Is Free Recall Actually Superior to Cued Recall?

Introducing the Recognized Recall Procedure to Examine the Costs

and Benefits of Cueing

by

Jason D. Ozubko

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Psychology

Waterloo, Ontario, Canada, 2011

© Jason D. Ozubko 2011

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

A vast literature and our own common sense tell us that free recall (i.e., recalling information without hints) is harder and less successful than cued recall (i.e., recalling information with hints). In this dissertation, I argue that in past work free and cued recall has not been directly comparable because cued recall procedures encourage guessing and the nature of the cues promotes accurate guesses. These biases often inflate cued recall performance above free recall, creating the illusion that cued recall is superior to free recall. To control for these issues, I introduce the recognized recall procedure. Recognized recall requires subjects to produce a word on every test trial and subsequently to recognize those produced words as "old" or "new." Across eight experiments with recognized recall, it is demonstrated that cueing does help subjects produce more studied words than in free recall, however, subjects are often unable to recognize those extra words produced. Worse yet, false memories are observed to rise in all cases of cueing. Three subsequent experiments demonstrate that cueing fails to improve recall consistently because cues do not always cue the same meaning of the word as was encoded at study. A final experiment demonstrates that free associates of studied words produced by subjects can be highly effective at improving memory if used as cues at test. It is concluded that cues can improve memory if they are specific to the study episode but can often lead to a rise in false memories. Thus, in terms of consistently optimizing accurate recall while minimizing false memories, free recall may actually be superior to cued recall.

Acknowledgments

I would like to thank Colin MacLeod and the University of Waterloo Department of Psychology,

Cognition Division, for their support. I would also like to thank Erica Elderhorst, Grace Hsiao,

and Merrick Levene for their assistance running the many experiments in this dissertation.

Finally, I would like to thank Mike Masson for his helpful discussions of this work.

Table of Contents

(Experiment 3 continued)

(Experiment 5 continued)

(Experiment 7 continued)

(Experiment 10 continued)

List of Tables

Introduction

The act of remembering involves recognition, recall, or both. For example, seeing a person at a distance, you might recognize their face and also recall their name. Recognition requires a decision about whether some stimulus or event is familiar when that stimulus or event occurs in your environment. Recall, in contrast, requires an act of retrieval because the desired stimulus or event is not present. Recall can be done unaided, as when you try to remember the items on the grocery list that you left at home; this is called *free recall*. Or, recall can be done with the addition of helpful hints, as when you think of categories of foods and try to recover the items that were in each category on your list; this is called *cued recall*. This dissertation is an investigation of the relation between free and cued recall and, more precisely, of the impact of cueing on memory.

Free and cued recall are indeed two of the most common measures of memory—both in the "real world" and in the laboratory. In laboratory recall studies, subjects often begin by studying a list of to-be-remembered stimuli presented one at a time. Characteristically, these stimuli are familiar words. Following the study phase, subjects are asked to recall as many of the studied items as possible. In free recall, that is the only instruction given and there are no "hints" given for individual studied items. In cued recall, subjects are given retrieval cues meant to remind them of studied items with the goal of assisting recall. In both cases, memory is measured as the number of items recalled correctly. Occasionally, the number of nonstudied items (i.e., intrusions) recalled is also reported, providing a measure of guessing or false memory.

Across a vast array of different procedures, investigations involving both free and cued recall have consistently found that cued recall performance is superior to that of free recall.

1

Providing retrieval cues that were encoded at the time of study certainly can help participants to retrieve and output more studied items (Paivio et al., 1994; Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971; Watkins & Tulving, 1975). But cues can be useful even when they were not present during study. Thus, semantic associates of studied items can cue memory (Bahrick, 1969; Bilodeau, 1967; Bilodeau & Blick, 1965; Fox et al., 1964; Humphreys & Galbraith, 1975; Lewis, 1974; Postman et al., 1955; Thomson & Tulving, 1970) as can category labels (Cohen, 1966; Dong, 1972; Dong & Kintsch, 1968; Earhard, 1967; Hudson & Austin, 1970; Lewis, 1974; Tulving & Psotka, 1971; Wood, 1967; Slamecka, 1972). Cues can even be presented after subjects have freely recalled as many items as they can, resulting in recall of additional items (Bahrick, 1970; Lauer & Battig, 1972; Mondani et al., 1973; Tulving & Pearlstone, 1966; and see Loess & Harris, 1968, for similar results in short-term memory). The necessary relation between cues and targets is not even restrictive; seemingly anything that can bring the target to mind can act as an effective cue. Cues can be homonyms or synonyms of studied words (Light, 1972); semantic, phonemic, or graphemic associates of studied words (Blaxton, 1989; Bregman, 1968; Ramponi et al., 2007); or even word-stems (i.e., the first few letters) of studied words (Allan et al., 1996; Allan & Rugg, 1998; Angel et al., 2009; Angel et al., 2010a, 2010b; Fay et al., 2005; Rugg et al., 1998; Schott et al., 2002). Finally, cues make recall easier for individuals with memory difficulties, as evidenced by the fact that cued recall performance is often more resistant to aging than is free recall (Perlmutter, 1979; Perry & Wingfield, 1994) and also more spared in amnesics (Buschke, 1984; Isaac & Mayes, 1999a, 1999b).

The finding that cued recall is superior to free recall is so well established in cognitive psychology that it generally is not even questioned. Indeed, Roediger (1973) provides an

example of the common sentiment when he states: "One of the most effective methods of improving memory for a series of briefly experienced episodes—rivaled only by imagery instructions and mnemonic devices—is the presentation of retrieval cues at the time of recall" (p. 644). And this assertion certainly aligns with intuition. Yet despite the overwhelming evidence that cued recall is superior to free recall, I will question this finding in this dissertation. I will argue that a direct comparison between these procedures has in fact yet to be conducted and, hence, almost none of this past evidence can be taken to show that cued recall is superior to free recall. That is, despite their apparent similarity, the free recall and cued recall procedures used in the literature have contained fundamental differences that make direct comparison of them misleading. Furthermore, the nature of cued recall paradigms importantly omits consideration of the subjective state of memorability, making it impossible to disambiguate guesses from true memory.

To eliminate the discrepancy between free recall and cued recall paradigms, and thereby to advance the study of recall, in this dissertation I introduce the procedure of *recognized recall*. This simple procedure better equates free recall and cued recall while still allowing for traditional measures of recall to be calculated. Additionally, recognized recall allows for a more direct measure of subjective memorability than traditional recall paradigms permit, and controls for factors—such as guessing—that are left uncontrolled in traditional recall paradigms.

Before discussing the recognized recall procedure in detail, however, I will first introduce a common framework for understanding how recall is accomplished: the generate-recognize model. Next, I will discuss whether, on theoretical grounds, we should expect cues to assist retrieval. At that point, I will return to the issue of directly comparing free recall and cued recall, and subsequently to the recognized recall procedure.

The Generate-Recognize Model of Recall

The most ubiquitous model of recall is the generate-recognize model. In fact, there exist numerous different versions of this model (e.g., Bahrick, 1970; Anderson & Bower, 1972; Haist et al. 1992; Jacoby & Hollingshead, 1990; Nobel & Shiffrin, 2001; Slamecka, 1972). Thus, the generate-recognize model of recall can be thought of as actually a class of models. Yet despite the details that differentiate the members of this class, they can be discussed in a general sense because all of them generally assume that recall is a two-step process.

In a generate-recognize framework, to recall a studied word, subjects must first generate a candidate. The details of generation vary from model to model, but all models agree that the generation phase is when subjects retrieve candidates for possible output. Furthermore, in all of the versions of the model, once a candidate has been determined, the next step is recognition. Again, the details of recognition can vary among models, but all agree that if a subject recognizes a candidate as a studied item, then it is output; if not, then another candidate is generated and the process continues until no other candidates are generated.

Generate-recognize models of recall have a long history in cognitive psychology. Indeed, William James (1890) even once described recall as a search involving a hierarchy of associations wherein if the search retrieved a desired response, it then had to be recognized before it could be output. This view of recall was well ensconced until, in the 1970's, a new phenomenon—*recognition failure of recallable words*—was put forth as direct evidence against generate-recognize models of recall (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973).

Recognition failure of recallable words, first studied by Tulving and colleagues, is a situation in which subjects apparently are able to recall words that they cannot subsequently

recognize as having been studied (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973; for a review, see Nilsson & Gardiner, 1993). In studies demonstrating recognition failure of recallable words, subjects would first study cue-target word pairs that were weakly semantically associated (e.g., TRAIN-BLACK). At test, subjects were shown strong associates of the targets that they had studied (e.g., WHITE-), and were asked to free associate words to these new cues. Not surprisingly, subjects often free associated targets that had been studied. Subjects were then asked to circle words that had been studied from those that they had produced by association. Finally, subjects were given a standard cued recall test using the weak associate cues that had originally been studied (e.g., TRAIN-). The key finding was that subjects were able to produce studied words via free association that they could not then recognize as having been studied. But when subsequently provided with the studied weak associate cues, subjects were able to recall many of the studied words, including ones that they had just failed to recognize.

According to generate-recognize models, recall is comprised of a generate stage and recognize stage. As such, recall actually subsumes the components of recognition. Because, to recall a word, that word must be both generated and recognized, any word that is recalled must necessarily have been recognized and, therefore, should subsequently be recognized on any other test. Consequently, the discovery that subjects could recall words during cued recall that they had just demonstrated that they were unable to recognize was seen as fundamentally incompatible with generate-recognize models of recall. On the basis of these results, Tulving and colleagues advocated abandoning generate-recognize models of recall in favour of alternative models (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973).

Despite their apparent contradiction with the observation of recognition failure of recallable words, generate-recognize models eventually were shown to be compatible with this phenomenon. One of the main reasons that recognition failure of recallable words was seen as irreconcilable with generate-recognize models, as argued by Tulving and Thomson (1973), was that most generate-recognize models at the time assumed transsituational identity of words (Bower, 1970; Bower et al., 1969; Fox & Dahl, 1971; Kintsch, 1970; Norman, 1968; Shiffrin & Atkinson, 1969; Slamecka, 1972; Underwood, 1972). Transsituational identity is the assumption that a word has but a single unique representation in the mind. By this account, when a word is studied, the unique corresponding representation can be strengthened or tagged. To recognize a word subsequently, subjects evaluate the strength or they attempt to retrieve the tags associated with each trace. Because recall includes both a generate and a recognize phase according to generate-recognize models, and because words were represented by a single node, it was argued that these types of models would predict that recognition failure of recallable words should be an impossible event.

Certainly, there were those who argued against transsituational identity, maintaining that words are all necessarily polysemous (see, e.g., Martin, 1975; Santa & Lamwers, 1974). But to explain recognition failure of recallable words themselves, Tulving and colleagues (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973) took a different tack and put forth the idea of *encoding specificity*. Encoding specificity is another framework which has a relatively long history in psychology. The notion here is that recall is maximized when test conditions match those of study. In simplest terms, as described by Tulving, encoding specificity suggests that only that which is actually stored can be retrieved. Hence, when encoding a study trial, all the elements of the study context will be encoded, to some degree,

along with that trial. If the test context varies substantially from the study context, and therefore contains little to none of the elements that were present at study, then retrieval of the study trial will fail. Such a notion is similar to the ideas of Hollingworth (1928), who suggested that the probability of recall is a function of how similar the test conditions are to the study conditions, and of Melton (1963), who suggested that retrieval success depends on reinstatement of study conditions. The most general form of this concept is transfer appropriate processing (Morris, Bransford, & Franks, 1977).

With regard to recognition failure of recallable words, encoding specificity suggested that every study event led to a relatively unique encoding that featured some but not all of an item's potential semantic elements. So for example, if GLUE-CHAIR were studied as a cue-target pair, the word CHAIR would be encoded in a different manner than if the pair ELECTRIC-CHAIR had been studied. In both cases, the word CHAIR is encoded, but the sense of the word CHAIR is slightly different in the two cases.

To subsequently recall or recognize that CHAIR was studied, some method for retrieving not just the word CHAIR but the specific encoding event (i.e., CHAIR in the context of GLUE) would be necessary. If, at test, CHAIR were generated based on a strong semantic associate that had not been present at study, such as TABLE, it would be entirely possible that this would not lead to the specific episodic trace of GLUE-CHAIR. In other words, according to the encoding specificity account, CHAIR and TABLE are semantically related, but for TABLE to be a good cue for the *specific* study episode of CHAIR that was experienced, TABLE would have had to have been *encoded*, either explicitly or implicitly, along with CHAIR during that study episode. Hence, encoding specificity would predict that the specific study episode including CHAIR should have a high probability of being retrieved when GLUE is used as a cue. When a strong

semantic associate like TABLE is used as a cue, it may lead subjects to generate CHAIR based on longstanding semantic knowledge, but subjects would not necessarily retrieve the specific study episode of GLUE-CHAIR, and hence could easily fail to recognize that they had actually studied CHAIR.

Overall, the notion of encoding specificity—or more generally, transfer appropriate processing—is well supported by our current knowledge of memory. However, even if transsituational identity and encoding specificity are mutually exclusive, neither is more or less compatible with the generate-recognize framework. Many researchers have argued that, although transsituational identity was a common component of early generate-recognize models, it was never a necessary component of any generate-recognize model (see, e.g., Martin, 1975; Santa & Lamwers, 1974, 1976). Indeed, in response to Tulving and colleagues' criticisms, many generate-recognize models were amended to do away with assumptions about transsituational identity, adopting an encoding-specificity-like, multi-trace set of assumptions (Anderson & Bower, 1974; Kintsch, 1974; Reder et al., 1974). Consequently, generate-recognize models that include episodic trace assumptions can easily explain the study-test interaction of recognition failure of recallable words (Jacoby & Hollingshead, 1990).

In sum, the prevailing framework—and the framework advocated here—for understanding recall is the generate-recognize model. Despite criticism of generate-recognize models in the past, they appear to be one of the most enduring, if not *the* most enduring, models of recall. Indeed, at a merely practical level, as Jacoby and Hollingshead (1990) state: "Generate/recognize models are too useful as descriptions of memory monitoring and other activities to be abandoned" (p. 452). With the framework of generate-recognize in mind, it is now time to consider the impact of cueing on recall.

<u>The Effect of Cues on Recall</u>

Putting aside the issue of recognition failure of recallable words temporarily, it seems reasonable on intuitive grounds to expect that cues should help retrieval. That is, when attempting to recall some information, being given hints or cues that relate to that information should be helpful. Indeed, at worst, we might expect cues to do nothing, but at best, they should lead to successful retrieval. From a generate-recognize perspective, what cues should do is very clear. That is, there are only two broad ways in which cues could conceivably help recall: (1) by assisting in the generation of candidates or (2) by assisting in the recognition of candidates.

The first step of recall in the generate-recognize framework is to generate a candidate. Regardless of the details concerning how generation typically occurs, cues may be helpful in generating potential candidates through simple free association. For example, if the desired target is CHAIR and the given cue is TABLE, subjects can attempt to free associate words to TABLE with the goal of finding candidates for recall. Indeed, as long as there is some reliable relation between the cue and the target, and particularly if subjects are aware of this relation, cues should be useful. Hence, semantic, graphemic, and phonemic associates, as well as word stems and category labels, should all help cue memory. In fact, these have all been shown to be effective cues (Allan et al., 1996; Allan & Rugg, 1998; Angel et al., 2009; Angel et al., 2010a, 2010b; Blaxton, 1989; Bregman, 1968; Cohen, 1966; Dong, 1972; Dong & Kintsch, 1968; Earhard, 1967; Fay et al., 2005; Hudson & Austin, 1970; Lewis, 1974; Light, 1972; Ramponi et al., 2007; Rugg et al., 1998; Schott et al., 2002; Slamecka, 1972; Tulving & Psotka, 1971; Wood, 1967).

Alternatively, cues could help in the generation of candidates via a more directed search. Craik (1983) referred to cued recall as providing more environmental support than free recall, in

the sense that cues could direct subjects to constrain their search through memory. In a similar vein, Tulving and Pearlstone (1966) also suggested that cues may act to help retrieval by providing subjects with some guidance in terms of where to look in memory. In any event, whether cues help via free association or a more guided search process, it is conceivable that cues could help the generation phase of retrieval.

The second step in recall under the generate-recognize framework is to recognize candidates as studied or unstudied. Here cues could be helpful in two key ways. First, candidates that are incompatible with the cue-target relation could be rejected without hesitation. That is, if it is known that cues and targets are semantically related and in response to the cue TABLE, the first generated candidate is ALLIGATOR or even FABLE, these candidates could immediately be rejected for being incompatible with the cue.

Another way in which cues could assist recognition is by making studied targets more accessible. That is, some researchers have suggested that the speed with which a candidate is generated may influence its recognition (Müller, 1913; Jacoby & Hollingshead, 1990). If a cue were able to help bring a candidate to mind quite quickly, that candidate might be accepted as studied, regardless of its level of familiarity. In a somewhat similar manner, if a cue does help a candidate come to mind more easily (compared to no cue), then this candidate would be expected to be processed more fluently than a subject might expect, and this could lead to better recognition of such candidates (cf. Whittlesea & Williams, 2001a, 2001b; Jacoby & Whitehouse, 1989; Yonelinas, 2002).

Given the preceding theoretical discussion and the long list of studies cited at the start of this dissertation, it certainly seems that there are both empirical and theoretical grounds on which

to accept that cued recall is inherently easier or better than free recall. Yet there are surprisingly numerous examples in the literature testifying as to the harmful nature of cues.

First of all, as has already been described in detailing recognition failure of recallable words, strong semantic associate cues that were not encoded at study do not always help recall and indeed can impair it (Funkhouser, 1968; Light, 1969; Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971; Watkins & Tulving, 1975). Part-list cueing is another form of cueing where some of the studied words are shown to prompt the retrieval of the rest of the set yet this form of cueing has been quite consistently shown to impair memory (Basden & Basden, 1995; Basden et al., 1977; Nickerson, 1984; Roediger, 1973, 1974; Slamecka, 1968, 1969; although see Lewis, 1971, for evidence that part-list cueing can help if words are blocked by category at study).

In a similar vein, studies attempting to examine the impact of cueing outside the laboratory have consistently found performance costs due to cueing. Although early studies found that pairs or groups of individuals could recall more than lone individuals, suggesting that individuals cross-cue one another effectively (Hoppe, 1962; Meudell, Hitch, & Kirby 1992; Perlmutter & De Montmollin, 1952; Stephenson, Abrams, Wagner, & Wade, 1986; Yuker, 1955), subsequent studies have demonstrated that the cues subjects are providing to one another actually impair memory (Basden et al., 1997b; Basden et al., 2000; Meudell et al., 1995; Weldon & Bellinger, 1997). That is, although groups as a whole recall more than individuals, individuals in groups recall less themselves than if they were isolated. Finally, and worst of all, cues can alter the very memories that they are attempting to retrieve (Loftus & Palmer, 1974; Loftus et al., 1978), creating significant misinformation effects and false memories.

Aside from the empirical evidence suggesting that cueing may not be as beneficial as it first seems, there are theoretical reasons to have reservations. Although the earlier theoretical discussion of how cues could improve memory in a generate-recognize framework still holds, all of the effects of cueing were considered to be acting only on studied words. In practice, however, cues are not necessarily selective to just the studied words. So, for example, cues may help subjects generate studied candidates but may just as easily help them generate unstudied candidates. As a trivial example, imagine that the studied target is the third strongest associate of the cue: The cue in this case would be much more likely to generate an item other than the target. During recognition, cues could cause subjects to dismiss some studied candidates because the subjects view them as incompatible with the cue. On the other hand, cues could lead subjects to accept unstudied candidates if the cues lead those unstudied candidates to be generated quickly or fluently.

The thesis underlying this dissertation is that cued recall is not necessarily superior to free recall. In some circumstances, cued recall may provide some benefits to retrieval, but the costs of cued recall have largely been overlooked until now. The argument is that cued recall has looked so beneficial for so long because of an artifact due to the methodological differences between free and cued recall. In fact, it is my assertion that current free and cued recall paradigms differ to such a degree that the impact of cueing has yet to be examined in unbiased conditions. We simply do not know the relation between the costs and the benefits of cueing.

Methodological Inequalities Between Traditional Free and Cued Recall Paradigms

Others have been aware of this issue between free and cued recall in the past: "…cued recall is not a simple addition to what is basically a free-recall task" (Humphreys & Galbraith, 1975, p. 710). Despite the superficial similarity between traditional free recall and cued recall

paradigms, they do differ significantly. In fact, it is not merely a matter of cues being present during cued recalls test and being absent during free recall tests. Two critical methodological differences between free and cued recall have received scant research attention: (1) There is significantly more response pressure in cued recall than in free recall paradigms, and (2) cued recall much more readily affords guessing. And because neither free recall nor cued recall paradigms currently assess subjective memorability beyond the produced responses, it is impossible to ascertain the extent to which cued recall performance has been biased by guessing.

Complicating matters, response pressure and guessing do not simply act randomly, biasing cued recall scores to be inaccurate. Instead, each of these factors biases performance in cued recall to be artificially superior to that in free recall. In their present form, therefore, it is likely that direct comparisons between free recall and cued recall are mischaracterizing the nature of cueing as more beneficial than it truly is. Before offering a solution, I will elaborate on each of these issues in turn.

