(8), 1242–1260.
<https://doi.org/10.1177/1747021820906117>
- Spanton, R. W., & Berry, C. J. (2022). Does variability in recognition memory scale with mean memory strength or encoding variability in the UVSD model? *Quarterly Journal of Experimental Psychology*, *76*(8), 2037–2052.
<https://doi.org/10.1177/17470218221136498>
- Squire, L. R. (2009). The legacy of patient HM for neuroscience. *Neuron*, *61*(1), 6–9.
<https://doi.org/10.1016/j.neuron.2008.12.023>
- Squire, L. R., & Dede, A. J. (2015). Conscious and unconscious memory systems. *Cold Spring Harbor perspectives in Biology*, *7*(3), a021667.
<https://doi.org/10.1101/cshperspect.a021667>
- Stark, C. E., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 945–972.
<https://doi.org/10.1037/0278-7393.26.4.945>
- Starns, J. J., Ratcliff, R., & McKoon, G. (2012a). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, *64*(1-2), 1–34.
<https://doi.org/10.1016/j.cogpsych.2011.10.002>
- Starns, J. J., Rotello, C. M., & Ratcliff, R. (2012b). Mixing strong and weak targets provides no evidence against the unequal-variance explanation of zROC slope: a comment on Koen and Yonelinas (2010). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(8), 793–801.
<https://doi.org/10.1037/a0027040>
- Stefanacci, L., Buffalo, E. A., Schmolck, H., & Squire, L. R. (2000). Profound amnesia after damage to the medial temporal lobe: A neuroanatomical and neuropsychological profile of patient ep. *Journal of Neuroscience*, *20*(18), 7024–7036.
<https://doi.org/10.1523/JNEUROSCI.20-18-07024.2000>
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, *26*(1), 1–12.
<https://doi.org/10.1037/h0080017>

- Uner, O., & Roediger, H. L. (2018). Are encoding/retrieval interactions in recall driven by remembering, knowing, or both? *Journal of Memory and Language*, *103*, 44–57.
<https://doi.org/10.1016/j.jml.2018.07.002>
- Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly journal of experimental psychology*, *67*(6), 1176–1190.
<https://doi.org/10.1080/17470218.2013.850521>
- Voss, J. L., & Gonsalves, B. D. (2010). Time to go our separate ways: opposite effects of study duration on priming and recognition reveal distinct neural substrates. *Frontiers in Human Neuroscience*, *4*, 227.
<https://doi.org/10.3389/fnhum.2010.00227>
- Voss, J. L., Reber, P. J., Mesulam, M.-M., Parrish, T. B., & Paller, K. A. (2007). Familiarity and Conceptual Priming Engage Distinct Cortical Networks. *Cerebral Cortex*, *18*(7), 1712–1719.
<https://doi.org/10.1093/cercor/bhm200>
- Ward, E., Berry, C., & Shanks, D. (2013a). Age effects on explicit and implicit memory. *Frontiers in psychology*, *4*, 639.
<https://doi.org/10.3389/fpsyg.2013.00639>
- Ward, E. V., Berry, C. J., & Shanks, D. R. (2013b). An effect of age on implicit memory that is not due to explicit contamination: implications for single and multiple-systems theories. *Psychology and Aging*, *28*(2), 429–442.
<https://doi.org/10.1037/a0031888>
- Ward, E. V., Berry, C. J., Shanks, D. R., Moller, P. L., & Czsiser, E. (2020). Aging predicts decline in explicit and implicit memory: A life-span study. *Psychological Science*, *31*(9), 1071–1083.
<https://doi.org/10.1177/0956797620927648>
- Watkins, P. C. (2002). Implicit memory bias in depression. *Cognition & Emotion*, *16*(3), 381–402.
<https://doi.org/10.1080/02699930143000536>
- Weidemann, C. T., & Kahana, M. J. (2021). Neural measures of subsequent memory reflect endogenous variability in cognitive function. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(4), 641–651.
<https://doi.org/10.1037/xlm0000966>
- Wilson, B. A., & Wearing, D. (1995). Prisoner of consciousness: A state of just awakening following herpes simplex encephalitis. In R. Campbell, & M. A. Conway (Eds.) *Broken Memories: Case Studies in Memory Impairment*, (pp. 14–30). Blackwell Publishing.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological review*, *114*(1), 152–176.
<https://doi.org/10.1037/0033-295X.114.1.152>
- Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological review*, *117*(4), 1025–1054.
<https://doi.org/10.1037/a0020874>
- Woollams, A. M., Taylor, J. R., Karayanidis, F., & Henson, R. N. (2008). Event-related potentials associated with masked priming of test cues reveal multiple potential contributions to recognition memory. *Journal of Cognitive Neuroscience*, *20*(6), 1114–1129.
<https://doi.org/10.1162/jocn.2008.20076>
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of experimental psychology: Learning, memory, and cognition*,