Cued Recall Involves More Response Pressure than Free Recall. In nearly all free recall experiments, after studying a list of words, subjects are asked to recall as many studied words as they can. Either when subjects feel that they can no longer recall any more studied words, or when a deadline is reached, they are allowed to terminate the testing procedure. Cued recall, however, works very differently. In cued recall, a set number of cues are given to subjects after the study phase. Subjects then proceed to at least attempt to recall a word on every cued trial. In some cued recall paradigms, subjects are encouraged or even forced to respond to every cue (Bahrick, 1971; Bodner et al., 2000; Nobel & Shiffrin, 2001; Ramponi et al., 2007; Roediger et al., 1973; Taconnat et al., 2008) whereas in other cued recall paradigms subjects can pass when cues bring nothing to mind (Paivio et al., 1994; Roediger, 1973; Tulving & Pearlstone, 1966).

Yet, regardless of which procedure is used, both forms of cued recall test inherently encourage more responses than free recall. That is, whereas in free recall a subject may give up after recalling 9 words even though 20 words were studied, in cued recall, after studying 20 words, being given 20 cues means that at least 20 retrieval attempts will be made, not 9. This is true even if strict instructions to make a response to every cue are not given: Subjects still would at least attempt 20 retrievals.

One immediate consequence of the fact that cued recall often applies greater response pressure than free recall is that cued recall may derive its entire advantage over free recall merely from reminiscence. That is, reminiscence studies have found that if subjects are asked to attempt to recall information on two separate occasions, they tend to produce a certain amount of new information on the second occasion that was not recalled on the first occasion (e.g., Roediger & Payne, 1982; Wheeler & Roediger, 1992). From these studies, it seems reasonable that having subjects continue to try to recall after a point where they may have given up could easily lead to more information ultimately being retrieved. And given that subjects are likely making more retrieval attempts, on average, in cued recall paradigms than in free recall paradigms, it is possible that the entire literature showing a cued recall advantage over free recall may simply be due to more efforts to retrieve.

Cued Recall Encourages Guessing. Aside from reminiscence, a second potential impact of greater response pressure is to promote more guessing. That is, by placing more pressure, either overtly or covertly, on subjects to make a response, subjects may become more prone to making guesses. And if the cues are good cues for studied items, some of those guesses will be correct—even if the subject does not recognize that they are correct. But the experimenter

14

scoring the data will not know whether a given correct response is a guess, so all correct responses in cued recall will be accepted.

It is the case, therefore, that cues not only encourage guessing but they also meet the necessary conditions for giving rise to consistently accurate guesses (Bahrick, 1969; Fox et al., 1964; Freund & Underwood, 1970). That is, in free recall, if no more words can be recalled, there is no real basis on which to start to make guesses. Indeed, subjects may actually be reluctant to produce items of which they are uncertain such that they not only do not guess but they may even set a sufficiently stringent criterion that items that they do recall are not produced. After recalling words that one is confident were studied, one could simply start outputting any word that comes to mind, but that would likely produce many intrusions with little benefit. In cued recall, on the other hand, cues have a necessary regular relation to targets (e.g., semantic or graphemic association). Because cues are predictive of their targets, educated guesses can be made from cues for which no appropriate studied target can be recalled. The fact that a regular relation exists between cues and targets means that these guesses are much more likely to be right than are guesses in free recall and, at the least, the temptation for subjects to guess should be greater in cued recall than in free recall.

Taking this a step farther, an additional way in which cued recall typically encourages guessing is that often only studied words are cued at test. By only providing cues for studied words, cued recall paradigms once again strongly bias subjects toward guessing: If subjects know or realize that only studied words are being cued, this further encourages them to "trust the cues," in the sense that guessing for some cues may yield correct responses (even if the subject actually does not remember having studied the word that they produce based on the cue.) Thus,

cues not only inherently encourage guessing but the usual practice of cueing only studied targets further encourages guessing.

The Consequences of Guessing: Intrusions. One of the consequences of guessing is that intrusion rates—recall of unstudied items—should rise in cued recall paradigms. Indeed, in cases where intrusion rates are reported and are not on floor, they often indicate that more intrusions and therefore more false memories or more guessing are occurring in cued recall than in free recall (Tulving & Osler, 1968; Tulving & Pearlstone, 1966; Slamecka, 1972). Yet, aside from a few notable examples where researchers have given intrusions serious consideration (e.g., Jacoby & Hollingshead, 1990; Loess & Harris, 1968; Nobel & Shiffrin, 2001; Roediger, 1973), intrusion rates are often given little, if any, attention (Bodner et al., 2000; Humphreys & Galbraith, 1975; Tulving & Osler, 1968) and many times are not reported at all (Allan et al., 1996; Allan & Rugg, 1998; Blaxton, 1989; Lewis, 1971, 1974; Paivio et al., 1994; Reder at al., 1974; Rugg et al., 1998; Schott et al., 2002; Taconnat et al., 2008; Thomson & Tulving, 1970; Tulving & Thomson, 1973; Watkins & Tulving, 1975).

The practice of regularly ignoring intrusion rates is regrettable. It is akin to the situation that would exist if recognition memory researchers regularly reported only hits and ignored or downplayed false alarms. Such a practice could lead to serious distortions of phenomena. For example, the pseudoword effect is the finding that pseudowords (e.g., HENSION) produce higher hit and false alarm rates than do words in recognition (Greene, 2004; Joordens, Ozubko, & Niewiadomski, 2007; Ozubko & Joordens, 2008, 2011; Ozubko & Yonelinas, under review). If accuracy were based solely on the hit rate, then we would claim that pseudowords are more memorable than words. However, the difference between hit and false alarm rates is often smaller for pseudowords than for words (Ozubko & Joordens, 2008, 2011; Ozubko & Yonelinas,

16

under review). Because memory is best reflected by taking into account both hit and false alarm rates, pseudowords are actually *less* memorable than words even though they often produce a *higher* hit rate than words. The higher hit and false alarm rates, then, are produced by bias (i.e., the tendency to call items "old" regardless of status), not actual sensitivity in memory.

The astute reader may wonder why intrusions should be a major consideration in recall data considering that intrusions can often be at floor. Interestingly, the fact that intrusions are often at floor may actually make the case for focusing on intrusions stronger, rather than weaker. That is, when two measures are at floor, as intrusion rates in free recall and cued recall often can be, of course they will not differ from one another. However, this does not mean that these two measures are necessarily equivalent. Instead, it may simply indicate that the current methodology is inadequate to assess intrusions. To argue that two measures truly do not differ, both measures must first be raised off floor. Thus, the fact that intrusions are often at floor is not a reason to ignore them, and may simply reflect insensitivity in the current methodology. Indeed, the cases where intrusion rates are reported and are not at floor tend to support the notion that cues do indeed inflate intrusion rates, and hence, that there is a difference to be found in intrusions between free recall and cued recall (Tulving & Osler, 1968; Tulving & Pearlstone, 1966; Slamecka, 1972).

The Consequences of Guessing: Correct Guesses. In a similar vein to the bias account of cueing, a second consequence of guessing is that cued recall scores likely are naturally contaminated with correct responses to which no subjective index of memorability was attached. In other words, correct guesses occur when subjects produce studied words in response to cues that the subjects do not actually remember studying, but that they infer might have been studied—perhaps especially when those items come easily to mind. Whereas free recall scores

are likely comprised primarily of true recall because there is no strong basis for guessing, cues give subjects a basis from which to initiate guessing. Cued recall scores could, therefore, easily be comprised of both true recalls and correct guesses. Such correct guesses would serve to inflate cued recall scores beyond free recall scores, creating the illusion that more items were being remembered in the presence of cues than in the absence of cues.

Using existing cued recall procedures, there is no way to separate correct guesses from true recalls. Yet measuring correct guesses is a relatively trivial matter. In simplest terms, after producing a word, a subject could be asked to decide whether that word is either "old" or "new." That is, subjects could be asked to recognize their own "recalls." Indeed, although it may seem redundant at first, such procedures are already commonplace among researchers examining the event-related potential (ERP) correlates of cued recall (Allan et al., 1996; Allan & Rugg, 1998; Angel et al., 2009, 2010a, 2010b; Fay et al., 2005; Rugg et al., 1998; Schott et al., 2002).

ERP is a highly sensitive technique for measuring the neurocorrelates of cognitive functions. As a result, ERP researchers understand the need to keep their data as clean and accurate as possible (Luck, 2005). Researchers in this area put a great deal of effort into keeping measures as pure as possible and minimizing contamination. The goal of these strict standards is for ERP waveforms to be tied as closely as possible to separate cognitive acts. As a result of these high methodological standards, researchers examining cued recall with ERP recognized that in standard cued recall paradigms, guessing was uncontrolled. If these researchers adopted existing cued recall paradigms, ERP waveforms representing true recall and those representing correct guesses would be merged into the same category and therefore would distort one another. To obtain more accurate ERP waveforms, subjects in these cued recall ERP studies are often asked to classify each word that they produce as "old" or "new." Words classified as old are

taken as recalls whereas words classified as new are taken as guesses. Using this technique, researchers have found notable differences between true recalls and correct guesses, suggesting that guessing does occur in cued recall paradigms and that, just because a studied word is produced, this does not necessarily correspond to the subjective experience of memory (Allan et al., 1996; Allan & Rugg, 1998; Angel et al., 2009, 2010a, 2010b; Fay et al., 2005; Rugg et al., 1998; Schott et al., 2002).

In sum, free and cued recall procedures differ notably in terms of response pressure, propensity for encouraging guessing, and the likelihood that recall scores are being inflated by correct guesses. Tulving and Osler (1968) once argued that the only way to compare free and cued recall was to keep everything constant up until the test. Any differences subsequently noted between free recall and cued recall could then be safely attributed to the retrieval cues. It is my argument that keeping free recall and cued recall equivalent up until the test is not enough: Free recall and cued recall paradigms must be equated as much as possible *during* the test as well. Intrusions should also play a more central role in theorizing about differences between free recall and cued recall, as changes in intrusion rates do indicate a change in the propensity to guess and/or in memory sensitivity. Finally, it is not enough to assume that guessing is not inflating cued recall performance; guessing must be actively measured and controlled. With all this in mind, it is time to turn to the paradigm of recognized recall, which has been designed to better equate free recall and cued recall procedures and to more accurately measure subjective memorability in both.

The Recognized Recall Paradigm

The recognized recall paradigm has been developed to control the testing conditions of free recall and cued recall in a more precise manner than is possible using traditional recall

paradigms. This new paradigm has as its goal to make free recall and cued recall measures more

directly comparable, and to discriminate between guesses and true recall. In simplest terms,

recognized recall involves an equivalent number of test trials for both free recall and cued recall.

On each test trial for both measures, subjects are asked to produce a word. Subjects are

instructed always to produce a studied word if one comes to mind, but failing that any word is

allowed. Once a word has been produced, subjects make a recognition decision, judging whether

the word just produced is "old" or "new." Thus, the number of test trials and the pressure to

respond on each trial is held constant in free recall and cued recall. Furthermore, in both

paradigms, subjects produce studied and unstudied words, and classify them as "old" or "new."

True recall is considered to have occurred only when a subject both produces a studied word *and*

recognizes it as studied. The only difference between free and cued recall in recognized recall

then, is that cues are provided before each test trial in cued recall, whereas nothing is provided

before each test trial in free recall.

Imagine that the word CHAIR had been studied. During test, in free recall, a subject

would be given no test probe and merely be asked to produce a word; in cued recall however, the

test probe might be the word TABLE. In either case, the subject could produce one of two

responses—the studied word CHAIR or some unstudied word such as LEG. If CHAIR was

produced, the subject could correctly recognize it as a studied word by saying "old" (a hit) or fail

to recognize it by saying "new" (a miss). If LEG was produced, the subject could incorrectly

recognize it by saying "old" (a false alarm) or deny it by saying "new" (a correct rejection).

The major benefit of the recognized recall procedure is that it provides a more detailed

assessment of subjective memorability while still providing the traditional measure of recall.

Traditional recall paradigms count the number of studied items produced as the measure of

recall. In contrast, recognized recall counts the number of studied items produced but further assesses whether subjects recognize those produced items as actually having been studied, allowing a more precise measure of recall along with classic measures.

Furthermore, by forcing subjects to produce a number of new items and to make recognition decisions about those items, recognized recall creates the conditions to adequately measure intrusions and guessing. Hits and false alarms can be separated in recognized recall. A significant rise in hits accompanied by a similar rise in false alarms can easily be spotted as an increase in bias to say "yes," rather than as an actual increase in memory sensitivity. Intrusions therefore play a prominent controlling role in recognized recall, allowing the estimation of the degree of guessing in cued recall.[1]

Finally, although not an inherent aspect of the recognized recall procedure per se, presenting cues for both studied and unstudied items at test is preferable. Presenting cues for studied and unstudied words at test doubles the number of test trials compared to the number of study trials. This provides ample opportunity for subjects to produce new items and, hence, ample opportunity to accurately measure false alarms and intrusions. If only cues for studied words were provided at test, were a subject to produce all of the studied words, there would be no way to measure potential intrusion rate and consequently no way to gauge potential bias. Furthermore, when cues for both studied and unstudied items are presented, subjects are discouraged from guessing liberally based on the cues because this would dramatically increase

---

[1] The question might be asked whether recognized recall is biased by the fact that subjects are forced to produce both old and new items during the test. That is, by being forced to produce new items, subjects are forced to guess during the production phase. Subjects may sometimes guess correctly, and produce old items that they never would have produced in a more traditional cued recall paradigm, where guessing is not forced. Recognized recall might then overestimate guessing in cued recall because it forces responses. There is some validity to this criticism although, as Experiment 8 will show, it does not bias the results of recognized recall. Furthermore, since the same requirements are placed on free recall and on cued recall, if recognized recall is biasing subjects, it should bias free recall and cued recall similarly.

their intrusion rate. Therefore, presenting cues for both studied and unstudied words is recommended when using recognized recall, both because it keeps the value of the cues themselves more neutral and because it allows ample opportunity to measure false alarms and intrusions. Even in traditional recall paradigms, the practice of using cues for both studied and unstudied words is advisable to minimize guessing.

Before continuing, it is worthwhile to note that recognized recall does have roots in past work. For example, a number of researchers have used recall paradigms where subjects must recognize their own responses as studied or new. Anderson and Bower (1972) did this using a 6-point confidence scale, whereas Bodner et al. (2000) used remember-know measures (see Gardiner, 1988; Rajaram, 1993; Tulving, 1985). Several other researchers have used elements of the recognized recall procedure, such as having subjects recognize their own produced words as old/new and forcing subjects to produce a word on every test trial (Bahrick, 1971; Bodner et al., 2000; Jacoby & Hollingshead, 1990; Ramponi et al., 2007; Roediger et al., 1993). Indeed, as stated previously, such procedures are becoming commonplace among researchers examining the ERP correlates of cued recall (Allan et al., 1996; Allan & Rugg, 1998; Angel et al., 2009, 2010a, 2010b; Fay et al., 2005; Rugg et al., 1998; Schott et al., 2002). Thus, the concepts behind recognized recall are not completely new to researchers studying cued recall. What is new is the application of recognized recall as a general recall protocol. To date, recognized recall has never been applied to both free recall and cued recall, never been tested and delineated in detail, and never been used to directly study the nature of cueing in and of itself. This is precisely the goal of this dissertation.

As has been highlighted thus far, a major motivation behind this thesis is to introduce and delineate the recognized recall procedure. Specifically, recognized recall represents a practical step forward in recall methodology. Just as receiver operating characteristic curves and process-dissociation procedure have arisen as next generation recognition protocols (above simple yes/no recognition), it is hoped that the work here will demonstrate recognized recall as a next generation recall protocol. This dissertation will show that recognized recall provides measures above and beyond those of traditional recall paradigms, and is able to measure nuanced cueing effects that are missed in traditional recall paradigms.

Although one of the primary goals of this dissertation is to introduce and delineate the recognized recall procedure, there are several theoretical goals of this work as well. At a grand level, the major impetus behind all of this work is to use recognized recall to better understand the cognitive processes of cueing. That is, how do cues alter the search of memory (if at all)? In what ways is the effect of cueing beneficial or costly? And, if there are noticeable costs to cueing, why were these costs overlooked in the past, in traditional recall paradigms?

A major theme throughout this dissertation will be that of the recognition failure of recallable words, and how this phenomenon may be the norm, not the exception. As a reminder, recognition failure of recallable words occurs when subjects are able to produce studied words that they cannot subsequently recognize as having been studied. Across many of the experiments here, it will be demonstrated that recognition failure of recallable words occurs frequently in cued recall but not in free recall. In past work, recognition failure of recallable words has been found only in specially contrived circumstances (i.e., when target words are studied in the context of other unrelated or weakly related words but are then cued at test with a

strong associate). Perhaps for this reason, the phenomenon was believed not to be a general occurrence in more typical recall settings. Thus, two major questions will be addressed are: Why does recognition failure of recallable words occur in typical recall procedures, and why has its occurrence been overlooked in the past? Experiments 5 through 11 will consider this issue in detail. By Experiment 11, an explanation for why recognition failure of recallable occurs so commonly will be put forth.

A second theme will be that of false memories. Although this will be a particular focus in Experiments 3, it will be an enduring theme across this dissertation. The notion of false memories in recall is not new to researchers of memory, however careful consideration of false memories in recall is conspicuously missing from considerations of memory performance. In a review of the field of memory research itself, Koriat, Goldsmith, and Pansky (2000) describe how memory researchers were historically concerned with quantity measures of memory (i.e., how many items were recalled or recognized). However, as the field of memory has grown, focus has shifted from quantity measures to accuracy measures. For example, current recognition researchers usually focus not just on hit rates, but on the difference between hit rates and false alarm rates. In other words, they focus on the relative accuracy of the responses, rather than simply on the number of responses. Interestingly, this shift in focus is often lacking in research on recall.

As has been mentioned already, many researchers using recall downplay intrusions (Bodner et al., 2000; Humphreys & Galbraith, 1975; Tulving & Osler, 1968) or simply do not report intrusions at all (Allan et al., 1996; Allan & Rugg, 1998; Blaxton, 1989; Lewis, 1971, 1974; Paivio et al., 1994; Reder at al., 1974; Rugg et al., 1998; Schott et al., 2002; Taconnat et al., 2008; Thomson & Tulving, 1970; Tulving & Thomson, 1973; Watkins & Tulving, 1975).

An area of emphasis in this dissertation will be the careful consideration of intrusions in recall: What factors give rise to intrusions? How do changes in intrusions illuminate the actual memory sensitivity of subjects? How should intrusions then best be interpreted? And can it be that a studied word that is produced at test and identified as "studied" is sometimes an intrusion? Although this last point may seem counter-intuitive at first, in the General Discussion the case for the existence of correct intrusions (i.e., "correct" recalls that may actually be false memories) is made.[2] And in fact, if such intrusions do exist, their study may be one of the future directions of recall research.

In sum, there are strong empirical and theoretical goals to this thesis. On the empirical side, this dissertation aims to introduce and delineate the recognized recall procedure. The goal is to make recognized recall accessible to recall researchers by describing it well, and to demonstrate its superiority to traditional measures of recall. In line with this second goal, the theoretical motivations of this dissertation aim to use recognized recall to examine recall and cueing more closely than traditional recall paradigms allow. Specifically, in this dissertation it will be demonstrated that recognition failure of recallable words occurs frequently in cued recall but not free recall, and the factors that give rise to this effect will be examined in detail. In terms of false memories, a close examination of the factors that give rise to false memories and the idea of correct intrusions will be examined. In this way, this dissertation aims to make both empirical and theoretical contributions through the use of the recognized recall procedure.

---

[2] Because of the speculative nature of correct intrusions at the current time, they are not discussed in detail throughout this dissertation. However, suffice to say that they represent an interesting step forward in our understanding of recall.

Experiment 1:

Free vs Cued Recall

Numerous studies have demonstrated that cued recall using extra-list semantic associates of studied items as cues leads to performance superior to that of free recall (Bahrick, 1969; Bilodeau, 1967, Bilodeau & Blick, 1965; Fox et al., 1964; Humphreys & Galbraith, 1975; Lewis, 1974; Postman et al., 1955; Thomson & Tulving, 1970). To date, only a few studies have used recognized recall, and they have focused on word-stem cues, have not compared cued recall to free recall, and have only sporadically considered intrusions (Allan et al., 1996; Allan & Rugg, 1998; Angel et al., 2009, 2010a, 2010b; Fay et al., 2005; Rugg et al., 1998; Schott et al., 2002). The aim of the first experiment was simply to examine free recall and cued recall using extra-list semantic associates as cues both in a traditional recall paradigm and in the recognized recall paradigm. The goal was to see how recognized recall compares to traditional recall as well as how free recall and cued recall compare within recognized recall.

Based on past work, more studied items should be produced in cued recall than in free recall, in both the traditional recall and the recognized recall paradigms. Further, because traditional recall paradigms count the number of studied items produced as the number of items "recalled," the number of studied items produced in recognized recall should be comparable to the number of items "recalled" in traditional recall. As for intrusions, with no good reason to think otherwise, the recognized recall paradigm should provide a similar estimate of intrusions to that provided by traditional recall paradigms. Although intrusion rates are often at floor in free recall and cued recall, in cases where they are not, cued recall ordinarily leads to more intrusions than free recall (e.g., Tulving & Osler, 1968; Tulving & Pearlstone, 1966; Slamecka, 1972). Given that recognized recall provides ample opportunity to measure intrusions, more intrusions

should be observed in cued recall than in free recall. As well, if intrusions truly are rarities in traditional recall paradigms, there should be more intrusions in recognized recall than in traditional recall in general (although if intrusions are simply underreported, there may be similar levels of intrusions across the two paradigms).

Finally, recognized recall will provide measures of recognition that traditional recall paradigms do not (i.e., hits and false alarms). Furthermore, for an item to be considered to be recalled in recognized recall, subjects not only must produce a studied item but they also must recognize it as "old." Whether more studied items are recalled (i.e., both produced *and* recognized as studied) in cued recall than in free recall under the recognized recall paradigm is an empirical question. Although past work has frequently shown a recall advantage for cued recall over free recall, in light of the discussion in the Introduction, there is no reason to expect any difference here.

Before proceeding to Experiment 1, it should be noted that the free recall and cued recall conditions using recognized recall will be referred to simply as the Free Recall and Cued Recall conditions. This is because recognized recall will be the recall paradigm of interest throughout this dissertation. Traditional recall is examined only in this first experiment as a bridge to the existing literature. As a result, to distinguish them, the free recall and cued recall conditions using the traditional recall paradigm will be referred to as the Traditional Free Recall and Traditional Cued Recall conditions.

Method

Participants. All subjects were students from the University of Waterloo who participated in exchange for course credit toward a psychology course. Twenty-five students participated in the Free Recall condition. To ensure reliability, the Cued Recall condition was

run twice, once with 28 students and once with 33 students. There were no significant differences between these replications, so they were collapsed into a single Cued Recall condition with 61 subjects. Finally, 37 students and 28 students participated in the Traditional Free Recall and Traditional Cued Recall conditions, respectively.

Materials. A word pool of 156 items was created from the free association norms of Nelson, McEvoy, and Schreiber (2004). Nelson et al. collected norms by showing participants a series of cue words and recording the targets produced by free association. For present purposes, the backward association norms compiled by Nelson et al. from their data were of principal interest. The backward association norms are arranged by target words instead of cue words. For each target word, a list of the cue words that gave rise to that target word is provided, as well as the probability that each cue word would give rise that that particular target word during free association. So for example, if RIGHT was the target word of interest, Nelson et al. have listed that LEFT, WRONG, CORRECT, and ACCURATE are all cue words which gave rise to RIGHT during free association, with probabilities of .93, .72, .23, and .16, respectively.

In selecting items from Nelson et al.'s (2004) norms, cases of repetition either between the target and cues or within the cues of different items were eliminated. For example, if PRINCESS was a target itself but for the target KING, PRINCESS was also a cue, then one of these items was eliminated so that there would be no repetition of words in this study. Additionally, if UNIVERSE was a cue for the target word WORLD but also for the target word GALAXY, then one of these items was eliminated, again to avoid repetitions of words in the stimulus set.

Cue words were also selected based on the probability with which they would retrieve their respective targets. In all cases, the cues selected were those with the highest probabilities of

giving rise to their respective targets. The end result of the repetition and cue trimming was a set

of 156 target words and the best cue word for each target based on free-association. On average,

the probability that the cues would give rise to their respective target words, based on the

norming data, was .30 (*SD* = .24).

Procedure. Participants were welcomed into the lab and, after reading and signing a

consent form, engaged in a serial presentation study phase on a computer. During this study

phase, words were individually presented in the center of the screen for 2 s with a 0.5 s inter-

stimulus interval. For each participant, 24 words were randomly selected from the pool of 156

target words to be used as study items. The study phase was identical in the Free Recall and

Cued Recall conditions.

After the study phase, participants were instructed as to the nature of the test phase. In

both Free Recall and Cued Recall conditions, participants were instructed that the test would

consist of a series of trials. On each trial, subjects were told to produce a word. Subjects were

told to produce a studied word if one came to mind, but failing that any new word could be

produced. Subjects were told not to repeat words (studied or unstudied), and to keep new words

relatively unrelated to one another. Subjects produced words by typing them on the keyboard,

followed by the ENTER key. Subjects were informed that, after producing a word, that word

would re-appear in the center of the screen along with labels "old" and "new" below it. Subjects

were told to press the M key to indicate that a word was old (i.e., studied) or the C key to

indicate that a word was new (i.e., unstudied). The test phase consisted of 48 trials.

The only difference between the Free Recall and Cued Recall conditions was that an

extra-list semantic associate cue was provided before each test trial in the Cued Recall condition.

The 24 cues for the studied words were randomly inter-mixed with 24 cues for unstudied words.

A single cue was presented to subjects before each test trial. Cues were presented for 2 s, followed immediately by the production prompt. Although cues to both studied and unstudied words were presented at test, subjects were not informed as to the nature of the cues.

Finally, the Traditional Free Recall and Traditional Cued Recall conditions were run identically to the Free Recall and Cued Recall conditions except that subjects were only asked to produce studied words. Subjects were not asked to produce new words on trials where no studied word could be produced; instead, they were told to press ENTER to skip those trials. Unlike in the recognized recall conditions, no recognition decision was requested of subjects on any trials in the Traditional Free Recall or Traditional Cued Recall conditions.

Results

An alpha level of .05 was used to identify results throughout all of the analyses reported in this dissertation. Effect size estimates were computed using Partial η^2 ($p\eta^2$), or Cohen's d where appropriate. The results of the Free Recall and Cued Recall conditions of Experiment 1 can be seen in Table 1.

It should be noted that all recall protocols here, in every experiment, used lenient scoring. Lenient scoring involves counting items as successfully recalled regardless of whether they were produced to the wrong cue, and is a practice that has been used many times in past literature (Tulving & Osler, 1968; Thomson & Tulving, 1970; Humphreys & Galbraith, 1975; Blaxton, 1989). Hence, although each cue at test was intended to cue a certain studied or unstudied word, scoring did not take account of the correspondence between cue-target pairs specifically (i.e., as opposed to strict scoring that would score targets as correct only when they were produced to the correct cue). Although there is nothing incorrect about strict scoring, lenient scoring gives cues the greatest chance to improve memory beyond free recall, as strict scoring procedures would not

Table 1. Mean number of words produced, hit and false alarm rates of produced words, number of words recalled, and number of intrusions in Experiments 1, 3, 4, and 8. Standard errors are shown in parentheses below means.

Condition	Number of Productions			p("old")		Overall Performance	
	Old	New	Repeats	Old	New	Recalls	Intrusions
Experiment 1							
Free Recall	8.72	37.32	1.48	0.97	0.06	8.52	2.20
	(0.72)	(0.84)	(0.47)	(.01)	(.01)	(0.74)	(0.57)
Cued Recall	10.54	34.69	2.56	0.78	0.13	8.31	4.62
	(0.46)	(0.65)	(0.34)	(.02)	(.01)	(0.49)	(0.53)
Experiment 3							
Cued Recall	12.43	32.14	3.07	0.77	0.29	9.82	9.50
8 Cues/ Trial	(0.79)	(0.92)	(0.40)	(.04)	(.03)	(0.92)	(1.12)
Experiment 4							
Cued Recall	15.62	30.14	2.00	0.78	0.16	12.14	4.67
Strong Associate Cues	(0.63)	(0.78)	(0.54)	(.03)	(.03)	(0.68)	(0.93)
Experiment 8							
Cued Recall	8.87	5.48	0.00	0.99	0.12	8.81	4.58
No Response Pressure	(0.65)	(1.49)	-	(.01)	(.03)	(0.65)	(1.08)

count words that were recalled to the wrong cues. Furthermore, given that lenient scoring is the more common method for scoring cued recall, it was adopted as the standard method throughout this dissertation.

Traditional Recall. In terms of the Traditional Free Recall and Cued Recall conditions, consistent with past work, more words were recalled in the cued recall condition ($M = 10.04$, $SE = 0.71$) than in the free recall condition ($M = 8.32$, $SE = 0.42$), $t(63) = 2.18$, $d = 0.55$. Also consistent with past work that has examined intrusions, more intrusions were noted in the cued recall condition ($M = 4.50$, $SE = 0.76$) than the free recall condition ($M = 0.92$, $SE = 0.20$), $t(63) = 5.12$, $d = 1.29$. The relatively high rate of intrusions observed in this experiment is in disagreement with past studies that have found relatively low levels of intrusions. Given that intrusions often are not reported (e.g., Allan et al., 1996; Allan & Rugg, 1998; Blaxton, 1989; Lewis, 1971, 1974; Paivio et al., 1994; Reder at al., 1974; Rugg et al., 1998; Schott et al., 2002; Taconnat et al., 2008; Thomson & Tulving, 1970; Tulving & Thomson, 1973; Watkins & Tulving, 1975), the results here suggest that intrusions may actually be more frequent than is usually assumed, and are merely under-reported in the existing literature. Especially considering that no special instructions were given that would have biased subjects toward producing new items (e.g., subjects were not encouraged to make responses or to guess), the relatively high level of intrusions is somewhat surprising.

Before considering the results of the recognized recall conditions in detail, it is important to consider how traditional and recognized recall compare. Specifically, the prediction was that the number of studied items produced in recognized recall should be the same as the number of items "recalled" in traditional recall. Examining Table 1, it is clear that this prediction was verified: The number of studied words produced in recognized recall corresponds closely to the

number of items recalled in traditional recall. Namely, the number of studied words produced in the Cued Recall condition did not differ from the number of words recalled in the Traditional Cued Recall condition, $t(87) = 0.60$, $p = .55$, $d = 0.13$. Similarly, the number of studied words produced in the Free Recall condition did not differ from the number of words recalled in the Traditional Free Recall condition, $t(60) = 0.51$, $p = .62$, $d = 0.13$.

Finally, the number of intrusions in traditional recall was expected to be fewer than or equal to the number of intrusions in recognized recall. In fact, both of these expectations were met: The number of intrusions in cued recall in both recognized recall and traditional recall did not differ, $t(87) = 0.13$, $p = .90$, $d = 0.03$, although in free recall more intrusions were noted in recognized recall than in traditional recall, $t(60) = 2.45$, $d = 0.62$. Given that in traditional free recall subjects are permitted to quit the recall test whenever they want, these results support the idea that recognized recall provides a more sensitive measure of false memory in free recall. However, the more important conclusion to draw from these results is that, generally speaking, recognized recall does appear to provide very similar measures to those of traditional recall. Especially when we consider the number of items "recalled," recognized recall can provide the same measures as traditional recall paradigms. However, the benefits of recognized recall over traditional recall paradigms are the recognition measures that it provides and its ability to estimate true recall more accurately. Analysis of those measures is next.

Recognized Recall. To examine recognized recall, the Free Recall and Cued Recall conditions in Experiment 1 were analyzed. First, in terms of the number of items produced, only the number of old items produced is of interest. That is, because the number of new items produced is equal to the number of test trials less the number of studied items produced (assuming that repetitions and non-responses are minimal), the number of new items produced

will always be linearly related to the number of old items produced. Hence, testing both the

number of old and new items produced would be redundant. The number of new items produced

in each experiment will be reported, along with the number of repetitions, however analyses of

repetitions will not be presented and only occasionally will tests of new items be reported, when

the fact that they are decreasing is relevant for understanding differences seen or not seen in

intrusion rates.

In traditional recall paradigms, the measure of recall is the number of studied items

produced. Consistent with past research, in the Cued Recall condition, subjects produced more

studied items than they did in the Free Recall condition, $t(84) = 2.12$, $d = 0.46$. Recognition

accuracy was first analyzed in a 2 (old vs new) X 2 (Free Recall vs Cued Recall) mixed

ANOVA. There were more hits than false alarms, $F(1, 84) = 1660.39$, $MSe = 0.01$, $p\eta^2 = .95$,

and more "old" responses in Free Recall than in Cued Recall in general, $F(1, 84) = 6.08$, $MSe =$

0.02, $p\eta^2 = .07$, however, these two factors interacted, $F(1, 84) = 48.79$, $MSe = 0.01$, $p\eta^2 = .37$.

Follow-up analyses indicated that this interaction was driven by the fact that there were both

more hits, $t(84) = 5.24$, $d = 1.14$, and fewer false alarms, $t(84) = 3.00$, $d = 0.66$, in the Free Recall

condition than in the Cued Recall condition.

Finally, recall and intrusion rates were first analyzed in a 2 (recall vs intrusion) X 2 (Free

Recall vs Cued Recall) mixed ANOVA. Of course, there were more recalls than intrusions

generally, $F(1, 84) = 56.38$, $MSe = 15.75$, $p\eta^2 = .40$. Despite the fact that there was no overall

difference between Free Recall and Cued Recall, $F(1, 84) = 3.22$, $MSe = 13.51$, $p = .08$, $p\eta^2 =$

$.04$, there was a borderline interaction, $F(1, 84) = 3.90$, $MSe = 15.75$, $p = .05$, $p\eta^2 = .04$.[3] A

priori follow-up analyses revealed that although Free Recall and Cued Recall had equivalent

---

[3] Although this borderline interaction may cause hesitation for some, these results will be replicated again and
again throughout this dissertation.

recall rates, $t(84) = 0.23$, $p = .82$, $d = 0.05$, there were significantly more intrusions in Cued

Recall than in Free Recall, $t(84) = 2.66$, $d = 0.60$.

Discussion

Experiment 1 examined free recall and cued recall in both recognized recall and

traditional recall paradigms. The first important finding was that recognized recall does indeed

provide the same measures of "recall" that traditional recall paradigms do. Namely, the number

of studied items produced in recognized recall matched closely with the number of items

"recalled" in traditional recall paradigms. Furthermore, the numbers of intrusions observed in

both paradigms were very similar, although more intrusions in free recall were observed in

recognized recall than in traditional recall. This difference most likely results from the fact that,

in traditional free recall paradigms, subjects may terminate the recall test whenever they desire,

meaning that opportunities to measure false memory are limited. Hence, recognized recall may

indeed be the more sensitive paradigm in this situation. Generally speaking, then, recognized

recall provides the same measures as traditional recall paradigms, while also providing more in

depth measures of memorability.

Specifically, recognized recall allows the measurement of recognition accuracy for

produced items. And this fact turns out to be non-trivial. That is, despite the fact that cues

increased the number of studied items that were produced compared to free recall, the

recognition accuracy for these items was impaired by the presence of the cues. The net result is

that there was no true rise in correct recall despite a significant rise in intrusions. Therefore,

using recognized recall has demonstrated that cues appear to provide no real benefit to actual

recall and merely increase false memories (i.e., intrusions).

Although there exists a large body of past work showing that cued recall is superior to free recall (Bahrick, 1969; Bilodeau, 1967, Bilodeau & Blick, 1965; Fox et al., 1964; Humphreys & Galbraith, 1975; Lewis, 1974; Postman et al., 1955; Thomson & Tulving, 1970), the remarkable aspect of the present work is that it does not contradict these past findings. That is, traditional recall paradigms take the number of items produced as their measure of recall. Indeed, cues do help subjects to produce more studied items than they do in free recall, and these benefits were identical to those of traditional recall paradigms. However, the discovery here is that subjects consistently fail to recognize studied items that they have produced as studied. This finding aligns with past work on recognition failure of recallable words (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973) and actually suggests that recognition failure may not be a special case of cueing but indeed may occur regularly in the presence of cues. All of the subsequent experiments will use recognized recall; having compared the two methods, traditional recall will not be tested again.

Experiment 2:

Providing Cues after Free Recall

In examining the effectiveness of cues, cued recall and free recall performance are usually assessed independently. However, many researchers have demonstrated that presenting cues after free recall can effectively increase recall scores (Bahrick, 1970; Lauer & Battig, 1972; Mondani et al., 1973; Tulving & Pearlstone, 1966). Although Experiment 1 demonstrated that cued recall did not lead to superior performance compared to free recall, it remains possible that cues can be effective after an opportunity for free recall. Thus, in Experiment 2, a free recall and then a cued recall test were presented to all subjects. As a reminder, Experiment 2 (and all subsequent experiments) used only the recognized recall procedure. Experiment 2 therefore will permit replication of the Free Recall condition of Experiment 1 and will allow assessment of whether a cued recall test provided after a free recall test can demonstrate a clear benefit of cueing.

Method

Participants. Thirty-five students from the University of Waterloo participated in Experiment 2 in exchange for course credit toward a psychology course.

Materials. Experiment 2 used the same stimulus set as Experiment 1.

Procedure. Experiment 2 was run identically to the Free Recall condition of Experiment 1, except that at the conclusion of the Free Recall test, subjects were given a cued recall test identical to the Cued Recall condition of Experiment 1.

Results

The overall results of Experiment 2 can be seen in Table 2. First, the free recall data of Experiment 2 did replicate the free recall data of Experiment 1. That is, the number of studied

Table 2. Mean number of words produced, hit and false alarm rates of produced words, number of words recalled, and number of intrusions in the free recall and cued recall tests in Experiment 2. Cued recall results are further conditionalized on whether they are repetitions from the free recall test (Across-Test Repetitions) or new productions (Novel Contributions). Standard errors are shown in parentheses below means.

Condition	Number of Productions			p("old")		Overall Performance	
	Old	New	Repeats	Old	New	Recalls	Intrusions
Free Recall	8.69	37.89	0.91	0.92	0.08	8.23	3.11
1st Test	(0.67)	(0.66)	(0.18)	(.02)	(.02)	(0.71)	(0.83)
Cued Recall	10.97	35.37	1.63	0.83	0.22	9.49	8.06
2nd Test	(0.72)	(0.78)	(0.22)	(.03)	(.03)	(0.75)	(1.05)

Old and New Items from Cued Recall Conditionalized on Free Recall

Condition	Number of Productions			p("old")		Overall Performance	
	Old	New	Repeats	Old	New	Recalls	Intrusions
Novel Contributions	5.17	33.43	-	0.70	0.21	3.86	7.17
	(0.44)	(0.80)	-	(.05)	(.03)	(0.43)	(0.96)
Across-Test Repetitions	5.80	1.97	-	0.95	0.45	5.63	0.86
	(0.63)	(0.27)	-	(.03)	(.08)	(0.62)	(0.16)

items produced during free recall did not differ in the two experiments, $t(58) = 0.03$, $p = .97$, $d = 0.01$. Furthermore, a 2 (old vs new) X 2 (Experiment 1 vs 2) mixed ANOVA examining recognition accuracy for free recall found no difference between experiments, $F(1, 58) < 1$, and no interaction, $F(1, 58) = 2.31$, $MSe = 0.02$, $p = .13$, $p\eta^2 = .04$. Similarly, a 2 (recalls vs intrusions) X 2 (Experiment 1 vs 2) mixed ANOVA for free recall also found no difference between experiment and no interaction, both $F(1, 58) < 1$.

In terms of cued recall, no direct comparison was made to Experiment 1 because the cued recall data from Experiment 2 are potentially contaminated from the prior free recall test. As a result, the cued recall results of Experiment 2 are not directly comparable to the Cued Recall condition of Experiment 1, and would not necessarily be expected to replicate that condition. Thus, the most appropriate question to ask about the cued recall results of Experiment 2 is not whether they replicate the cued recall results of Experiment 1, but rather whether cues helps subjects recall words that were not recalled during free recall. As a result, cued recall results were conditionalized on whether they were novel contributions to recall performance or repetitions from free recall (i.e., across-test repetitions). These conditionalized results are presented in Table 2. Novel contributions and across-test repetitions were compared to the free recall results of Experiment 2.

In terms of the number of studied words produced, cued recall did lead to production of a significant number of novel studied items that were not produced during free recall, $t(34) = 11.64$, $d = 3.99$. Thus, in line with past work, cued recall after free recall does help subjects to produce further studied words (Bahrick, 1970; Lauer & Battig, 1972; Mondani et al., 1973; Tulving & Pearlstone, 1966). However, free recall and cued recall do not overlap perfectly, as seen by the fact that across-test repetitions did not equal the number of studied items produced in

free recall, $t(35) = 9.34$, $d = 1.60$. Thus, cued recall did not simply boost standard free recall performance. Some items that were produced during free recall were produced again with cues, and some were not.

Recognition accuracy was analyzed in a 2 (old vs new) X 2 (novel contribution vs free recall) within-subjects ANOVA. Overall there were more hits than false alarms, $F(1, 34) = 273.11$, $MSe = 0.06$, $p\eta^2 = .89$, and although there was no global difference between conditions, $F(1, 34) = 2.53$, $MSe = 0.03$, $p = .12$, $p\eta^2 = .07$, there was a significant interaction, $F(1, 34) = 57.47$, $MSe = 0.02$, $p\eta^2 = .63$. This interaction indicated that the novel contributions in the cued recall test had both a lower hit rate and a higher false alarm rate than items produced in the free recall test, $t(34) = 4.60$, $d = 0.88$ and $t(34) = 5.68$, $d = 0.99$, respectively. Thus, as in Experiment 1, cueing once again impaired the ability to recognize items produced from cues. Subjects missed more targets and accepted more lures in the presence of cues than in their absence.