- 20(6), 1341–1354.
<https://doi.org/10.1037/0278-7393.20.6.1341>
- Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6), 1415–1434.
<https://doi.org/10.1037/0278-7393.25.6.1415>
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, 46(3), 441–517.
<https://doi.org/10.1006/jmla.2002.2864>
- Yonelinas, A. P., Kroll, N., Dobbins, I., Frederick, C., Knight, R., & Gazzaniga, M. (2001). Visual implicit memory in the left hemisphere: evidence from patients with callosotomies and right occipital lobe lesions. *Psychological Science*, 12(4), 293–298.
<https://doi.org/10.1111/1467-9280.00354>
- Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: a review. *Psychological Bulletin*, 133(5), 800–832.
<https://doi.org/10.1037/0033-2909.133.5.800>
- Yonelinas, A. P., Ramey, M. M., & Riddell, C. (2022). Recognition memory: The role of recollection and familiarity. In M. Kahana, & A. Wagner (Eds.) *Oxford Handbook of Human Memory: Foundations and Applications*. Oxford University Press.
- Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., & Kahana, M. J. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 792–804.
<https://doi.org/10.1037/0278-7393.32.4.792>

Appendices

Appendix A

Parameter Estimation

All signal detection models were fitted to experimental data using maximum likelihood estimation (Dunn 2010). There were twenty repetitions of the fitting procedure. In the first ten, randomly sampled starting parameters were generated from normal distributions, with means estimated from the simulated data. In the latter ten repetitions, starting parameters were randomly sampled from distributions with means fixed to plausible a priori values. These sampling techniques resulted in a range of different parameter values being optimised in each repetition of the fitting procedure. This increased the chance of the fitted parameters being reflective of the global minimum log-likelihood. The starting parameters in each repetition of the fitting procedure were used to optimise the negative log-likelihood of the model using the Nelder-Mead algorithm, implemented in the `optim` function in R. The negative log-likelihood for each model was defined as the negative sum of the logarithms of the likelihood in each trial. Of the twenty model fits, that with the minimum negative log-likelihood was chosen as having the best fit to the data. The parameter values from the optimised log-likelihood function were then recovered as the best-fitting parameters for that dataset. The likelihood equations for each model are described below.

A.1 Experiments 1 and 2

In Experiment 1, the SS model likelihood equation for a given recognition confidence response (Z) and identification response time (ID) is almost identical to Equation A2 of Berry et al. (2012). The only difference is that in the present experiments, σ_0 was free to vary, rather than being fixed to

equal 1. With this modification, the likelihood function is

$$L(Z, ID|I) = [\Phi(\lambda_{j+1}|\mu_{J_r|ID,I}, \sigma_{J_r|ID}^2) - \Phi(\lambda_j|\mu_{J_r|ID,I}, \sigma_{J_r|ID}^2)] \times \phi(ID|b - s\mu_{p|I}, \sigma_{ID}^2) \quad (\text{A.1})$$