Discussion

Overall, cues did lead to a significant number of novel studied items being produced in cued recall beyond those produced in free recall. Yet, the number of novel items produced was not exceptional. Furthermore, cues did impair the ability to recognize studied and unstudied items. These two factors combined, in this case, to ultimately produce more novel intrusions than true recalls, $t(34) = 3.00$, $d = 0.54$. Thus, despite replicating past work showing that cues can help in the production of novel studied items after a free recall test, cueing ultimately has a negative impact, insomuch as it leads to a dramatic rise in intrusions. Furthermore, cueing was not a simple addition to free recall, as evidenced by the fact that some words produced in free recall were not reproduced during cued recall. This fits well with the literature on fluctuations in recall (e.g., Tulving, 1967).

Experiment 1 and Experiment 2 are in agreement that cues may not be as beneficial as they have been characterized in the past. Experiment 1 shows that cues appear to provide no real benefit to memory and seem to act only to increase intrusions. Experiment 2 demonstrates that, although providing cues after a free recall test can help subjects to produce some novel studied items, ultimately adding cues as a second test may actually result in worse overall memory compared to simply providing cues on a single test, or not providing cues at all. Although the obvious conclusion may be to claim that cueing is not an effective mnemonic, it remains possible that the cues used thus far have simply not been effective enough to demonstrate the superiority of cued recall over free recall. Hence, the next two experiments are direct attempts to increase the effectiveness of cueing.

Experiment 3:

Increasing the Number of Cues

If the lack of a cued recall advantage thus far is due to the fact that the cues selected here have not been effective in cueing the relevant target sufficiently, one way to rectify this situation could be to provide multiple cues on each test trial. That is, instead of providing a single extra-list semantic associate cue for each target, provide multiple extra-list semantic associate cues for each target before each test trial. Increasing the number of cues for each item in this manner should cue subjects more selectively to the appropriate target for which the cues were intended and, perhaps, show that cues can provide a memory benefit over free recall.

One potential drawback to this method is that providing multiple cues to studied targets means providing multiple cues to unstudied targets as well. Past researchers have shown that the fluency with which information is processed can give rise to subjective feelings of familiarity (cf. Whittlesea & Williams, 2001a, 2001b; Jacoby & Whitehouse, 1989; Yonelinas, 2002). Given that multiple cues should prime, either explicitly or implicitly, a target word much more effectively than a single cue, targets in Experiment 3 might be processed more fluently than those in previous experiments. Then, if unstudied targets also are processed more fluently, they should give rise to a greater sense of subjective familiarity. Consequently, subjects should be more likely to false alarm to these unstudied targets and, hence, to produce more intrusions.

Thus, although providing multiple cues at test may help subjects to arrive at studied targets more easily, these cues may also prime unstudied targets more effectively, increasing both false alarm and intrusion rates. Whether the mnemonic advantage to multiple cues (if any) outweighs the potential rise in false memories is an empirical question, one that will be

addressed by comparing the number of studied items produced and recognized in Experiment 3 to the previous experiments.

Method

Participants. Twenty-eight students from the University of Waterloo participated in Experiment 3 in exchange for course credit toward a psychology course.

Materials. Experiment 3 used the same word pool as Experiment 1. This word pool contained lists of 156 target items and the probabilities with which the top associates gave rise to those targets. In the Cued Recall condition of Experiment 1, cues were always the top associate for the relevant target. In Experiment 3, the top 8 associates of each target were selected to be used as cues. Thus, in Experiment 3, more cues were used on each test trial than in Experiment 1, although these extra cues were weaker, on average, than the top associate. The mean probability that the any one of the top 8 associate cues would give rise to their respective target words was .13 ($SD = .13$).

Procedure. Experiment 3 was run identically to the Cued Recall condition of Experiment 1 except that instead of a single strongest semantic associate cue being provided before each test trial, the top 8 semantic associates of the intended target were used as cues. The order with which the 8 cues were presented was randomized for each trial, and each cue was presented for 2 s with no inter-stimulus interval.

Results

The results of Experiment 3 can be seen in Table 1. For comparison purposes, because the data did not differ, the Free Recall condition in Experiment 1 and the free recall data from Experiment 2 were collapsed into a single free recall comparison set. The Cued Recall condition from Experiment 1 was used as the cued recall comparison set. Here, and in all subsequent

experiments, unless specified otherwise, whenever a condition is compared to free recall, the comparison is to this free recall comparison set, and whenever a condition is compared to cued recall, the comparison is to this cued recall comparison set. These sets will be referred to as the free recall control condition and the cued recall control condition.

Compared to Free Recall. In Experiment 3, where cueing used 8 cues, more studied items were produced than in the free recall control condition, $t(87) = 4.23$, $d = 0.91$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Experiment 3 vs free recall) mixed ANOVA. More hits were observed than false alarms, $F(1, 87) = 808.51$, $MSe = 0.02$, $p\eta^2 = .90$, and although there was no overall performance difference between Experiment 3 and the free recall control, $F(1, 87) = 1.53$, $MSe = 0.02$, $p = .22$, $p\eta^2 = .02$, old/new status and condition interacted, $F(1, 87) = 66.51$, $MSe = 0.02$, $p\eta^2 = .43$. Follow-up comparisons showed that hits were lower and false alarms were higher in Experiment 3 compared to the free recall control, $t(87) = 4.40$, $d = 0.94$, and $t(87) = 7.67$, $d = 1.64$, respectively.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 3 vs free recall) mixed ANOVA. Recall and intrusion rates were higher in Experiment 3 than in free recall, $F(1, 87) = 43.97$, $MSe = 15.23$, $p\eta^2 = .34$, and although there were more recalls than intrusions in general, $F(1, 87) = 12.92$, $MSe = 25.38$, $p\eta^2 = .13$, this interacted with condition, $F(1, 87) = 10.24$, $MSe = 25.38$, $p\eta^2 = .11$. Specifically, although there were more intrusions in Experiment 3 than in the free recall control, $t(87) = 6.31$, $d = 1.35$, there was no difference in terms of correct recalls, $t(87) = 1.61$, $p = .11$, $d = 0.34$.

Compared to Cued Recall. In Experiment 3, more studied items were produced than in the cued recall control condition, $t(87) = 2.17$, $d = 0.47$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Experiment 3 vs cued recall) mixed ANOVA. More hits were observed

than false alarms, $F(1, 87) = 588.47$, $MSe = 0.02$, $p\eta^2 = .87$, and although there was an overall

difference between Experiment 3 and the cued recall control, $F(1, 87) = 6.72$, $MSe = 0.03$, $p\eta^2 = $

.07, old/new status and condition interacted, $F(1, 87) = 12.99$, $MSe = 0.02$, $p\eta^2 = .13$. Follow-up

comparisons showed that, although false alarm rates were higher in Experiment 3 compared to

the cued recall control, $t(87) = 5.25$, $d = 1.13$, hit rates did not differ, $t(87) = 0.14$, $p = .89$, $d = $

0.03.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment

3 vs cued recall) mixed ANOVA. Recall and intrusion rates were higher in Experiment 3 than in

the cued recall control, $F(1, 87) = 22.04$, $MSe = 17.76$, $p\eta^2 = .20$, and although there were more

recalls than intrusions in general, $F(1, 87) = 6.80$, $MSe = 22.71$, $p\eta^2 = .07$, this interacted with

condition, $F(1, 87) = 4.79$, $MSe = 22.71$, $p\eta^2 = .05$. Specifically, although there were more

intrusions in Experiment 3 than in the cued recall control, $t(87) = 4.47$, $d = 0.96$, there was no

difference in terms of correct recalls, $t(87) = 1.58$, $p = .12$, $d = 0.33$.

Discussion

The results of Experiment 3 suggest that providing multiple cues at test can increase the

likelihood of subjects producing the desired target. Unfortunately, these cues also seem to act to

increase the fluency of unstudied words, making subjects more likely to false alarm to these

items (cf. Whittlesea & Williams, 2001a, 2001b; Jacoby & Whitehouse, 1989; Yonelinas, 2002).

Worse yet, despite the fact that more studied items were produced in Experiment 3 compared to

both free recall and cued recall conditions, there was no significant increase in the number of

words recalled. Hence, the rise in intrusions and false alarms from providing multiple cues

comes with no benefit to overall levels of recall. In fact, in Experiment 3, subjects were

producing just as many recalls as intrusions, $t(27) = 0.20$, $p = .84$, $d = 0.04$.

It is worth noting that although there was no significant increase in the number of words recalled in Experiment 3 compared to the free recall and cued recall controls, numerically there were more recalls. One possibility is that there was insufficient power to demonstrate a difference between Experiment 3 and the controls. Although this is certainly possible, an issue yet to be considered is that of correct intrusions, which can artificially inflate recall scores. As the notion of correct intrusions is somewhat tangential to the current discussion, and because it requires the discussion of later experiments, it will be addressed in the General Discussion, not here. For now, suffice it to say that the trend toward more recalls in Experiment 3 compared to the free recall and cued recall controls may actually be an artifact of relatively high false alarm rates, indicative of a lax response criterion.

Experiment 4:

Increasing the Cue-Target Association

Providing multiple cues at test helped subjects produce more studied words but also acted to inflate false alarm rates. One reason for this inflation is that, although the multiple cues would help subjects to arrive at studied targets relatively easily, those same cues could also have helped subjects to arrive at unstudied targets occasionally. Given that processing fluency can affect subjective feelings of familiarity (cf. Whittlesea & Williams, 2001a, 2001b; Jacoby & Whitehouse, 1989; Yonelinas, 2002), unstudied targets produced from multiple cues may have felt more familiar, on average, than in the cued recall control. As a result, false alarms would have been expected to increase in Experiment 3.

The goal of Experiment 4 was again to increase the effectiveness of cues but this time without influencing the false alarm and intrusion rates. Consider that the results of Experiments 1 and 2 have demonstrated that when a word is produced in response to a cue, it appears to be produced more fluently than if no cue were provided. Experiment 3 further demonstrates that as more cues are used to cue a target, fluency is further enhanced. Hence, the number of cues provided appears to be the driving factor that influences the fluency with which targets are produced. To increase the effectiveness of cues without influencing false alarm and intrusion rates then, single cues were used in Experiment 4, rather than multiple cues as in Experiment 3. To make cues more effective at cueing their intended targets, cue-target association strength was increased in Experiment 4, compared to previous experiments. If cues affect the fluency of targets to the same degree as in Experiment 1 yet are more likely to cue to intended studied target than in Experiment 1, we should see a rise in the number of studied words produced in

Experiment 4 yet no difference in terms of recognition accuracy. The net result may be an increase in the number of words recalled with no similar increase in intrusions.

To be clear before continuing, however, cue-target association strength is merely an index of how reliably a cue leads to the free association of its target across individuals. For example, HORSE-SADDLE is a strongly related pair because, in response to HORSE, a majority of people would free associate SADDLE. HORSE-BEACH is not a strongly related pair because in response to HORSE, few people would free associate BEACH. The phrasing "association strength" may suggest that targets to strongly associated cues should be generated more fluently than those to weakly associated cues, but this is not so. Cue-target association strength only indicates how reliably cues lead to their intended targets, not the fluency with which those targets are generated.

As Experiment 3 demonstrated and as will be confirmed in the results of Experiment 4, it is the number of cues provided—not the cue-target association strength—that can affect the fluency with which cues lead to their targets. Thus, using single, stronger associates as cues in Experiment 4 should increase the effectiveness of cues in helping with the production of their intended targets, without affecting recognition accuracy and, consequently, without inflating false alarms or intrusions.

Method

Participants. Twenty-one students from the University of Waterloo participated in Experiment 4 in exchange for course credit toward a psychology course.

Materials. Experiment 4 used the same word pool as Experiment 1 but, to maximize the strength of cues in retrieving studied items, the pool of 156 words from the previous experiments was trimmed. The 79 cues with the highest probability of leading to their target word were used

in Experiment 4. The average probability that the new cues would give rise to their respective target words, based on the norming data, was .49 (SD = .20), significantly higher than the probability of .30 from Experiment 1, $t(233)$ = 6.42, d = 0.83.

Procedure. Experiment 4 was run identically to the Cued Recall condition of Experiment 1.

Results

The results of Experiment 4 can be seen in Table 1.

Compared to Free Recall. In Experiment 4, more studied items were produced than in the free recall control condition, $t(80)$ = 7.69, d = 1.72. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Experiment 4 vs free recall) mixed ANOVA. More hits were observed than false alarms, $F(1, 80)$ = 1026.16, MSe = 0.02, $p\eta^2$ = .93, and although there was no overall difference between Experiment 4 and the free recall control, $F(1, 80)$ = 2.96, MSe = 0.01, p = .09, $p\eta^2$ = .04, old/new status and condition interacted, $F(1, 80)$ = 27.28, MSe = 0.02, $p\eta^2$ = .25. Follow-up comparisons showed that hits were lower and false alarms were higher in Experiment 4 compared to the free recall control, $t(80)$ = 4.62, d = 1.03, and $t(80)$ = 3.11, d = 0.70, respectively.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 4 vs free recall) mixed ANOVA. Recall and intrusion rates were higher in Experiment 4 than in the free recall control, $F(1, 80)$ = 21.53, MSe = 12.38, $p\eta^2$ = .21, and although there were more recalls than intrusions in general, $F(1, 80)$ = 69.75, MSe = 18.93, $p\eta^2$ = .47, this interacted with condition, $F(1, 80)$ = 1.57, MSe = 18.93, p = .21, $p\eta^2$ = .02. Specifically, there were significantly more recalls in Experiment 4 than in the free recall control, $t(80)$ = 4.05, d = 0.90, and there was only a borderline increase in intrusions, $t(80)$ = 1.88, p = .06, d = 0.42.

Compared to Cued Recall. In Experiment 4, more studied items were produced than in the cued recall control, $t(80) = 5.81$, $d = 1.30$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Experiment 4 vs cued recall) mixed ANOVA. Although more hits were observed than false alarms, $F(1, 80) = 786.29$, $MSe = 0.02$, $p\eta^2 = .91$, there was no overall difference between Experiment 4 and the cued recall control and no interaction, both $F(1, 80) < 1$. Thus, recognition accuracy did not reliably differ in Experiment 4 compared to the cued recall control.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 4 vs cued recall) mixed ANOVA. Generally there were more recalls than intrusions, $F(1, 80) = 60.75$, $MSe = 16.03$, $p\eta^2 = .43$, and although there were more recalls and intrusions in Experiment 4 than in the cued recall control, $F(1, 80) = 7.75$, $MSe = 15.13$, $p\eta^2 = .09$, these two factors interacted, $F(1, 80) = 6.99$, $MSe = 17.76$, $p\eta^2 = .08$. Specifically, there were significantly more recalls in Experiment 4 than in the cued recall control, $t(80) = 4.11$, $d = 0.92$, and a similar number of intrusions, $t(80) = 0.04$, $p = .97$, $d = 0.01$.

Discussion

The strong associates used in Experiment 4 were successful in demonstrating a benefit of cued recall. Namely, the strong associates increased the number of studied items produced beyond that of the free recall and cued recall control conditions. Furthermore, although cueing once again impaired recognition rates compared to free recall, as evidenced by lower hit rates and higher false alarm rates, there was a significant rise in the number of items recalled. Hence, Experiment 4 demonstrates that cues can provide a mnemonic benefit.

However, cues did not provide a mnemonic benefit to recall in the absence of any cost. Most notably, cues both impaired hit rates and inflated false alarm rates compared to free recall.

In fact, even though there was only a borderline increase in intrusions in Experiment 4 compared to the free recall control, it should be noted that because so many more studied items were being produced in Experiment 4 than in free recall, the number of new items produced was significantly reduced, $t(80) = 7.91$, $d = 1.77$. As the number of new items produced decreases, intrusion rates must also decline. Hence, the more appropriate measure to consider between Experiment 4 and the free recall control is the false alarm rate, which certainly was higher in Experiment 4 than in the free recall control.

That cues inflate false alarm rates is not too surprising. As shown in Experiment 3, presenting multiple cues before each test trial inflates false alarm rates further than does presenting a single cue at test. Most likely, then, the inflated false alarm rates from cueing come from the fact that items produced from a cue are produced more fluently than without a cue. Because cues can act to increase the fluency of both old and new items, new items should be false alarmed to more in cued recall than in free recall. What is more perplexing about cueing however, is that it consistently impairs hit rates. Across all four experiments so far, cues have significantly impaired hit rates compared to free recall. The pertinent question is: Why?

Experiment 5:

Status-Identified Cueing

One explanation for the hit rate disadvantage due to cueing observed thus far is that cues to both studied and unstudied words have been presented at test. Subjects may realize that some cues are cueing them to studied words and others are not. Furthermore, when a cue for an unstudied word is presented, it may lead to an unstudied candidate being produced quite fluently (at least more so than for an unstudied word produced without a cue). If subjects were simply to accept these fluently produced words as actually having been studied, they would false alarm to many unstudied words. Thus, in an attempt to minimize false alarms, subjects may adopt a stricter recognition criterion, which acts to decrease their hit rates while controlling false alarms from becoming extreme.

Such a discounting-strategy account would imply that a major reason why cueing has been impairing hit rates has been the presentation of cues for both studied and unstudied words at test. If, instead, subjects were informed which cues were meant for studied words and which were not, or only cues for studied words were presented, subjects would not need to discount fluency as much, and hit rates should increase. Thus, in Experiment 5 the types of cues were identified to subjects item by item by presenting cues for studied word in blue and cues for unstudied words in white and informing subjects of this manipulation (status-identified cueing). In Experiment 6, only cues for studied words were presented to subjects at test. These two experiments serve to both test this discounting-strategy account and to ascertain what consequences, if any, arise from the practice of presenting cues for both studied and unstudied words at test.

Method

        Participants. Twenty students from the University of Waterloo participated in

Experiment 5 in exchange for course credit toward a psychology course.

        Materials. Experiment 5 used the same word pool as Experiment 1.

        Procedure. Experiment 5 was run identically to the Cued Recall condition of Experiment

1 except that cues for studied words were presented in blue and cues for unstudied words were

presented in white. Subjects were informed of this manipulation and told that the white cues

likely would not be useful memory aids for recalling studied words.

Results

        The results of Experiment 5 can be seen in Table 3. The overall results of Experiment 5

were first compared to the free recall and cued recall control conditions. Subsequently, a

detailed analysis of trials where studied words were cued (studied-cued) and trials where

unstudied words were cued (unstudied-cued) was performed.

        Compared to Free Recall. In Experiment 5, more studied items were produced than in

the free recall control condition, $t(79) = 2.49$, $d = 0.56$. Recognition accuracy was first analyzed

in a 2 (old vs new) X 2 (Experiment 5 vs free recall) mixed ANOVA. More hits were observed

than false alarms, $F(1, 79) = 947.68$, $MSe = 0.02$, $p\eta^2 = .92$, and although there was no overall

difference between Experiment 5 and the free recall control, $F(1, 79) < 1$, old/new status and

condition interacted, $F(1, 79) = 20.22$, $MSe = 0.02$, $p\eta^2 = .20$. Follow-up comparisons showed

that hits were lower and false alarms were higher in Experiment 5 compared to the free recall

control, $t(79) = 3.19$, $d = 0.72$, and $t(79) = 3.73$, $d = 0.84$, respectively.

        Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment

5 vs free recall) mixed ANOVA. Recall and intrusion rates were higher in Experiment 5 than in

Table 3. Mean number of words produced, hit and false alarm rates of produced words, number of words recalled, and number of intrusions in Experiment 5. Trials are further conditionalized on whether cues were for studied or unstudied targets. Standard errors are shown in parentheses below means.

Condition	Number of Productions			p("old")		Overall Performance	
	Old	New	Repeats	Old	New	Recalls	Intrusions
Cued Recall	11.05	33.65	3.10	0.82	0.18	9.35	6.15
Status-Identified Cues	(0.81)	(1.08)	(0.43)	(.03)	(.03)	(0.91)	(1.23)
Studied Word Cued Trials	9.30	12.70	1.95	0.83	0.24	7.85	3.10
	(0.56)	(0.77)	(0.34)	(.03)	(.04)	(0.62)	(0.58)
Unstudied Word Cued Trials	1.75	20.95	1.15	0.77	0.15	1.50	3.05
	(0.40)	(0.44)	(0.23)	(.10)	(.03)	(0.40)	(0.70)

the free recall control, $F(1, 79) = 12.22$, $MSe = 12.66$, $p\eta^2 = .13$, and there were more recalls than intrusions in general, $F(1, 79) = 24.64$, $MSe = 23.27$, $p\eta^2 = .24$. There was no interaction, $F(1, 79) = 1.75$, $MSe = 23.27$, $p = .19$, $p\eta^2 = .02$. Despite the nonsignificant interaction, comparisons between recall and intrusion rates have been a consistent aspect of all previous analyses and were planned, a priori, here as well. These follow-up tests confirmed that although there was a significant increase in intrusions in Experiment 5, $t(79) = 3.00$, $d = 0.67$, there was no significant increase in words recalled, $t(79) = 1.06$, $p = .29$, $d = 0.24$.

Compared to Cued Recall. In Experiment 5, no more studied items were produced than in the cued recall control condition, $t(79) = 0.55$, $p = .59$, $d = 0.12$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Experiment 5 vs cued recall) mixed ANOVA. More hits were observed than false alarms, $F(1, 79) = 727.59$, $MSe = 0.02$, $p\eta^2 = .90$, however there was no overall difference between Experiment 5 and the cued recall control and no interaction, $F(1, 79) = 2.37$, $MSe = 0.03$, $p = .13$, $p\eta^2 = .03$, and $F(1, 79) < 1$, respectively. Thus, overall recognition accuracy did not reliably differ in Experiment 5 compared to the cued recall control.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 5 vs cued recall) mixed ANOVA. Generally there were more recalls than intrusions, $F(1, 79) = 17.58$, $MSe = 20.33$, $p\eta^2 = .18$, however there was only a borderline increase in the number of recalls and intrusions in Experiment 5 compared to the cued recall control, $F(1, 79) = 3.21$, $MSe = 15.45$, $p\eta^2 = .04$, furthermore there was no interaction, $F(1, 79) < 1$. Once again, however, given the a priori plan to directly compared recall and intrusion rates, follow-up analyses were carried out. These analyses confirmed the ANOVA results, finding no difference in either the number of items recalled or intrusions between Experiment 5 and the cued recall control, $t(79) = 4.11$, $d = 0.92$, and $t(80) = 0.04$, $p = .97$, $d = 0.01$, respectively.

<u>Cues for Studied vs Unstudied Words.</u> The results of comparisons against free recall and cued recall suggest that Experiment 5 merely replicated the previous cued recall results. That is, status-identified cueing did not alter subjects' performances in recognized recall. However, before accepting this conclusion, a detailed analysis of the studied-cued and unstudied-cued trials was performed.

First, data from studied-cued and unstudied-cued trials were directly compared. More studied words were produced on studied-cued trials than on unstudied-cued trials, $t(19) = 14.10$, $d = 3.27$. Only 14 of the 20 subjects in Experiment 5 produced studied words on unstudied-cued trials. Because of this missing data, an overall ANOVA was not conducted on recognition accuracy scores. Instead, to maximize power, hit and false alarm rates were compared separately via paired sample t-tests. These analyses confirmed that although hit rates did not differ, $t(13) = 1.00$, $p = .33$, $d = 0.32$,[4] there were more false alarms on studied-cued than on unstudied-cued trials, $t(19) = 3.78$, $d = 0.39$.

Finally, hit and false alarm rates for studied-cued and unstudied-cued trials were compared back to the cued recall control condition. To maintain consistency with the above

---

[4] One concern that could be raised here is that the nonsignificant difference in hit rates was due to insufficient power. Given the small numerical difference observed for the hit rates, however, it was unclear that testing more subjects would be worthwhile. Indeed, power analyses indicated that data from 108 subjects would need to be gathered to have an 80% chance of finding a difference between the hit rates. However, given that only 70% of subjects provided useful data for both studied- and unstudied-cued trials, this translates to approximately 155 subjects being required to search for a hit rate difference here. Although power analyses indicated that the hit rate differences here were likely not statistically significant, or at the least not practically significant, I nonetheless tried two methods of correcting for these missing hit rate data points, in an attempt to ascertain whether there was any evidence that hit rates differed for studied-cued vs unstudied-cued trials. First, missing data points for unstudied-cued hit rates were completed with the mean for unstudied-cued hit rates based on the other subjects. The net result was that the new mean for unstudied-cued hit rates was .77 (*SE* = .07). Second, a linear regression model predicting unstudied-cued hit rates from unstudied-cued false alarm rates and studied-cued hit and false alarm rates was derived and, based on this model, predicted unstudied-cued hit rate scores were inserted into missing data points. The result of this adjustment was that the new mean for unstudied-cued hit rates was .81 (*SE* = .07). Although both of these corrections make assumptions that certainly could be questioned, neither changed the results. In both cases, there was no significant difference between hit rates for studied-cued vs unstudied-cued trials, $t(19) = 0.78$, $p = .44$, $d = 0.19$, and $t(19) = 0.23$, $p = .82$, $d = 0.05$, respectively. Thus, there is some assurance that the nonsignificant difference between studied-cued and unstudied-cued hit rates was not simply an artifact of incomplete data for some subjects.

analyses and due to the missing data points, paired sample *t*-tests were used in lieu of overall ANOVAs. Hit and false alarm rates in the unstudied-cued condition were identical to those in the cued recall control, $t(73) = 0.04$, $p = .97$, $d = 0.01$, and $t(79) = 0.48$, $p = .63$, $d = 0.11$, respectively. For the studied-cued results, however, although hit rates did not differ, $t(79) = 1.24$, $p = .22$, $d = 0.28$, false alarm rates were higher than in the cued recall control condition, $t(79) = 3.15$, $d = 0.71$.

Discussion

The goal of Experiment 5 was to ascertain whether the hit rate disadvantage for cued recall might have been due to the fact that cues for both studied and unstudied words were being provided at test. Experiment 5 found no evidence that hit rates were being affected differentially by cues to studied vs unstudied words. To the contrary, it seems as though identifying cues to subjects as cueing studied vs unstudied words may actually have inflated false alarm rates.

One reason why identifying the cues to subjects might inflate false alarm rates is that it introduces response pressure. Recall that one of the major motivations for using cues for both studied and unstudied words at test was to minimize response pressure. When subjects know that cues correspond to studied words, they may feel more obligated to identify the produced item as "old," regardless of actual memory. In contrast, when subjects do not know which cues correspond to studied words, there is less pressure to say "old" on any given trial, and subjects can respond based more on their subjective sense of memory. Hence, identifying cues as cueing studied vs unstudied words to subjects likely re-introduced response pressure, despite the fact that cues for both studied and unstudied words were still being presented.

Finally, as in Experiment 3, there was some hint that total number of recalls was greater in Experiment 5 than in the free recall control condition. Once again, although it seems as if this

may be the case and may warrant further attention, ultimately consideration of correct intrusions

in the General Discussion will cast doubt on whether the number of recalls was actually

increasing.

Experiment 6:

Cues for Studied Words Only

According to a discounting-strategy account of the hit rate disadvantage for cued recall, hit rates are lower in cued recall than in free recall because subjects know that they are being cued for both studied and unstudied items at test. Because they realize that a produced item could feel familiar due to the enhanced fluency afforded by cues, subjects adopt a stricter recognition criterion. This tempers their hit rates while preventing false alarms from becoming excessive. Experiment 5 found no evidence that status-identified cues reduced the hit rate disadvantage for cued recall; however, in Experiment 5 cues for both studied and unstudied words were still being presented. In Experiment 6, therefore, only cues for studied words were presented at test. Consequently, Experiment 6 serves as a second test of the discounting-strategy account of the hit rate disadvantage for cued recall, as well as specifically gauging the effect of presenting only cues for studied words at test. To this point, it has been argued that presenting cues for both studied and unstudied words at test provides a less biased testing environment. Experiment 6 directly investigates this issue.

Method

Participants. Twenty-six students from the University of Waterloo participated in Experiment 6 in exchange for course credit toward a psychology course.

Materials. Experiment 6 used the same word pool as Experiment 1.

Procedure. Experiment 6 was run identically to the Cued Recall condition of Experiment 1 except that instead of presenting cues for both studied and unstudied words at test, only cues for studied words were presented. Hence, because subjects studied 24 words, only 24 test trials were given in Experiment 6.

Results

The results of Experiment 6 can be seen in Table 4. Because there were only 24 test

trials in Experiment 6, there were necessarily fewer opportunities to produce new words in

Experiment 6 than in previous experiments. To make the free recall control condition data

comparable to those of Experiment 6, they were truncated to include only the first 24 test trial

responses (see Table 4). Subjects in the free recall condition were unaware of how many test

trials to expect, and so these first 24 trials should be unbiased and should produce perfectly the

data expected had only 24 test trials been given. Finally, because the previous cued recall

control condition contained more opportunities to produce new items than was the case in

Experiment 6, no comparison of intrusion rates was carried out between these two conditions,

although the number of studied items produced and the recognition rates can be readily

compared.

Compared to Free Recall. In Experiment 6, more studied items were produced than in

the truncated free recall control condition, $t(86) = 3.12$, $d = 0.67$. Recognition accuracy was first

analyzed in a 2 (old vs new) X 2 (Experiment 6 vs truncated free recall) mixed ANOVA. More

hits were observed than false alarms, $F(1, 86) = 827.55$, $MSe = 0.02$, $p\eta^2 = .91$, there was only a

borderline difference between Experiment 6 and the truncated free recall control, $F(1, 86) = 3.37$,

$MSe = 0.02$, $p = .07$, $p\eta^2 = .04$, and old/new status and condition interacted, $F(1, 86) = 25.56$,

$MSe = 0.02$, $p\eta^2 = .23$. Follow-up comparisons showed that hits were lower and false alarms

were higher in Experiment 6 compared to the truncated free recall control, $t(86) = 5.35$, $d = 1.15$,

and $t(86) = 2.63$, $d = 0.57$, respectively.

Table 4. Mean number of words produced, hit and false alarm rates of produced words, number of words recalled, and number of intrusions in the free recall conditions of Experiments 1 and 2 combined and truncated to 24 trials and Experiment 6. Standard errors are shown in parentheses below means.

Condition	Number of Productions			p("old")		Overall Performance	
	Old	New	Repeats	Old	New	Recalls	Intrusions
Free Recall	8.50	14.53	0.73	0.96	0.13	8.26	1.82
Truncated to 24 Trials	(0.51)	(0.53)	(0.17)	(.01)	(.02)	(0.52)	(0.28)
Cued Recall	11.27	11.73	0.96	0.80	0.22	9.27	2.65
Studied Word Cues Only	(0.64)	(0.72)	(0.23)	(.03)	(.03)	(0.77)	(0.38)

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 6 vs truncated free recall) mixed ANOVA. Recall and intrusion rates were higher in Experiment 6 than in the truncated free recall control, $F(1, 86) = 4.16$, $MSe = 7.48$, $p\eta^2 = .05$, and there were more recalls than intrusions in general, $F(1, 86) = 114.21$, $MSe = 13.66$, $p\eta^2 = .57$. There was no reliable interaction, $F(1, 86) < 1$. Despite the nonsignificant interaction, follow-up a priori tests confirmed that although there was no significant rise in recalls, $t(86) = 1.06$, $p = .29$, $d = 0.23$, there was a borderline increase in intrusions in Experiment 6 compared to the truncated free recall control, $t(86) = 1.67$, $p = .09$, $d = 0.36$. Although only a borderline result, given that so many fewer new items were produced in Experiment 6 than in the truncated free recall control, $t(86) = 2.96$, $d = 0.64$, it is relatively remarkable that any trend in intrusions was noted. That is, false alarm rates were so high in Experiment 6 that they overpowered the significant decrease in new items produced to at least balance the level of intrusions in Experiment 6 and in the truncated free recall control.

Compared to Cued Recall. In Experiment 6, no more studied items were produced than in the cued recall control condition, $t(86) = 0.89$, $p = .38$, $d = 0.19$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Experiment 6 vs cued recall) mixed ANOVA. More hits were observed than false alarms, $F(1, 86) = 679.86$, $MSe = 0.02$, $p\eta^2 = .89$, and there was a borderline difference between Experiment 6 and the cued recall control, $F(1, 86) = 3.86$, $MSe = 0.03$, $p = .05$, $p\eta^2 = .04$, but no interaction, $F(1, 86) = 1.89$, $MSe = 0.02$, $p = .17$, $p\eta^2 = .02$. This borderline effect was actually driven by the fact that false alarm rates were higher in Experiment 6 than in the cued recall control, $t(86) = 3.03$, $d = 0.65$, although there was no difference in hit rates, $t(86) = 0.49$, $p = .63$, $d = 0.11$. Finally, there was no difference in terms of overall recall between Experiment 6 and the cued recall control, $t(86) = 1.06$, $p = .30$, $d = 0.23$.

<u>Discussion</u>

Overall, the results of Experiment 6 align with those of Experiment 5. Providing cues only for studied words in Experiment 6 did not correct the hit rate disadvantage for cued recall. In fact, the results from Experiment 6 look remarkably similar to those of the cued recall control except that false alarms noticeably increased. This false alarm rate increase was also observed on trials were cues for studied words were presented in Experiment 5, and suggests that cued recall is more susceptible to false memories and intrusions when only cues to studied words are presented or when subjects know which cues are meant for studied items. Once again, this is most likely due to the increased response pressure placed on subjects on those trials. When subjects know that a cue is meant for a studied word, there is more of an onus to respond "old," and this acts to drive up false alarm rates. Although this explanation would also reasonably expect hit rates to increase as well, there was in fact a small increase in hit rates in both Experiments 5 and 6 compared to the cued recall control. This hit rate difference was not significant in either case, however, even if it were it is not inconsistent with the response pressure explanation.

Yet, regardless of why false alarm rates increased in Experiments 5 and 6, the clear conclusion is that using cues only for studied words at test indeed biases cued recall performance. Hence, the recommendation that cues for both studied and unstudied words should be used regularly in cued recall paradigms finds empirical support in these two experiments. Finally, there was once again a small numerical increase in the number of recalls in Experiment 6 compared to the cued recall control; once again, detailed consideration of this effect will be deferred until discussion of the issue of correct intrusions in the General Discussion.

63

Experiment 7:

Delayed Recognition

Two attempts to investigate the discounting-strategy account have failed to find any evidence that the hit rate disadvantage for cued recall is due to subjects discounting their responses: Both when cues are identified as cueing studied or unstudied words and when only cues to studied words are presented, there was no evidence that the hit rate impairment in cued recall was reduced. Experiment 7 seeks to investigate the discounting-strategy account using a different methodology. If subjects are discounting studied targets because there are too many fluent unstudied targets being produced, then delaying the recognition test may eliminate this fluency differential. That is, instead of having subjects decide "old"/"new" immediately after producing each item, recognition will be delayed in Experiment 7, such that potential fluency effects from the cues may have time to dissipate. Thus, subjects will first do a phase in which they produce studied and unstudied words and subsequently will do a recognition test phase for those words that they produced in the first phase. Delaying the recognition test may thus act to improve hit rates in the cued recall condition, insofar as any enhanced sense of familiarity afforded by cues may have dissipated by the time the recognition test occurs, and therefore there would be less motivation to discount.

In addition to testing the discounting-strategy account, Experiment 7 provides an opportunity to gauge the impact, if any, of requiring immediate recognition vs delayed recognition. Hence, both a Free Recall and Cued Recall condition were run in Experiment 7, to properly assess this factor.

Method

Participants. Sixty-one students from the University of Waterloo participated in Experiment 7 in exchange for course credit toward a psychology course. In the Free Recall condition, 35 participated; in the Cued Recall condition, 26 participated.

Materials. Experiment 7 used the same word pool as Experiment 1.

Procedure. Experiment 7 was run identically to Experiment 1 except that instead of being given a recognition test trial immediately after producing each word, the recognition trials were saved for a delayed test. After producing 48 words, subjects were given instructions informing them that there would be one more test: a recognition test. Recognition was explained identically to Experiment 1 and they then proceeded to see the 48 words that they had produced, in a new random order, and decided "old" or "new" for each word.

Results

The results of Experiment 7 can be seen in Table 5. The primary question of interest was whether the Free Recall and Cued Recall conditions of Experiment 7 would replicate the free recall and cued recall control conditions. Hence, the Free Recall condition was compared to the free recall control and the Cued Recall condition was compared to the cued recall control. The Free Recall and Cued Recall conditions of Experiment 7 were then directly compared.

Free Recall. In the Free Recall condition of Experiment 7, the same number of words were produced as in the free recall control, $t(94) = 0.11$, $d = 0.02$. Recognition accuracy was analyzed in a 2 (old vs new) X 2 (Experiment 7 Free Recall vs free recall) mixed ANOVA. Although more hits were observed than false alarms, $F(1, 94) = 2564.67$, $MSe = 0.01$, $p\eta^2 = .97$, there was no difference between conditions and no interaction, both $F(1, 94) < 1$. Similarly, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 7 Free Recall vs

Table 5. Mean number of words produced, hit and false alarm rates of produced words, number of words recalled, and number of intrusions in free recall and cued recall in Experiment 7. Standard errors are shown in parentheses below means.

Condition	Number of Productions			p("old")		Overall Performance	
	Old	New	Repeats	Old	New	Recalls	Intrusions
Free Recall	8.57	37.77	3.31	0.94	0.06	8.09	2.26
Delayed Recognition	(0.62)	(0.77)	(0.69)	(.02)	(.01)	(0.63)	(0.42)
Cued Recall	11.00	33.69	3.92	0.79	0.17	8.73	5.81
Delayed Recognition	(0.65)	(0.70)	(0.85)	(.03)	(.03)	(0.62)	(0.93)

free recall) mixed ANOVA. And similarly, although there were more recalls than intrusions, $F(1, 94) = 84.46$, $MSe = 16.97$, $p\eta^2 = .47$, there was no difference between conditions and no interaction, both $F(1, 94) < 1$. Thus, the results of the Free Recall condition in Experiment 7 replicate those of the free recall control: No differences were observed.

Cued Recall. In the Cued Recall condition of Experiment 7, the same number of words were produced as in the cued recall control, $t(94) = 0.55$, $d = 0.11$. Recognition accuracy was analyzed in a 2 (old vs new) X 2 (Experiment 7 Cued Recall vs cued recall) mixed ANOVA. Although more hits were observed than false alarms, $F(1, 94) = 851.24$, $MSe = 0.02$, $p\eta^2 = .91$, there was no difference between conditions and no interaction, $F(1, 94) = 1$, and $F(1, 94) < 1$, respectively. Similarly, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 7 Cued Recall vs cued recall) mixed ANOVA. And similarly, although there were more recalls than intrusions, $F(1, 94) = 20.84$, $MSe = 19.12$, $p\eta^2 = .20$, there was no difference between conditions and no interaction, $F(1, 94) = 1.78$, $MSe = 13.21$, $p = .19$, $p\eta^2 = .02$, and $F(1, 94) < 1$, respectively. Thus, the results of the Cued Recall condition in Experiment 7 replicate those of the cued recall control: No differences were observed.

Free vs Cued Recall. Given that neither the Free Recall nor the Cued Recall condition of Experiment 7 differed from the free recall and cued recall controls, the results of Experiment 1 were expected to replicate here. Indeed, the Cued Recall condition in Experiment 7 led to more studied items being produced than did the Free Recall condition, $t(59) = 2.66$, $d = 0.69$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Free Recall vs Cued Recall) mixed ANOVA. More hits were observed than false alarms, $F(1, 59) = 1487.51$, $MSe = 0.01$, $p\eta^2 = .96$, and although there was no overall difference between Free Recall and Cued Recall in Experiment 7, $F(1, 59) < 1$, there was a significant interaction, $F(1, 59) = 44.69$, $MSe = 0.01$, $p\eta^2$

= .43. Follow-up analyses indicated that there were both fewer hits and more false alarms in the

Cued Recall condition in Experiment 7 compared to the Free Recall condition, $t(59) = 4.53$, $d =$

1.18, and $t(59) = 4.05$, $d = 1.05$, respectively.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Free Recall

vs Cued Recall) mixed ANOVA. More recalls than intrusions were observed in general, $F(1, 59)$

= 36.98, $MSe = 15.45$, $p\eta^2 = .39$, and although more responses were observed in Cued Recall

than in Free Recall, $F(1, 59) = 13.24$, $MSe = 9.92$, $p\eta^2 = .18$, these two factors interacted, $F(1,$

59) = 4.08, $MSe = 15.45$, $p\eta^2 = .07$. Namely, although there were more intrusions in the Cued

Recall condition than in the Free Recall condition, $t(59) = 3.79$, $d = 0.99$, there was no difference

in terms of items recalled, $t(59) = 0.71$, $p = .48$, $d = 0.18$. Thus, Experiment 7 replicates the

results of Experiment 1.

Discussion

The results of Experiment 7 completely replicated those of Experiment 1. First, the Free

Recall and Cued Recall conditions here did not differ from the free recall and cued recall control

conditions. In particular, cues were noted to increase the number of studied items produced but

to impair recognition. Both a hit rate decline and false alarm rate inflation were observed. The

net result was an increase in intrusions with no corresponding increase in the number of items

recalled.

The goal of Experiment 7 was to provide an alternative test of the discounting-strategy

hypothesis of the hit rate disadvantage for cued recall. No evidence was found that delaying the

recognition component of recognized recall affected performance whatsoever. Hence, no

evidence was obtained that attempting to separate the fluency of production from the recognition

decision improves cued recall performance.

Experiment 8:

Eliminating Production Pressure

Experiment 8 is a final attempt to ascertain whether the hit rate disadvantage in cued recall could be due to some kind of discounting strategy. If subjects are discounting some of their produced studied items due to the fact that they are producing both studied and unstudied items at test, then eliminating the pressure to produce new items should rectify this situation.[5] That is, if subjects are asked to produce only studied items, a strong strategy-discounting account would suggest that the number of new items produced should drop to zero, but the number of studied items produced should remain the same. As a result, there would be no need for subjects to discount hit rates anymore and hit rates should increase. Thus, in Experiment 8, the instructions to free associate new items if a studied item could not be retrieved were removed.