where $I = \text{old, new}$; Φ is the cumulative normal distribution function; ϕ is the normal density function; $\mu_{J_r|RT,I}$ and $\sigma_{J_r|RT}^2$ are the mean and variance of the conditional distribution of J_r given RT, $j = Z = 1 \dots 6$, and $\lambda = \{-\infty, C_1, C_2, \dots, C_5, \infty\}$, a vector of criteria with upper and lower bounds of positive and negative infinity. Also reflecting the newly added unequal variance assumption, the mean of the condition distribution, $J_r|ID$, is calculated as

$$\mu_{J_r|RT,I} = \mu_{r|I} - \frac{s\sigma_I^2(RT - b + s\mu_{p|I})}{s^2\sigma_I^2 + \sigma_p^2} \quad (\text{A.2})$$

and the variance, $\sigma_{J_r|ID}^2$, as

$$\sigma_{J_r|ID}^2 = \sigma_I^2 + \sigma_r^2 - \frac{s^2\sigma_I^4}{s^2\sigma_I^2 + \sigma_p^2}. \quad (\text{A.3})$$

In Experiment 2, the likelihood of recognition responses on new item trials without identification RTs was calculated as

$$L(Z|I = \text{new}) = \Phi(\lambda_{j+1}, \mu_{J_r,\text{new}}, \sigma_{J_r,\text{new}}) - \Phi(\lambda_j, \mu_{J_r,\text{new}}, \sigma_{J_r,\text{new}}) \quad (\text{A.4})$$

where $\mu_{J_r,\text{new}} = 0$ and $\sigma_{J_r,\text{new}} = \sqrt{\sigma_r^2 + \sigma_{\text{new}}^2}$.

To fit the DPSD1 model, it was assumed that recollection could only occur for old items in line with Yonelinas (1994). The likelihood of a familiarity-driven response to a given old item was therefore

$$L(Z < 6, ID|\text{old}) = (1 - R)[\Phi(\lambda_{j+1}, \mu_{J_r|ID,I}, \sigma_{J_r|ID}) - \Phi(\lambda_j)] \quad (\text{A.5})$$

Equation A.1 was used to calculate the likelihood of a recognition and identification response for a given new item in the DPSD1 model. For new item trials without identification RTs in Experiment 2, $L(Z|I = \text{new})$ was calculated as stated above for the SS model.

A.2 Experiment 3

In Experiment 3, the likelihood of a recognition response Z to an item type I (old-high, old-low, new-high, new-low) in the extended UVSD model was

$$L(Z, I) = \Phi(\lambda_{j+1}, \mu_I, \sigma_I) - \Phi(\lambda_j, \mu_I, \sigma_I) \quad (\text{A.6})$$

For the extended DPSD model, the likelihood of a new item trial receiving a given recognition response was also given by Equation A.6. For old items receiving a recognition confidence rating between 1 and 5, the following equation was used

$$L(Z < 6|\text{old}) = [1 - R][\Phi(\lambda_{j+1}, \mu_{J_r|RT,I}, \sigma_{J_r|RT}) - \Phi(\lambda_j, \mu_{J_r|RT,I}, \sigma_{J_r|RT})] \times \phi(ID|b - s\mu_{p,I}, \sigma_{ID}^2) \quad (\text{A.7})$$

The extended DPSD model likelihood function for old items that received a 6 rating was

$$L(Z = 6|\text{old}) = (1 - R_I) \times [1 - \Phi(\lambda_5, \mu_{J_r|RT,I}, \sigma_{J_r|RT})] \times \phi(ID|b - s\mu_{p,I}, \sigma_{ID}^2) + R \times \phi(ID|b - s\mu_{p,I}, \sigma_{ID}^2) \quad (\text{A.8})$$

A.3 Experiment 4

In Experiment 4, the SS and DPSD1 models were extended to fit data across four item type conditions, rather than two (old, new). The SS model likelihood function can therefore be expressed as Equation A.1 where I represents a vector of item types old-high, old-low, new-high, and new-low. With this definition of I , Equations A.7 and A.8 can be used for the extended DPSD1 model.