Experiment 8 therefore serves as a final test of the discounting-strategy account while also assessing the impact in previous experiments of forcing subjects to produce both studied and unstudied items at test. It is possible that recognized recall has mischaracterized the benefit of cueing up to this point due to the pressure to produce both studied and unstudied items at test. For example, Roediger (1993) found that subjects who are forced to guess during production judged more new items to have been studied than did subjects who were not forced to guess. In essence, it is possible that the instructions to guess in recognized recall may have inflated false alarm rates and hence, intrusions, in previous cued recall experiments in this dissertation (although see Jacoby & Hollingshead, 1990, for just the opposite finding). By removing the pressure to guess during production, Experiment 8 will test this hypothesis.

---

[5] Thanks to Mike Masson for suggesting this experiment.

Method

Participants. Thirty-one students from the University of Waterloo participated in Experiment 8 in exchange for course credit toward a psychology course.

Materials. Experiment 8 used the same word pool as Experiment 1.

Procedure. Experiment 8 was run identically to the Cued Recall condition in Experiment 1 except that subjects were instructed to produce only studied items. If a studied item did not come to mind, subjects were told to press ENTER to pass. Furthermore, to maintain consistency with past studies and allow calculation of hit and false alarm rates, subjects were still asked to decide "old" and "new" after producing a response. Subjects were told that the recognition decision was there "just in case" they changed their mind, although once again, they were told never to intentionally produce a new item and hence, were instructed that it was perfectly acceptable to never indicate "new" on any recognition decision during the entire test.

Results

The results of Experiment 8 can be seen in Table 1. False alarm rates here were calculated as the number of intrusions divided by the number of test trials where no studied item, repetition, nor non-response was produced. Note that this superficially differs from past experiments, where false alarms were intrusions divided by the number of new items produced, but is functionally equivalent. Because subjects were encouraged not to produce new items in this experiment, if false alarms were calculated as the number of intrusions divided by the number of new items produced, the resulting values would have been nonsensical, yielding false alarm rates near 1.0. Such false alarms would not have been comparable to any of the previous experiments.

Compared to Free Recall. In Experiment 8, no more studied items were produced than in the free recall control, $t(90) = 0.26$, $d = 0.05$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Experiment 8 vs free recall) mixed ANOVA. More hits were observed than false alarms, $F(1, 90) = 1950.27$, $MSe = 0.02$, $p\eta^2 = .96$; furthermore, both more hits and more false alarms were observed in Experiment 8 than in the free recall control, $F(1, 90) = 10.31$, $MSe = 0.01$, $p\eta^2 = .10$. There was no interaction, $F(1, 90) < 1$.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 8 vs free recall) mixed ANOVA. There were more recalls than intrusions in general, $F(1, 90) = 42.45$, $MSe = 23.02$, $p\eta^2 = .32$, and although there was a main effect of condition, $F(1, 90) = 3.96$, $MSe = 15.19$, $p\eta^2 = .04$, there was no interaction, $F(1, 90) < 1$. Despite this non-significant interaction, a priori comparisons between Experiment 8 and the free recall control showed a borderline increase in intrusions in Experiment 8, $t(90) = 1.78$, $p = .08$, $d = 0.38$, despite no difference in terms of recalls, $t(90) = 0.65$, $p = .52$, $d = 0.14$.

Compared to Cued Recall. In Experiment 8, fewer studied items were produced than the cued recall control, $t(90) = 2.09$, $d = 0.44$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Experiment 8 vs cued recall) mixed ANOVA. Although more hits were observed than false alarms, $F(1, 90) = 1564.92$, $MSe = 0.02$, $p\eta^2 = .95$, and there was an overall difference between Experiment 8 and the cued recall control, $F(1, 90) = 17.71$, $MSe = 0.02$, $p\eta^2 = .16$, a significant interaction was observed, $F(1, 90) = 35.31$, $MSe = 0.02$, $p\eta^2 = .28$. Follow-up analyses indicated that although hit rates were higher in Experiment 8 than in the cued recall control, $t(90) = 6.47$, $d = 1.37$, false alarm rates did not differ, $t(90) = 0.52$, $p = .61$, $d = 0.11$.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 8 vs cued recall) mixed ANOVA. Although there were more recalls than intrusions, $F(1, 90) =$

31.50, $MSe = 20.44$, $p\eta^2 = .26$, there was no effect between conditions and no interaction, both $F(1, 90) < 1$. Confirming these results, a priori follow-up comparisons revealed no difference between recalls or intrusions in Experiment 8 compared to the cued recall control condition, $t(90) = 0.59$, $p = .56$, $d = 0.12$, and $t(90) = 0.04$, $p = .97$, $d = 0.01$, respectively.

Discussion

Experiment 8 demonstrates that when the response pressure during production is eliminated, cueing actually appears less useful. Indeed, hit rates did increase in Experiment 8 (even beyond that of the free recall control condition) however, unlike previous cued recall experiments, subjects were not producing more studied items in Experiment 8 than in free recall. Hence, even though there was no hit rate disadvantage in Experiment 8, because there was no increase in the number of studied items produced over free recall, there was similarly no increase in recalls compared to either the free recall or cued recall control conditions.

Worse yet, intrusion rates may have been increasing. A borderline increase in intrusions was noted compared to the free recall control condition, and no significant decrease in intrusions was noted compared to the cued recall control condition. It is important to remember that the number of observed intrusions is a function partly of the intrusion rate (i.e., false alarm rate) and partly of the number of new items produced. Hence, the number of intrusions observed can sometimes be ambiguous between conditions. To more carefully compare between experiments, false alarms are the more appropriate measure, as these provide the rate of false memories, and are not dependent on the number of new items produced. Indeed, false alarm rates were higher in Experiment 8 than in the free recall control, and were equivalent to those in the cued recall control condition. Thus, in Experiment 8, cues provided no memory benefit and simply increased false memories.

One potential criticism of Experiment 8 could be that, had there been stronger cues at test, a recall advantage for cueing might have been found. That is, Experiment 4 did demonstrate a recall advantage over free recall using strong semantic associates of studied words as cues. Examining Experiments 4 and 8 together, the expectation might be that, had strong cues been used in Experiment 8, more studied items would have been produced than was actually observed. And if more studied items were being produced, the number of words recalled should have similarly increased. Hence, a recall advantage might have been seen in Experiment 8 compared to the control conditions.

Although this would be the expected outcome, it is noteworthy that, regardless of whether an absolute cued recall advantage is observed, intrusions did rise in Experiment 8 compared to free recall. Furthermore, eliminating response pressure acted to increase hit rates precisely because subjects stopped producing words that they could not recognize. Neither of these facts would be likely to change had stronger semantic associate cues been used in Experiment 8 and, hence, using stronger cues here might have changed the absolute number of words recalled yet altered none of the critical conclusions.

A second issue worth noting is that past studies of cued recall have not always used instructions that encourage subjects to respond on every trial. Hence, past work often finds a cued recall advantage even when no onus exists to produce items by free associating. Consequently, because Experiment 8 did not find this effect, it could be argued that Experiment 8 was flawed in some way. And if Experiment 8 is flawed, it cannot be taken as a good measure of cueing in the absence of guessing.

Once again, using stronger associate cues it may be possible to show a recall advantage for cues even when there is no response pressure. However, it is also important to point out that

the recognized recall paradigm itself was designed to minimize or measure guessing as much as possible. Showing cues for both studied and unstudied words was meant to minimize guessing, as was forcing subjects to recognize their own responses as "old" or "new." Thus, past cued recall studies that did not force subjects to respond on every trial, but that lacked the controls present in recognized recall, *did* encourage guessing. That all reasons to guess were eliminated in Experiment 8 but past cued recall findings were not replicated is not a problem. Instead, it merely highlights the present argument that when guessing is adequately controlled in cued recall paradigms, many of the advantages of cued recall disappear.

The primary purpose of Experiment 8 was to investigate the discounting-strategy account one last time. There was, once again, no support for this account. Instead, the results of Experiment 8 demonstrate that the hit rate disadvantage observed in cued recall thus far may in fact be due to cues helping subjects to produce items that they cannot subsequently recognize. These extra productions appear to arise via free association: When the requirement to free associate new items is removed (as in Experiment 8), so too are these extra productions. Thus, the reason that hit rates decline in cued recall is that cues lead subjects to produce more studied items than they can recognize. When those extra studied items are removed, cued recall hit rates are equivalent to those of free recall.

Related Paired-Associates

If cues can help subjects free associate studied words, why do subjects not recognize those words? Indeed, early theories of cueing suggested that cues act to help memory by directing subjects toward the location of targets in memory (Tulving & Pearlstone, 1966). Nearly all of the experiments thus far have demonstrated that cues do help subjects to produce studied words. Yet, if subjects are able to produce more studied words than in free recall, why are these extra studied words often unrecognizable?

The belief that just because a word can be produced, it should be recognizable stems from the implicit assumption of transsituational identity (cf. Martin, 1975). As described in the introduction to this dissertation, transsituational identity is the notion that a word has only one, or at least one primary, representation. In a transsituational identity account, a ROSE is always a ROSE. The word ROSE can have different meanings and definitions, but it has only one representation in the mind to encompass its entirety. The single representation of a word is accessed every time that word is perceived, thought of, recalled, or recognized.

In contrast to the transsituational identity view, many researchers have pointed out that a single word can have multiple distinctive meanings (Light & Carter-Sobell, 1970; Martin, 1975; Katz & Fodor, 1963). Thus, SEAL can refer to either a marine carnivore or an embossed emblem. Hence, the word SEAL encountered shortly after thinking about a zoo will carry quite a different meaning than were it encountered while daydreaming about official office documents (as I am sure we all frequently do). Moreover, this fact extends beyond words with multiple dictionary definitions. Even words with a single dictionary definition have multiple senses (Anisfield & Knapp, 1968; Atkinson & Shiffrin, 1968; Buschke & Lenon, 1969; Brown &

McNeill, 1966; Clark & Stafford, 1969; Collins & Quillan, 1969; Reder et al., 1974). For

example, the word WATER, despite having only one real dictionary definition, has many senses.

WATER is viewed differently when thought about in the context of a lake, a waterfall, a shower,

sweat, condensation, etc.

To formalize the distinction here, Weiss (1967) differentiated between the nominal and

the functional status of a stimulus. The nominal stimulus is that which the experimenter

manipulates. For example, ROSE is a nominal stimulus and technically only contains four

objective letters and nothing else. The functional stimulus on the other hand is the way in which

a subject perceives and treats the nominal stimulus, which can vary greatly. ROSE may bring to

mind images of flowers and gardens, friends who are named Rose, the idea that something has

risen, the word ROWS, etc. In defining these two terms, Weiss pointed out that the nominal

stimulus and the functional stimulus are not always the same. That is, the stimulus ROSE may

be meant, by the experimenter, to bring to mind flowers, and hence FLOWERS would be the

subsequent cue chosen for ROSE. However, subjects could very easily interpret ROSE in a

completely different sense (e.g., as in thinking of the past tense of RISE). Thus, the nominal

word ROSE is giving rise to functionally distinct senses for the experimenter and the subject

respectively. In essence, the subject is not experiencing the ROSE that the experimenter

intended.

A major contribution to our understanding of memory was made in the 1970's by

researchers who found strong support for this nominal-functional distinction. Namely, when a

word is encoded, it is encoded in a specific functional sense. Subsequently, cues that cue that

specific functional sense are effective at retrieving that memory trace (Light & Carter-Sobell,

1970; Thomson & Tulving, 1970; Watkins & Tulving, 1975). Cues that cue another mental

context, even if they cue the same nominal word through semantic association, are not effective at retrieving the specific memory trace that was encoded (Light & Carter-Sobell, 1970; Thomson & Tulving, 1970; Winograd & Conn, 1971). Hence, studied words that are produced based on cues that cue a different functional sense of a studied word will not be recognized because in essence they actually have not been retrieved at all. As has been stated previously, this principle was dubbed encoding specificity (Thomson & Tulving, 1970, 1971; Watkins & Tulving, 1975).

In an encoding specificity framework, semantic associate cues would be expected to frequently lead subjects to produce studied words that cannot be recognized, because there is no guarantee that the functional sense of the target word implied by a cue is the same as that experienced during the study phase. And in fact, this hit rate disadvantage has replicated again and again in the cued recall experiments in this dissertation. Even in cases where semantic associate cues were a useful mnemonic (as in Experiment 4), subjects were still producing more studied items than could be recognized. Thus, how can the hit rate disadvantage for cued recall be eliminated? The answer from the encoding specificity framework is that, if studied items can be forced to be interpreted in the functional sense that the semantic associate cue will later cue, the hit rate disadvantage should be eliminated.

To eliminate the hit rate disadvantage for cued recall, then, Experiment 9 attempted to cue subjects not just to the nominal stimuli, but to the subjective stimuli that they had experienced at study. One way of accomplishing this is to force the encoding of each stimulus at study into the context of the eventual retrieval cue. Keeping with the theme of using semantic associates as cues, in Experiment 9, subjects will be required to encode pairs of semantically related words at study, and at test the first item of that pair will be used as the cue for the second. In a practical sense then, Experiment 9 is virtually identical to Experiment 1, except that cues are

present both at study and at test, rather than just at test. The main theoretical difference is that

the presence of cues at study should encourage subjects to encode targets in the same functional

sense as they will be cued at test and, hence, an encoding specificity account would predict that

the hit rate disadvantage for cued recall should be eliminated in Experiment 9.

Method

Participants. Sixty-nine students from the University of Waterloo participated in

Experiment 9 in exchange for course credit toward a psychology course. In the Free Recall

condition, 34 participated; in the Cued Recall condition, 35 participated.

Materials. Experiment 9 used the same word pool as Experiment 1.

Procedure. Experiment 9 was run identically to Experiment 1 except that subjects studied

cue-target pairs, rather than just targets. Cue-target pairs were determined in the same manner as

in Experiment 1. Furthermore, during the study phase, subjects were informed only to remember

both words from each pair, and that the two words appeared together. Due to the increased

difficulty in remembering pairs vs single items (based on pilot testing), cue-target pairs were

shown for 6.5 s (rather than the 2 s exposure used in previous experiments) and only 18 cue-

target pairs were shown at study (rather than the 24 targets used in previous experiments). As a

result of the shortened study list, only 36 test trials were given (rather than 48 test trials as in

previous experiments).

Test instructions in the Cued Recall condition were modified to inform subjects that the

first word from each word pair would be shown at test to cue them to the second word.

Furthermore, subjects were informed that they would see cues not present during the study

phase. Finally, Free Recall instructions asked subjects to recall only the second word from each

studied word pair. Beyond these changes, test instructions were identical to Experiment 1.

<u>Results</u>

The results of Experiment 9 can be seen in Table 6. The number of cues accidentally produced during test are reported here, however these were not analyzed.[6] Due to the methodological differences between Experiment 9 and the free recall and cued recall controls, no direct overall comparisons against the controls were conducted, although hit rates are contrasted where appropriate.

More studied words were produced in the Cued Recall condition than in the Free Recall condition, $t(67) = 13.19$, $d = 3.22$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Free Recall vs Cued Recall) mixed ANOVA. More hits were observed than false alarms, $F(1, 67) = 8016.57$, $MSe = 0.004$, $p\eta^2 = .99$, and although there was no overall difference between Free Recall and Cued Recall, $F(1, 67) = 1.27$, $MSe = 0.003$, $p = .26$, $p\eta^2 = .02$, there was a significant interaction, $F(1, 67) = 16.83$, $MSe = 0.004$, $p\eta^2 = .20$. Follow-up analyses indicated that there were both fewer hits and more false alarms in the Cued Recall condition in Experiment 9 compared to the Free Recall condition, $t(67) = 2.13$, $d = 0.52$, and $t(67) = 4.00$, $d = 0.98$, respectively. Despite the fact that hit rates were significantly lower in the Cued Recall condition than in the Free Recall condition of Experiment 9, it should be noted that hit rates were exceptionally high. That is, hit rates in the Cued Recall condition of Experiment 9 were higher than the hit rates observed in both the free recall and cued recall control conditions, $t(93) = 2.06$, $d = 0.43$, and $t(67) = 6.50$, $d = 1.35$, respectively.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Free Recall vs Cued Recall) mixed ANOVA. More recalls than intrusions were observed in general,

---

[6] No explicit predictions were made regarding production and/or recognition of these items, and subjects were explicitly asked not to produce cues. Furthermore, direct comparison of cue production rates between the Free Recall and Cued Recall conditions would likely be biased due to the fact that cues are exposed twice in Cued Recall (once at study, and once during the test) but only once in Free Recall.

Table 6. Mean number of words produced, hit and false alarm rates of produced words, number of words recalled, and number of intrusions in Experiments 9 through 11. Standard errors are shown in parentheses below means.

Condition	Number of Productions				p("old")		Overall Performance	
	Old	New	Cue	Repeats	Old	New	Recalls	Intrusions
Experiment 9								
Free Recall	5.76	28.38	0.97	0.47	0.98	0.03	5.65	0.71
Related Pairs	(0.41)	(0.53)	(0.21)	(0.15)	(.01)	(.01)	(0.41)	(0.15)
Cued Recall	14.26	21.54	0.06	0.09	0.95	0.08	13.60	1.74
Related Pairs	(0.49)	(0.49)	(0.04)	(0.05)	(.01)	(.01)	(0.53)	(0.27)
Experiment 10								
Free Recall	5.35	27.00	2.44	0.53	0.94	0.04	5.03	1.15
Unrelated Pairs	(0.39)	(0.65)	(0.40)	(0.17)	(.02)	(.01)	(0.39)	(0.23)
Cued Recall	7.85	26.32	0.88	0.91	0.97	0.12	7.68	3.35
Unrelated Pairs	(0.62)	(0.62)	(0.23)	(0.18)	(.01)	(.02)	(0.63)	(0.50)
Experiment 11								
Cued Recall	7.85	25.00	1.56	1.37	0.82	0.09	6.33	2.22
Weak-Cue Study/	(0.52)	(0.62)	(0.33)	(0.34)	(.04)	(.02)	(0.46)	(0.49)
Strong-Cue Test								

$F(1, 67) = 420.10$, $MSe = 5.79$, $p\eta^2 = .86$, and although more responses were observed in Cued

Recall than in Free Recall, $F(1, 67) = 194.70$, $MSe = 3.58$, $p\eta^2 = .74$, these two factors interacted,

$F(1, 67) = 71.21$, $MSe = 5.79$, $p\eta^2 = .52$. Follow-up analyses found that cues increased both

recalls and intrusions, $t(67) = 11.90$, $d = 2.91$, and $t(67) = 3.34$, $d = 0.82$, respectively. This

interaction actually indicated that cues led to larger increase in recalls than in intrusions.

Although this finding is favorable for cueing, it should be remembered that intrusion rates were

necessarily limited by the fact that fewer new items were produced in cued recall than free recall,

$t(67) = 9.50$, $d = 2.32$. Hence, the more appropriate indicator of false memories here is false

alarm rate, and this indeed was higher in the Cued Recall than in the Free Recall condition.

Discussion

The results of Experiment 9 support the encoding specificity account. Namely, cues in

Experiment 9 were found to aid in the production of studied items, as in all previous

experiments, yet no large hit rate disadvantage was observed here. In Experiment 9, subjects

encoded all of the studied words in the presence of the cues. This should have encouraged

subjects to view and encode those studied words in the specific functional senses that were

compatible with the cues. When the cues subsequently helped subjects to produce those studied

items, those items would have had the same functional sense as when studied and, therefore,

subjects were able to recognize those words as studied.

It can be noted that, technically, there was a hit rate disadvantage for cues. That is, hit

rates were significantly lower in the Cued Recall condition of Experiment 9 than in the Free

Recall condition. This difference is significant but trivial. The size of the hit rate difference was

much smaller than has been observed in the past experiments in this dissertation and,

furthermore, hit rates in the Cued Recall condition of Experiment 9 were higher than in either the

free recall or cued recall control conditions. Thus, the claim here is that, in Experiment 9, the hit rate disadvantage for cues was significantly reduced, if not functionally eliminated.

Finally, it is worthwhile to point out that although Experiment 9 demonstrated a clear mnemonic benefit of cues, this benefit was not cost free. That is, in Experiment 9, both false alarms and intrusions were significantly higher in the Cued Recall condition than in the Free Recall condition. Hence, even though cues were helpful in terms of increasing the number of items recalled, they also increased the number of false memories, a result that has occurred repeatedly in this dissertation.

Unrelated Paired-Associates

As noted, the results of Experiment 9 were in line with an encoding specificity account of cueing. As a second test of this account, Experiment 10 used unrelated word pairs as cue-target pairings at study. If the encoding specificity account is correct, then as long as targets are encoded in reference to their eventual retrieval cues, cueing will be effective at helping subjects produce studied items at test, while leaving hit rates unaffected. Therefore, the fact that unrelated words are used as retrieval cues here should not matter, given that they were present at study. Past work by Tulving and colleagues suggests that weakly related cues can act to cue memory in this way (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973), so Experiment 10 should qualitatively replicate Experiment 9.

Method

Participants. Sixty-eight students from the University of Waterloo participated in exchange for course credit toward a psychology course, 34 in the Free Recall condition and 34 in the Cued Recall condition.

Materials. Experiment 10 used the same word pool as Experiment 1. This time, however, cue-target pairings were created by randomly pairing two targets from the word pool together. No words identified as cues in the word pool were used in Experiment 10.

Procedure. Experiment 10 was run identically to Experiment 9 except for the modified stimuli just described.

Results

The results of Experiment 10 can be seen in Table 6. As in Experiment 9, the number of cues produced during test are reported here, however they were not analyzed. Also as in

Experiment 9, due to the methodological differences between Experiment 10 and the free recall and cued recall control conditions, no direct overall comparisons against the controls were conducted. After contrasting Free Recall and Cued Recall within Experiment 10, these conditions were compared to Experiment 9.

Free vs Cued Recall within Experiment 10. More studied words were produced in the Cued Recall condition than in the Free Recall condition, $t(66) = 3.41$, $d = 0.84$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Free Recall vs Cued Recall) mixed ANOVA. More hits were observed than false alarms, $F(1, 66) = 3563.61$, $MSe = 0.01$, $p\eta^2 = .98$, and although there was an overall difference between Free Recall and Cued Recall, $F(1, 66) = 15.74$, $MSe = 0.01$, $p\eta^2 = .19$, there was no significant interaction, $F(1, 66) = 2.72$, $MSe = 0.01$, $p = .10$, $p\eta^2 = .04$. Despite the lack of significant interaction, a priori planned comparisons showed that there was no difference in hit rates, $t(66) = 1.55$, $p = .13$, $d = 0.38$, although there were more false alarms in Cued Recall than in Free Recall, $t(66) = 4.03$, $d = 0.99$.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Free Recall vs Cued Recall) mixed ANOVA. More recalls than intrusions were observed in general, $F(1, 66) = 57.90$, $MSe = 9.89$, $p\eta^2 = .47$, and although more responses were observed in Cued Recall than in Free Recall, $F(1, 66) = 42.35$, $MSe = 4.73$, $p\eta^2 = .39$, these two factors did not interact, $F(1, 66) < 1$. A priori follow-up analyses confirmed the main effect of condition, finding that cues increased both recalls and intrusions, $t(66) = 3.55$, $d = 0.87$, and $t(66) = 4.01$, $d = 0.99$, respectively.

Free Recall in Experiment 9 vs Experiment 10. There was no difference in the number of studied words produced in the Free Recall conditions of Experiments 9 and 10, $t(66) = 0.72$, $p = .47$, $d = 0.18$. Free recall recognition accuracy was first analyzed in a 2 (old vs new) X 2

(Experiment 9 vs Experiment 10) mixed ANOVA. More hits were observed than false alarms, $F(1, 66) = 8086.99$, $MSe = 0.004$, $p\eta^2 = .99$, and although there was no overall difference between experiments, $F(1, 66) = 1.39$, $MSe = 0.004$, $p = .24$, $p\eta^2 = .02$, there was a significant interaction, $F(1, 66) = 8.18$, $MSe = 0.004$, $p\eta^2 = .11$. This interaction indicated that although false alarms were equivalent, $t(66) = 1.62$, $p = .11$, $d = 0.40$, hit rates were lower in Experiment 10, $t(66) = 2.29$, $d = 0.56$.

Finally, overall free recall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 9 vs Experiment 10) mixed ANOVA. More recalls than intrusions were observed in general, $F(1, 66) = 188.89$, $MSe = 3.54$, $p\eta^2 = .74$, however there was no difference between experiments and no interaction, $F(1, 66) < 1$, and $F(1, 66) = 2.69$, $MSe = 3.54$, $p = .11$, $p\eta^2 = .04$, respectively. A priori follow-up analyses confirmed these results, finding no difference between experiments in either recalls or intrusions, $t(66) = 1.09$, $p = .90$, $d = 0.27$, and $t(66) = 1.62$, $p = .11$, $d = 0.40$, respectively. Thus, apart from the fact that hit rates were slightly lower in the Free Recall condition of Experiment 10 compared to Experiment 9, there were no differences between experiments.

Cued Recall in Experiment 9 vs Experiment 10. More studied items were produced in Experiment 9 compared to Experiment 10, $t(67) = 8.12$, $d = 2.00$. Cued recall recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Experiment 9 vs Experiment 10) mixed ANOVA. More hits were observed than false alarms, $F(1, 67) = 3546.57$, $MSe = 0.01$, $p\eta^2 = .98$, and although there was an overall difference between experiments, $F(1, 67) = 6.29$, $MSe = 0.01$, $p\eta^2 = .09$, there was no significant interaction, $F(1, 67) < 1$. Planned, a priori follow-ups found that although hit rates were equivalent across experiment, $t(67) = 1.21$, $p = .23$, $d = 0.30$, false alarm rates were higher in Experiment 10, $t(66) = 2.09$, $d = 0.51$.

Finally, overall cued recall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment 9 vs Experiment 10) mixed ANOVA. More recalls than intrusions were observed in general, $F(1, 67) = 187.47$, $MSe = 12.04$, $p\eta^2 = .74$, and although there was a difference between experiments, $F(1, 67) = 31.30$, $MSe = 5.13$, $p\eta^2 = .32$, these two factors interacted, $F(1, 67) = 40.64$, $MSe = 12.04$, $p\eta^2 = .38$. Follow-up analyses found that although more recalls were observed in the Cued Recall condition of Experiment 9, $t(67) = 7.20$, $d = 1.78$, more intrusions were observed in the Cued Recall condition of Experiment 10, $t(67) = 2.85$, $d = 0.70$.

Discussion

Several conclusions can be drawn from Experiment 10. First, hit rates were equivalent in the Free Recall and Cued Recall conditions of Experiment 10. This supports the idea that the hit rate disadvantage seen for cued recall in earlier experiments was due to the fact that cues were not cueing studied items in the same functional or subjective sense as they were studied. Hence, although in past experiments cueing was leading subjects to produce nominally studied words, these words were frequently functionally distinct from study, and hence, unrecognizable. To the subject, they essentially were not the same words as had been studied.

Second, the Free Recall condition in Experiment 10 replicated that of Experiment 9, albeit with a slightly higher false alarm rate (and hence more intrusions). This higher intrusion rate is interesting considering that cue-target pairs were unrelated in Experiment 10. Specifically, if intrusions were the result of subjects fluently producing targets, one might imagine that when cue-target pairs were unrelated, as in Experiment 10, cues could not lead subjects to produce targets fluently. Cues may have still led to fluent free association of related words, but subjects should have rejected these related words on the basis of their knowing that cues and targets were unrelated in Experiment 10. Hence, overall, cues should have led to fewer

fluent productions that were compatible (i.e., unrelated) with the cue-target relation, and intrusions should have declined.

However, this explanation assumes that subjects are being relatively savvy in their recall and recognition strategies. Although it is completely possible for subjects to be implementing a strategy that involves evaluating the cue-target compatibility, and rejecting incompatible targets, it remains speculative whether or not this is the case. Especially when cues and targets are randomly paired together, there will certainly be some cue-target pairs that are highly unrelated (e.g., ZEBRA-BOTTLE), however others may appear to be related to some degree (e.g., DOG-PARK). From the view of the subject then, although some cue-target pairs are unrelated, others are related. At test then, it would not be useful to reject targets that are compatible with cues because (from the subject's perspective), some of the targets *should* be compatible with the cues. Indeed, when reviewing a list of randomly paired words, there definitely is a subjective sense where many of the pairs appear to be related. Thus, in Experiment 10, subjects were not necessarily viewing functionally unrelated cue-target pairings, and hence, if cues at test helped subjects fluently produce a semantically related candidate, subjects could still have accepted this candidate, regardless of the fact that it violated the nominal (lack of) relation between cues and targets in the study list

Returning to the conclusions that can be drawn from Experiment 10: Third, and most important, although the unrelated cues in Experiment 10 helped subjects to produce studied words and ultimately led to more items being recalled than in the Free Recall condition, the unrelated cues here were less effective than the related cues in Experiment 9. That is, unrelated cues were not as helpful in producing studied words, which ultimately led to a smaller recall advantage in Experiment 10 than Experiment 9. Thus, even though unrelated paired-associate

cues can be useful cues at test, insofar as they can help subjects to produce studied words without necessarily impairing recognition, not surprisingly they are decidedly less advantageous than related paired-associate cues.

Although they were less useful than the related cues in Experiment 9, the unrelated cues in Experiment 10 did confirm the encoding specificity account of the hit rate disadvantage for cueing. Namely, when targets are encoded in the context of cues at study, those cues will be effective cues at test.

Weakly Related Cues at Study, Strongly Related Cues at Test

The results of Experiment 9 and 10 both suggest that cues are effective when they cue the functional sense of the targets that was encoded at study. This implies that the hit rate disadvantage for cued recall seen previously was due to the fact that cues were cueing subjects to a different subjective sense of the targets than was encoded at study. Experiment 11 serves to test this hypothesis directly by explicitly cueing a different subjective sense of the targets than was encoded during study. In Experiment 11, subjects encoded targets in reference to weak semantic associates (e.g., TRAIN-BLACK), and at test were given strong semantic associates as cues (e.g., WHITE-?). This setup means that Experiment 11 closely resembles recognition failure of recallable words paradigm (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973). If the results of this experiment parallel those of the previous cued recall experiments in this dissertation (e.g., Experiment 1), insomuch as semantic associates that were used as cues were not necessarily cueing words in the same functional sense as they were studied, then there should be a production advantage accompanied by a hit rate disadvantage in Experiment 11, as seen in those previous cued recall experiments.

Before proceeding, it is worth noting that the study conditions of Experiment 11 may actually closely resemble those of standard, individual item study phases. Although this may seem counter-intuitive at first, when association by contiguity is considered, it becomes apparent that in typical study conditions subjects likely do encode unrelated words as pairs and sets. That is, work on association by contiguity has shown that subjects often encode items in the order in which they appear (Howard & Kahana, 1999; Kahana, 1996; McDaniel & Bugg, 2008; Ozubko & Joordens, 2007). Hence, even in a study list where individual words are presented on each

trial, subjects are likely encoding items that appear together, as sets. Thus, if DUCK was

followed by PLATE, subjects may encode these two words in almost the exact same fashion as

in Experiment 11, where DUCK-PLATE is presented on a single trial. The point here is that

although Experiment 11 forces subjects to encode pairs of words together, this may in fact be

very similar to how subjects encode random lists of words anyway. Thus, once again,

Experiment 11 is expected to provide analogous results to the Free Recall and Cued Recall

conditions of Experiment 1.

Method

Participants. Twenty-seven students from the University of Waterloo participated in

Experiment 11 in exchange for course credit toward a psychology course.

Materials. Experiment 11 used the same word pool as Experiment 1. At study, cues for

the cue-target pairings were always the 8th strongest associate of each target. At test, the cues

were always the 1st strongest associate of each target (both studied and unstudied).

Procedure. Experiment 11 was run identically to Experiment 9 except for the modified

stimuli as described above.

Results

The results of Experiment 11 can be seen in Table 6. As in Experiment 9 and 10, the

number of cues produced during test is reported, however these were not analyzed. The results

of Experiment 11 were compared to the Free Recall condition of Experiment 10 because, in

Experiment 10 subjects studied unrelated cue-target pairs at study and then performed a free

recall test. Hence, the study conditions of Experiment 11 (weak cue-target pairs) are the closest

match to those of Experiment 10.[7] The comparison between Experiment 11 and Experiment 10

then closely resembles the comparison between free recall and cued recall in Experiment 1,

except that weakly/unrelated items are studied along with the studied items initially, thus

simulating encoding study items in an idiosyncratic context.

More studied words were produced in Experiment 11 than in the Free Recall condition of

Experiment 10, $t(59) = 3.92$, $d = 1.02$. Recognition accuracy was first analyzed in a 2 (old vs

new) X 2 (Experiment 11 vs Experiment 10 Free Recall) mixed ANOVA. More hits were

observed than false alarms, $F(1, 59) = 1835.15$, $MSe = 0.01$, $p\eta^2 = .97$, and although there was no

overall difference between Experiment 11 and Free Recall in Experiment 10, $F(1, 59) = 2.65$,

$MSe = 0.02$, $p = .11$, $p\eta^2 = .04$, there was a significant interaction, $F(1, 59) = 17.66$, $MSe = 0.01$,

$p\eta^2 = .23$. Subsequent comparisons showed that there were both fewer hits and more false

alarms in Experiment 11 compared to Free Recall in Experiment 10, $t(59) = 3.17$, $d = 0.83$, and

$t(59) = 2.29$, $d = 0.60$, respectively.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Experiment

11 vs Experiment 10 Free Recall) mixed ANOVA. More recalls than intrusions were observed

in general, $F(1, 59) = 93.90$, $MSe = 5.12$, $p\eta^2 = .61$, and although more responses were observed

in Experiment 11 than in Free Recall in Experiment 10, $F(1, 59) = 10.39$, $MSe = 4.10$, $p\eta^2 = .15$,

these two factors did not interact, $F(1, 59) < 1$. A priori follow-up analyses confirmed the main

effect of condition, finding that cues increased both recalls and intrusions, $t(59) = 2.18$, $d = 0.57$,

and $t(59) = 2.14$, $d = 0.56$, respectively.

---

[7] Note that practically, however, because the Free Recall results did not differ in Experiments 9 or 10, there was no qualitative difference in results if Experiment 9 was used instead. The results of the comparison between Experiment 11 and Experiment 10 only are reported because they provide the best fit conceptually.

<u>Discussion</u>

Although recall was increased in Experiment 11 compared to the Free Recall condition in Experiment 10, false alarms, and hence intrusions, once again rose. Thus, in every cued recall study so far, false alarms have increased. More importantly, however, the results of Experiment 11 mimic those of previous cued recall experiments in this dissertation that found a hit rate deficiency (e.g., Experiment 1). These results suggest that the reason that cues impaired hit rates in Experiment 1 and other experiments is that studied words were being encoded idiosyncratically in subjective contexts that were not always related to strong semantic associates of those studied words. As a result, subjects were producing studied words from strong semantic associates that led to a different functional sense of the targets than was encoded, and hence the produced words were functionally not the words that were studied. Subjects therefore failed to recognize these nominally studied stimuli and this impaired hit rates.

At a theoretical level, Experiment 11 highlights an important point: Namely, that in standard study conditions, where subjects study a random list of words with words presented individually, subjects are likely encoding words in unusual or atypical functional senses. This effect arises because each studied word sets a context for each subsequent word to be studied. Because words usually are unrelated (at least in the typical recall experiments), each word is therefore being encoded in an unusual or atypical context, hence fostering an unusual or atypical functional sense of that word. Effectively, rather than random word lists ensuring that words are each studied in relative isolation, with their most common meanings coming to light, Experiment 11 suggests that instead, just the opposite occurs.

Finally, it should be noted here, too, that Experiment 11 closely parallels Tulving and colleagues' studies on recognition failure of recallable words (Thomson & Tulving, 1970;

92

Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973). Namely, Tulving and colleagues found that strong semantic associates were ineffective cues if words had been encoded in the presence of weak semantic associates. The present findings directly support this result, and demonstrate that cues do lead to a recognition impairment, as Tulving and colleagues found. Yet, although the results from Experiment 11 could be viewed as a simple extension of the recognition failure of recallable words paradigm, interpreted in the presence of the earlier studies here, a more sweeping conclusion can be drawn: Recognition failure of recallable words may not be an isolated phenomenon constrained to a specific paradigm, but rather may be a regularity of traditional cued recall paradigms. This fact has been missed in past studies of cued recall, however, because subjects were rarely asked to recognize their own "recalls."

One counter to such an argument is to point out that in Experiment 8, when subjects were encouraged not to produce words that they would not be able to recognize subsequently, the hit rate disadvantage (i.e., unrecognizable recalls) disappeared. This suggests that when the production pressure of recognized recall is eliminated there is no hit rate disadvantage to speak of, and hence past studies involving cued recall may not have included many unrecognizable "recalls." However, it is important to remember that although unrecognizable "recalls" may be the result of guessing, and there are many other aspects to traditional cued recall paradigms that encourage guessing beyond whether responses are or are not forced.

In traditional cued recall paradigms, only cues for studied words are presented at test and subjects know that they will not need to justify any word that they produce as "old." Both of these facts should encourage more guessing in traditional recall paradigms than in recognized recall, regardless of whether responses are forced. For example, the fact that subjects are required to explicitly recognize their own productions as "old" or "new" in recognized recall

should actively discourage guessing, as subjects may worry about getting "caught" when they refuse to recognize a production as "old." Similarly, when only cues for studied words are presented at test, subjects have a basis to make educated guesses when no items can be recalled, and indeed may adopt this strategy. Furthermore, the instructions in Experiment 8 explicitly discouraged guessing. None of these controls exist in typical cued recall paradigms. As a result of the methodological controls for guessing in Experiment 8, the fact that no unrecognizable "recalls" (i.e., unrecognizable produced studied items) were observed in Experiment 8 could have been due to the fact that numerous other factors which normally encourage guessing, and so encourage unrecognizable "recalls," were controlled.

Indeed, Jacoby and Hollingshead (1990) represent the only case where traditional cued recall is compared to a recognized recall-like paradigm. In their work, Jacoby and Hollingshead examined cued recall performance in a condition where subjects were asked to recall words to cues and to pass if no studied word came to mind vs a condition where subjects were encouraged to give a response to every cue and subsequently to recognize those responses as "old" or "new." Jacoby and Hollingshead found that 77% of studied items were produced as recalls in traditional cued recall paradigms whereas only 51% of studied items were both produced and recognized. These results demonstrate that subjects may be producing more words in traditional cued recall paradigms than they are capable of recognizing. Furthermore, the fact that subjects were not forced to respond in Jacoby and Hollingshead's cued recall condition speaks against the idea that unrecognizable "recalls" are merely the result of forcing responses to cues.

Whether one accepts the suggestion that recognition failure of recallable words may be widespread and that unrecognizable "recalls" often contaminate cued recall scores in typical cued recall paradigms, it should be clear that the practice of gauging subjects' ability to

recognize their own output is a valuable one. If collecting such data were commonplace, we

could easily look back to an array of past literature to assess how often recognition failure

occurs. This fact alone supports the notion that it may be advisable to adopt recognized recall as

a standard method for assessing recall performance.

Experiment 12:

Subjective Cues

The past three experiments have demonstrated that cues which are encoded at study are effective at assisting production at test, while leaving recognition unharmed. These results are completely in line with Tulving and colleagues' past work on recognition failure of recallable words (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973) in that they demonstrate the importance of encoding specificity. However, if cues are only effective to the degree that they cue subjects to the same subjective sense of the stimulus that was experienced at study, then it is possible that subject-produced cues may be the most effective form of cueing. In essence, Experiments 9 through 11 have all relied on subjects studying pairs of words in the hopes that the cue-target pairing would encourage the target to be encoded in reference to the cue. Although unrelated cues were effective in Experiment 10 at helping subjects to produce extra studied words without impairing the ability to recognize those words, they were less effective than the stronger semantic associate cues used in Experiment 9. Thus, although Experiments 9 through 11 demonstrate that cues encoded at study can be useful cues at test, they also demonstrate that some cues are more effective than others when it comes to production.

That strong semantic associate cues are more effective than unrelated cues indicates that the relation between the cue and the target matters. Cues will be more effective to the degree that they are related to the target. To explain this finding, consider the fact that the strength of association between cue-target pairs is determined by the probability with which a cue will give rise to a target, via free association. Normed free association data are, by definition, based on responses from numerous subjects. Although such data can reveal regular relations between cue-

target pairs, and so indicate which cue-target pairs are most likely to be related, they miss idiosyncratic relations that could exist between cue-target pairs for some subjects but not others. Hence, strong semantic associates are better cues than are unrelated words with regard to helping subjects to produce studied words because they have a high propensity, on average, to encourage a random subject to think of the target. But it is likely that the ideal cue-target pairing, for any given subject, varies substantially.

If the above explanation of cues is correct, it suggests that there may be an even more effective method for cueing than using strong semantic associates that were encoded at study. Namely, if cues are effective to the degree that they bring to mind the target, why not simply have subjects free associate a word to each studied word and then use those response words as cues? A free associate to a studied word should be extremely compatible with that studied word for the subject who produced it because what is produced will be determined by the subjective manner in which that subject views that word. Furthermore, free associates should naturally be encoded along with studied items as they will be determined by how the studied items are viewed by subjects; this fact was shown to be important in Experiments 9 through 11 to keep from impairing the hit rates.

For example, if GENERAL is the studied word in question, and a subject imagines a general in an underground bunker issuing commands, he or she may produce COMMAND as a free associate. Not only is COMMAND semantically related to GENERAL, but it is specifically semantically related to the subject's own sense of the word GENERAL (at that time) and thus, may be more semantically related to the studied word GENERAL, for that subject, than any other word at that moment (e.g., ARMY, SOLDIER, SPECIFIC). Furthermore, COMMAND was also naturally present at study for the subject in question, therefore there is no need to force

97

this cue to be encoded along with GENERAL: It was in fact encoded by default. Based on the fact that free associates to studied items should be highly subjectively related to the studied words, they should have a high propensity to give rise to those targets during the production phase at test. Furthermore, because free associates to studied words are likely encoded along with studied words at the time of study, an encoding specificity account would expect their later use as cues to leave hit rates for the studied targets unaltered, as in Experiments 9 and 10.