Appendix B

Parameter Recovery Simulations

Each of the parameter recovery simulations for the models in Chapter 3 followed a similar procedure. First, sets of true parameters were generated from uniform distributions with bounds reflecting a range of plausible estimates. These true parameters were then used to generate simulated datasets from each model specification. The relevant model was then fitted to each simulated dataset according to the model fitting procedure detailed in Appendix A. After each simulation, the differences between the true parameter values and the fitted parameter values were analysed. The following sections describe the generation of starting parameter estimates, including the estimation of parameter values from the data, and the results of each simulation.

B.1 Extended UVSD Model

75 sets of "true" parameters were generated. The bounds on the distributions used to generate these parameters are found in Table B.1. 75 datasets with 60 trials in each item characteristic condition, for a total of 240 trials per dataset, were then simulated from the extended UVSD model specification.

Initial estimates of μ_{oh} and μ_{ol} were calculated as

$$\widehat{\mu_{o,K}} = \Phi[P(H_K)] - \Phi[P(FA_K)] \quad (\text{B.1})$$

where $P(H)$ and $P(FA)$ are hit and false alarm rates, and K represents the level of item characteristic variability (high, low). Estimates of μ_{nh} were set to equal zero. To estimate the old item σ parameters, z -ROC slopes were calculated using the same distributions used to estimate each old

Table B.1: Generative lower and upper bounds a and b on the uniform distributions used to simulate true parameters, the means of the true and estimated parameters given by the extended UVSD model (SDs in parentheses), and the Bayes Factors of the comparisons between true and estimated parameters.

Parameter	a	b	True Mean		Mean Estimate		BF
μ_{nh}	-1	1	-0.08	(0.56)	-0.02	(0.73)	0.20
σ_{nh}	0.5	4	2.30	(0.98)	2.50	(1.43)	0.27
μ_{oh}	0	4	2.14	(1.08)	3.13	(6.59)	0.37
σ_{oh}	0.5	4	2.45	(1.02)	2.57	(1.45)	0.20
μ_{ol}	0	4	1.94	(1.19)	2.94	(5.71)	0.47
σ_{ol}	0.5	4	2.33	(1.03)	2.67	(1.90)	0.40
C_1	-1	0.2	-0.42	(0.35)	-0.45	(0.43)	0.20
C_2	$C_1 + 0.01$	$C_1 + 1$	0.09	(0.42)	0.09	(0.48)	0.18
C_3	$C_2 + 0.01$	$C_2 + 1$	0.61	(0.53)	0.63	(0.58)	0.18
C_4	$C_3 + 0.01$	$C_3 + 1$	1.18	(0.58)	1.23	(0.67)	0.20
C_5	$C_4 + 0.01$	$C_4 + 1$	1.68	(0.64)	1.75	(0.75)	0.21

item μ parameter. If a z -ROC slope could not be calculated due to zero or unit response probabilities, the relevant σ parameter was estimated to equal 1.25. The lowest criterion, C_1 , was estimated with the formula

$$\widehat{C}_1 = \Phi\left[1 - \left(\frac{\sum X_{2|N}}{n}\right)\right] \quad (\text{B.2})$$

The difference between each successive criterion and the previous was then calculated as

$$\widehat{\Delta}_{C_i} = \Phi\left[1 - \left(\frac{\sum X_{i+1|N}}{n}\right)\right] - \Phi\left[1 - \left(\frac{\sum X_i|N}{n}\right)\right] \quad (\text{B.3})$$

The sampling distributions for the starting values of the fitting routine are found in Table B.1.

Results

Estimates of each parameter from fits to simulated data were compared with the true parameter values that generated each simulated dataset using Bayesian t -tests. As seen in Table B.1, most of the Bayes Factors from these t -tests were less than 0.33, indicating strong evidence for there being no difference between the true and recovered parameters. There were only three exceptions; inconclusive evidence for a difference was found for μ_{oh} , μ_{ol} , and σ_{ol} , although the Bayes Factors for each of these comparisons were still relatively low. This indicates that the extended UVSD model's parameters can be successfully recovered from simulated data with the same number of trials and datasets as Experiment 3. Some outlying parameter estimates were produced when fitting the UVSD model to these simulated data. These are reflected in the means and standard

Table B.2: Generative lower and upper bounds a and b on the uniform distributions used to simulate true parameters of the extended DPSD model, the means of the true and estimated parameters given by the model (SDs in parentheses), and the Bayes Factors of the comparisons between true and estimated parameters.

Parameter	a	b	True Mean	Mean Estimate	BF
μ_{nh}	-1	1	0 (0.56)	-0.01 (0.84)	0.18
μ_{oh}	0	4	1.94 (1.06)	4.73 (6.91)	0.53
μ_{ol}	0	4	1.90 (1.16)	4.71 (7.05)	1.44
R_h	.1	.6	.37 (.15)	.46 (.30)	0.18
R_l	.1	.6	.35 (.15)	.35 (.28)	0.18
C_1	-1	0.2	-0.38 (0.36)	-0.54 (0.52)	0.18
C_2	$C_1 + 0.01$	$C_1 + 1$	0.14 (0.52)	0.20 (0.76)	0.18
C_3	$C_2 + 0.01$	$C_2 + 1$	0.70 (0.60)	1.06 (1.06)	0.18
C_4	$C_3 + 0.01$	$C_3 + 1$	1.22 (0.68)	1.96 (1.56)	0.20
C_5	$C_4 + 0.01$	$C_4 + 1$	1.77 (0.72)	3.27 (2.00)	1.11

deviations of some parameters, particularly μ_{oh} and μ_{ol} . Although these outliers make the means less representative of the model's general predictions about data, the model can regardless still produce a generally accurate account of the generative parameters of simulated data.

B.2 Extended DPSD Model

75 datasets with 60 trials per item characteristic variability condition were simulated from the model specification. The initial estimates of the old item μ parameters and the decision criteria were calculated from the data in the same way as in the UVSD model. The starting estimates from the remaining free parameters in the model (μ_{nh}, R_h, R_l) were estimated from a uniform distribution in every repetition of the model fitting procedure (see Table B.2).

The results from the fit procedure are also found in Table B.2. As with the UVSD model, there was evidence for an absence of a difference between the true and estimated values for all but three parameters. Bayes Factors for μ_{oh}, μ_{ol} and C_5 indicated inconclusive evidence for a difference between the generative parameters and the parameter estimates given by the DPSD model. In each case, this was likely the result of outlying values given by the fit procedure which, like in the UVSD model, are reflected in the mean and standard deviation for each of these parameters. Despite this, the rest of the model's parameters, including its two recollection parameters, were recovered well.

Table B.3: Generative lower and upper bounds a and b on the uniform distributions used to simulate true parameters of the extended SS model, the means of the true and estimated parameters given by the model (SDs in parentheses), and the Bayes Factors of the comparisons between true and estimated parameters.

Parameter	a	b	True Mean		Mean Estimate		BF
μ_{nh}	-1	1	0.07	(0.54)	0.1	(0.57)	0.18
μ_{oh}	0	2	1.05	(0.61)	1.08	(0.66)	0.19
μ_{ol}	0	2	0.95	(0.56)	1.04	(0.70)	0.83
σ_{nh}	0.5	2.5	1.60	(0.60)	1.64	(0.86)	0.15
σ_{oh}	0.5	2.5	1.51	(0.57)	1.47	(0.74)	0.17
σ_{ol}	0.5	2.5	1.59	(0.59)	1.71	(1.16)	0.26
b	7	9	8.07	(0.59)	8.06	(0.59)	0.14
s	0.1	0.3	0.19	(0.06)	0.19	(0.10)	0.13
σ_p	0.5	1.5	1.05	(0.30)	1.04	(0.30)	0.31
C_1	-1	-0.2	-0.01	(0.11)	0.01	(0.20)	0.20
C_2	$C_1 + 0.01$	$C_1 + 1$	0.51	(0.29)	0.53	(0.31)	0.22
C_3	$C_2 + 0.01$	$C_2 + 1$	0.98	(0.43)	1	(0.45)	0.16
C_4	$C_3 + 0.01$	$C_3 + 1$	1.51	(0.53)	1.55	(0.61)	0.26
C_5	$C_4 + 0.01$	$C_4 + 1$	2.01	(0.64)	2.04	(0.70)	0.19

Table B.4: Generative lower and upper bounds a and b on the uniform distributions used to simulate true parameters of the extended DPSD1 model, the means of the true and estimated parameters given by the model (SDs in parentheses), and the Bayes Factors of the comparisons between true and estimated parameters.

Parameter	a	b	True Mean		Mean Estimate		BF
μ_{nh}	-1	1	-0.02	(0.59)	-0.01	(0.59)	0.16
μ_{oh}	0	2	1.09	(0.55)	1.14	(0.60)	0.54
μ_{ol}	0	2	1.05	(0.58)	1.06	(0.63)	0.14
R_h	0	0.6	0.31	(0.17)	0.30	(0.17)	0.18
R_1	0	0.6	0.30	(0.17)	0.30	(0.18)	0.13
b	7	9	8.09	(0.62)	8.10	(0.60)	0.22
s	0.1	0.3	0.21	(0.06)	0.24	(0.13)	0.86
σ_p	0.5	1.5	0.99	(0.29)	0.97	(0.30)	0.70
C_1	-1	-0.2	0	(0.12)	0.02	(0.17)	0.25
C_2	$C_1 + 0.1$	$C_1 + 1$	0.52	(0.29)	0.55	(0.32)	0.45
C_3	$C_2 + 0.1$	$C_2 + 1$	1.08	(0.40)	1.11	(0.45)	0.31
C_4	$C_3 + 0.1$	$C_3 + 1$	1.64	(0.50)	1.69	(0.60)	0.79
C_5	$C_4 + 0.1$	$C_4 + 1$	2.18	(0.53)	5.26	(13)	0.94

B.3 Extended Models of Recognition and Priming

The extended SS and DPSD1 models were each fitted to 75 simulated datasets with 60 trials per item characteristic variability condition. The parameters of the generative distributions of the true parameters for these models, along with the mean estimated parameters and Bayes Factors comparing these estimates can be found in Tables B.3 and B.4.

For the SS model, Bayesian *t*-tests gave evidence for no difference between all true and estimated parameter values with the exception of μ_{ol} , for which inconclusive evidence was found. Given this and the relatively low inconclusive Bayes Factor for μ_{ol} (BF = 0.83), the extended SS model recovers generative parameters well. For the DPSD1 model, two model fits were excluded for returning large outlying values of the C_5 parameter. Bayesian *t*-tests on the remaining 73 parameter values and model fits indicated no difference between true values and estimated values of seven parameters; μ_{nh} , μ_{ol} , R_h , R_l , b , C_1 , and C_3 . The remaining six parameters yielded inconclusive Bayes Factors, the largest being C_5 with a BF = 0.94.

Overall, the model was able to recover most of its generative parameters, albeit with less precision than the extended SS model. A small number of C_5 parameter estimates were still notably larger than their corresponding true values, as seen in the larger mean and standard deviation for these estimates. This behaviour where C_5 converges to a higher value is likely driven by a lack of responses in some categories in each item characteristic variability condition. In simulations with a higher number of trials and no missing data in any of the response categories, this behaviour is not apparent. For instance, running the same parameter recovery simulation with 200 trials per item characteristic condition yields a mean estimated $C_5 = 2.36$ ($SD = 1.31$), close to the true mean 2.17 ($SD = 2.17$). This means that although the model is recoverable in principle, it may output some psychologically implausible parameter estimates for data with a smaller number of trials. These outliers are rare, however, and can therefore be dealt with by exclusion.