Thus, Experiment 12 evaluates the effectiveness of using free associates to studied items as subjective cues at test. Before proceeding, however, it should be noted that the act of free associating words to studied items may actually lead to those studied words being encoded more effectively than in past experiments, where subjects were simply asked to remember studied words. That is, free associating to studied words encourages deeper processing, generation, and possibly mental imagery, all of which are known to enhance encoding (Craik & Lockhart, 1972; Craik & Tulving, 1975; Paivio, 1969, 1971; Slamecka & Graf, 1978). Given this fact, in Experiment 12 both a free recall and a cued recall condition were examined and directly compared to each other, rather than to previous experiments. In both conditions, subjects produced free associates to items at study. Hence, the results from the Cued Recall condition in Experiment 12 can be compared to the Free Recall condition without any differences in encoding.

Method

Participants. Thirty students from the University of Waterloo participated in Experiment 12 in exchange for course credit toward a psychology course. Fifteen students participated in each of the Free Recall and the Cued Recall conditions.

Materials. Experiment 12 used the same word pool as Experiment 1.

Procedure. Experiment 12 was run identically to Experiment 1 except that, during study, subjects had to type in a free associate to each study item. Subjects were given an unlimited time to do so, but in practice took no longer than a few seconds on each study word. Subjects were informed to come up with associates that did not contain the study words themselves (e.g., FROGS is not an acceptable associate of FROG, and neither would FUNDAMENTAL be an acceptable associate of FUN or vice versa).

At test, in the Cued Recall condition, subjects' free associate cues were randomly inter-mixed with 24 cues for unstudied words, exactly as was done in Experiment 1. That is to say that the cues for unstudied words were selected from the same pool as Experiment 1. In essence, Experiment 12 replicates Experiment 1 except that subjects generate free associates at study, and those associates replace the cues for studied words that would have been used in Experiment 1. In the Free Recall condition, of course, no cues were shown at test.

Results

The results of Experiment 12 can be seen in Table 7. Given that, in Experiment 12, subjects produced a free associate at study to each studied word, it was inappropriate to compare Experiment 12 to the previous free recall and cued recall controls. That is, any differences between Experiment 12 and previous conditions could be due to either the subjective cues at test or the enhanced encoding afforded from the deeper processing of Experiment 12. Hence, direct comparisons with past experiments were not performed and instead the Free Recall and Cued Recall conditions of Experiment 12 were directly compared to each other.

In Experiment 12, more studied words were produced in the Cued Recall condition than in the Free Recall condition, $t(28) = 9.73$, $d = 3.68$. Recognition accuracy was first analyzed in a 2 (old vs new) X 2 (Free vs Cued Recall) mixed ANOVA. More hits were observed than false

Table 7. Mean number of words produced, hit and false alarm rates of produced words, number of words recalled, and number of intrusions in free recall and cued recall in Experiment 12. Standard errors are shown in parentheses below means.

Condition	Number of Productions			p("old")		Overall Performance	
	Old	New	Repeats	Old	New	Recalls	Intrusions
Free Recall	9.73	37.67	0.47	1.00	0.05	9.73	1.80
	(0.68)	(0.82)	(0.24)	-	(.01)	(0.68)	(0.28)
Cued Recall	19.73	26.53	1.60	0.997	0.07	19.67	2.13
Subjective Cues	(0.77)	(0.87)	(0.46)	(.003)	(.02)	(0.77)	(0.65)

alarms, $F(1, 28) = 7126.62$, $MSe = 0.002$, $p\eta^2 = .996$. However, there was no difference between

Free Recall and Cued Recall and no interaction, $F(1, 28) = 1.06$, $MSe = 0.002$, $p = .31$, $p\eta^2 = .04$,

and $F(1, 28) = 1.74$, $MSe = 0.002$, $p = .20$, $p\eta^2 = .06$, respectively.

Finally, overall performance was analyzed in a 2 (recalls vs intrusions) X 2 (Free Recall

vs Cued Recall) mixed ANOVA. There were more recalls than intrusions in general, $F(1, 28) =$

284.53, $MSe = 8.55$, $p\eta^2 = .91$, and although there was a main effect of condition, $F(1, 28) =$

130.10, $MSe = 3.04$, $p\eta^2 = .82$, there was a significant interaction, $F(1, 28) = 40.43$, $MSe = 8.55$,

$p\eta^2 = .59$. Follow-up analyses indicated that there were more recalls in the Cued Recall

condition than in the Free Recall condition, $t(28) = 9.70$, $d = 3.67$, but intrusions did not differ,

$t(28) = 0.47$, $p = .64$, $d = 0.18$.

Discussion

Subjective cues are by far the most effective method of cueing observed yet in this

dissertation. Subjective cues combine all the benefits of free recall with all those of cued recall.

In essence, subjective cues help subjects to produce far more studied items than in free recall,

however subjective cues do not impair subjects' ability to recognize those items. The net result

is that subjective cues significantly improve recall without increasing intrusions.

Moreover, it is worth emphasizing that, in Experiment 12, subjective cues did not

significantly increase intrusion rates beyond that of free recall. This is the first time in this

dissertation that provision of cues has not inflated false memories. Most likely, subjective cues

help subjects control false memories because of their subjectively memorable and effective

nature. Simply put, subjects should often recognize their own cues when they see them, since

they themselves generated them (cf. Slamecka & Graf, 1978). Because subjective cues are likely

so memorable, new cues should stand out to subjects, and subjects can be assured that they will

likely not produce a studied word on these trials. Therefore, subjects can readily reject items produced in response to new cues. Furthermore, because subjects' own cues are highly effective cues in the sense that they are very subjectively specific to the studied words, they should be highly effective at helping subjects to produce a studied, and subsequently recognized, word. Hence, because subjective cues are so memorable and effective, there should be very few opportunities for intrusions to subjective cues. The net result is that few false memories are expected in response to subjective cues.

At a practical level, however, although the results of Experiment 12 are encouraging for cued recall advocates, they do suggest that the most effective cue comes from the subject whom you are cueing. In essence, if someone cannot retrieve some information, and they have not provided you with a cue to give back to them, then there may be little you can do. If you already knew what an individual was trying to retrieve, Experiment 4 would suggest that you could use highly specific cues, and this could help jog someone's memory. Experiments 9 and 10 further suggest that you could use cues that were specific to what was studied, and this could help jog someone's memory. Yet, both of these scenarios presuppose that one already knows what they are trying to get someone else to provide. In other words, why would you bother cueing someone if you already had the answer? Obviously in laboratory contexts, cueing is interesting in and of itself but, in more realistic settings, it would seem that as a whole the results of these studies suggest that cues will reliably increase false memories, unless they are subjective.

For easy reference, a summary of the results from each of the experiments in this dissertation is reported in Table 8. The goal of this dissertation has been to challenge the long held view that there is a unilateral benefit of cued recall beyond free recall. Specifically, this dissertation has argued that although cued recall often leads to apparently superior performance to free recall in traditional recall paradigms, response pressures and guessing in these paradigms frequently underlie this difference. To control for these differences, the recognized recall procedure was developed and implemented.   This procedure is an active attempt to better equate testing conditions of free recall and cued recall, specifically by controlling for response pressure and both controlling and measuring guessing. Recognized recall better equates response pressure and removes much of the onus for unmeasured guessing in cued recall. Furthermore, not only does recognized recall allow for finer-grained measures of subjective memory than do traditional recall paradigms, but it does so while still permitting traditional measures of recall to be obtained (i.e., the number of studied items produced).

Across 12 experiments using recognized recall, the major result shown is that cued recall is not a uniformly effective mnemonic. When cues are strong (Experiment 4), specific to the encoding context (Experiments 9 and 10), or subjective (Experiment 12), a recall benefit can be seen for cued recall over free recall. However, except for the case of subjective cues, cues increased false alarms in all cases. This increase in false alarms is significant because it indicates that memory sensitivity was declining, despite the fact that subjects were producing more studied words in cued recall than in free recall. Indeed, this heightened false alarm rate in cued recall is worrisome for anyone seeking to use cueing in more realistic settings because it

Table 8. Summary of the Experiments.

Experiment	Label	Description	Result
1	Free and Cued Recall	Free and cued recall were examined in separate groups for both recognized recall and traditional recall paradigms. Only recognized recall was examined in all subsequent experiments.	Cues helped the production of studied words but impaired recognition: In cued recall, hits were lower and false alarms were higher than in free recall. The number of studied words produced provided the same measure of "recall" as traditional recall paradigms.
2	Cued after Free Recall	Cues were provided after a free recall test.	Cues did lead to some novel contributions above free recall but also caused a large increase in intrusions.
3	8 Cues/Trial	Instead of providing a single cue before each test trial, 8 cues were provided, all pointing to the same target.	Providing multiple cues per test trial increased both the number of studied items produced and the false alarm rate above previous cued recall results. The net result, however, was a rise in intrusions.
4	Strong Associate Cues	The semantic associate cues used here were more strongly related to their targets than in other experiments.	Stronger associate cues did increase the number of studied items produced above previous cued recall results. This ultimately led to more recalls despite the fact that recognition was still impaired.
5	Status-Identified Cues	Cues were identified as cues for studied words (in blue) or cues for new words (in white).	Identifying cues to subjects increased false alarm rates above previous cued recall results.
6	Studied Word Cues Only	Only cues for studied words were presented at test.	False alarm rates increased compared to previous cued recall results.
7	Delayed Recognition	Rather than recognizing each word as it was produced, recognition was delayed until all productions were completed. Both free and cued recall were examined.	Delaying recognition increased false alarm rates in both free recall and cued recall above previous free recall and cued recall results.

Experiment	Label	Description	Result
8	No Response Pressure	Subjects were instructed never to produce a word unless they thought it was "old."	The results were identical to previous cued recall results except that cues no longer helped subjects to produce more studied words than in free recall.
9	Related Paired-Associates	Related cue-target pairs were studied and those same cues were used at test. Both free recall and cued recall were examined.	When cues were studied with targets, they no longer impaired hit rates. False alarm rates were still inflated, however, as in previous cued recall results.
10	Unrelated Paired-Associates	Unrelated cue-target pairs were studied and those same cues were used at test. Both free recall and cued recall were examined.	When cues were studied with targets, they no longer impaired hit rates. False alarm rates were still inflated, however, as in previous cued recall results.
11	Weakly/Strongly Related Paired-Associates	Weakly related cue-target pairs were studied but, at test, strong associates of the targets were used as cues.	As in previous cued recall results, cues did help subjects produce more studied words, however recognition was impaired. There was some increase in the number of words recalled compared to free recall, but a significant rise in intrusions.
12	Subjective Cues	Subjects produced associates at study, to study words, that were subsequently used as cues at test, inter-mixed with novel words as new cues. Both free recall and cued recall were examined.	Subjective cues were highly effective, helping subjects to produce many more studied items while not impairing recognition. The net result was a rise in number of words recalled in response to cues, with no other change.

suggests that the reliability of information produced in response to cues is lower than when no cues were provided at all.

Furthermore, in many cases, cues were found not to provide a mnemonic benefit over free recall despite the fact that the false alarm costs were still present (Experiments 1 through 7, and 11). That is, there often was a cost without a benefit as a result of cueing. Interestingly, the reason that cues appeared to be relatively ineffective was that they frequently led subjects to produce words, through free association, that lacked subjective memorability. In essence, subjects were producing words that they could not recognize, a finding that aligns with past work on recognition failure of recallable words (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973). Thus, the results here suggest that recognition failure of recallable words may actually be a frequent occurrence—and one that is frequently overlooked— in traditional cued recall paradigms.

Interestingly, the results of Experiment 11 confirm that recognition failure of recallable words occurs when words are encoded in the context of an unrelated or weakly related associate at study and a strong associate is used as a cue at test. However, as recognition failure of recallable words was seen across most of the cued recall conditions where only single words, and not pairs, were being studied, this suggests that single word study phases may actually induce subjects to encode words in a manner consistent with more atypical or unusual meanings of those words. For example, in isolation, the word DUCK may bring to mind a duck in a pond. However, if DUCK is being presented in a study list and is seen directly after ADVENTURE, DUCK may instead bring to mind a cartoon duck character from childhood television. In essence, single-item study phases may actually be bad at bringing about the common or most typical meanings of words because each word that is studied is preceded by an unrelated word,

which sets an atypical context in which to interpret that current word. And when words are encoded in the context of atypical or unusual meanings, strong associates (such as BIRD) can give rise to recognition failure of recallable words, precisely what was observed in many of the single-item study phase cued recall experiments in this dissertation.

In addition to examining the impact of cueing, the experiments reported here have delineated many aspects of the recognized recall procedure itself. Whether recognition occurs immediately after production or at a delay makes little difference (Experiment 7). Furthermore, presenting cues to both studied and unstudied words is actually a less biased practice than presenting only cues to studied words (Experiments 5 and 6). Finally, items that subjects are able to produce but not recognize are eliminated in recognized recall when guessing is strictly discouraged at test (Experiment 8). Given its success in examining subjective memorability in free recall and cued recall more closely than is possible in traditional recall paradigms, its current adoption as a standard in examining cued recall in ERP settings (Allan et al., 1996; Allan & Rugg, 1998; Angel et al., 2009; Angel et al., 2010a; Angel et al., 2010b; Fay et al., 2005; Rugg et al., 1998; Schott et al., 2002), and the other important properties that have been delineated in detail here, recognized recall appears to be preferable to traditional recall paradigms—especially for researchers interested in directly studying the effects of cues.

Interpreting Intrusions

One issue that deserves more attention here is that of intrusions. In the recognized recall procedure, there are two measures of false memory: false alarms (i.e., rates of intrusions) and intrusions themselves. Although intrusions do provide some measure of false memory, it must always be remembered that intrusions are dependent on both the number of new items produced and the false alarm rates. In conditions where subjects are producing more studied items, they

necessarily must also be producing fewer new items. In these conditions, the number of intrusions produced tends to decline, regardless of whether the propensity for false memory has declined. Hence, when dealing with data from recognized recall, false alarms—rather than intrusions per se—are the more appropriate measure to compare between experiments, to look for increasing levels of false memory or guessing.

At first glance, it may seem like a flaw of the recognized recall procedure that the number of new items produced is dependent on the number of studied items produced. And therefore, one criticism of adopting recognized recall over traditional recall paradigms could be that it has such dependencies. However, it should be pointed out that traditional cued recall paradigms share this dependency. That is, in traditional cued recall paradigms, subjects often are given a series of cues and asked to recall a single word to each cue. Necessarily, the number of studied items produced here limits the number of new items that can be produced. Hence, the assumption that intrusions are not dependent on recalls in traditional recall paradigms is often false.

Thus, recognized recall is indeed superior to traditional recall paradigms when it comes to measuring false memories. Because, whereas intrusions in both traditional cued recall paradigms and recognized recall are limited by the number of studied items produced, intrusion rates (i.e., false alarms) can be measured in recognized recall and should be relatively invariant with respect to the actual number of intrusions. Indeed, across many of the present experiments, false alarm rates were observable and comparable even when the raw number of intrusions was quite low. False alarms, as measured in recognized recall, therefore provide a more reliable measure to compare false memories between conditions where the number of studied items produced is changing. The presentation of intrusions in the present experiments was mainly to

allow for more direct comparisons with intrusion rates as measured by traditional recall paradigms in the past literature.

Finally, when originally discussing the issue of intrusions in the introduction to this dissertation, I noted that intrusions are often quite low, although there tend to be more intrusions in cued recall than in free recall (e.g., Tulving & Osler, 1968; Tulving & Pearlstone, 1966; Slamecka, 1972). However, in the present experiments using recognized recall, intrusions occurred regularly and were not at floor. Thus, one potential issue with recognized recall is whether it causes intrusions.

First, it could be that forcing subjects to produce both studied and unstudied words at test fosters intrusions (e.g., as found by Roediger et al., 1993). However, Experiment 6 directly tested this notion and found the opposite: Presenting only cues for studied words at test actually increased false alarms. Experiment 8 further tested the notion that forcing subjects to produce words was the culprit behind the relatively high number of intrusions, but found that intrusion rates were no different than when there was response pressure. Indeed, one reason for observing more intrusions in recognized recall than are typically reported likely is because estimates of intrusion rates based on past literature may underestimate actual intrusion rates. That is, many times intrusions are simply not reported (Allan et al., 1996; Allan & Rugg, 1998; Blaxton, 1989; Lewis, 1971, 1974; Paivio et al., 1994; Reder at al., 1974; Rugg et al., 1998; Schott et al., 2002; Taconnat et al., 2008; Thomson & Tulving, 1970; Tulving & Thomson, 1973; Watkins & Tulving, 1975). Given the sheer number of studies that do not report intrusion rates, it becomes difficult to ascertain whether the intrusion rates in the present experiments truly differ from the typical rates because it is difficult to ascertain what the typical rates actually are. Even assuming, however, that traditional recall paradigms typically observed fewer intrusions than

were observed in the present experiments, one of the major reasons for using recognized recall was to examine free recall and cued recall with a higher resolution procedure. Hence, it seems reasonable to suggest that recognized recall may not just be more accurate at measuring recalls than past paradigms; it may also be more accurate at measuring intrusions. A primary reason for this is the fact that traditional cued recall paradigms do not allow for as much opportunity to produce intrusions as is the case in recognized recall, hence making it more difficult to measure intrusions.

Indeed, as previously described, the number of intrusions observed is inversely related to the number of studied items produced, even in traditional cued recall paradigms. Given this relation, and that experimenters often only present cues for studied items, traditional cued recall paradigms necessarily put limits on the number of intrusions that could be observed. Therefore, it is unsurprising that intrusion rates are lower in previous work than they were in this dissertation.

Why Care about Intrusions?

Regardless of whether one agrees with the present claims about intrusions, one may ask why intrusions even matter. That is, if our goal is to measure what a subject can recall, why do intrusions even need to be considered? Intuitively at least, it feels as though intrusion rates may be interesting and worth examining, but may not have much bearing on recall scores. This intuition likely results from the fact that usually very few intrusions are observed. And even when intrusions are observed, recalls still appear meaningful.

As already discussed in the introduction, the notion that intrusions rates are often low, and so can be ignored, is flawed. In fact, that intrusion rates are often so low may actually be cause for concern. That is, if intrusions are at floor, then any differences between free recall and

cued recall may be misrepresenting the cued recall advantage. In essence, cued recall may lead to a rise in both recalls and intrusions, but because intrusion rates are often at floor, this effect would frequently be missed. Thus, the notion that intrusions can be ignored simply because they may be infrequent is not valid, and may instead suggest the need for more accurate measures of memorability.

Another reason why it might seem reasonable to ignore or to downplay intrusions is that recalls still seem to indicate memory, regardless of intrusions. In recognition memory, neither hit nor false alarm rates are interpretable in isolation. That is, a high hit rate does not necessarily indicate strong memory, if the false alarm rate is also high. Similarly, a low hit rate does not necessarily indicate weak memory, if there are no false alarms. Hence, recognition memory accuracy is completely dependent on the relation between the hit and false alarm rates. In recall, however, this strong relation does not hold.

In recall, the number of words recalled is interpretable, to some degree, regardless of exactly how many intrusions were observed. For example, in Experiment 3, an equivalent number of recalls and intrusions was observed. However, this fact alone would certainly not allow the claim that subjects had no memory in Experiment 3, unlike the case where recognition hits and false alarms were equivalent. To successfully produce and recognize as many studied words as was seen in Experiment 3, subjects would need to have some memory for the words that had been studied. Thus, to some degree, regardless of the number of intrusions observed, recalls still provide a measure of memory.

Although intrusions do not invalidate recalls, they do provide an important measure of false memory. Especially when it comes to comparisons between conditions, if increases in intrusions are noted in the absence of increases in recall, it can certainly be said that memory

performance is worse in the condition with more intrusions. Furthermore, intrusions can have real consequences, in the sense that intrusions reduce the reliability of a subject's output. Hence, if absolute accuracy is paramount, especially in real world settings (e.g., eyewitness testimony), a retrieval method that saw both recalls and intrusions rise might actually be considered less accurate than a condition with fewer recalls but also fewer intrusions. The point here is not that intrusions have a clear relation to how recall scores should be interpreted, as false alarms do to hits in the study of recognition, but merely that intrusions can add to the interpretation of how recall performance changes due to experimental manipulations. Intrusions can provide an important measure of guessing and false memory, even if they do not invalidate memory as measured by recall scores.

As a case in point, intrusions are of current interest to neuroscientists working with patients. Intrusions are common in patients with Alzheimer's (Fox et al., 1998; Fuld et al., 1982; Kern et al., 1992; Lowenstein et al., 1991; Manning et al., 1996), with frontal lobe dysfunction (Dalla Barba et al., 1995; Kapur & Coughlan, 1980), and with hippocampal damage (Deweer et al., 1995). Researchers examining Alzheimer's patients have found that intrusions are also even more common in cued recall than in free recall (Dalla Barba et al., 1995; Dalla Barba & Wong, 1995), supporting the idea that cues may be inflating recall scores in traditional cued recall paradigms through nothing more than guessing and response pressure. More importantly, however, current work examining these intrusions has found not only quantitative but also qualitative differences between free recall and cued recall. Desgranges et al. (2002) found that cued recall intrusions were almost uniformly semantically related to studied words (99% of intrusions were related to studied words) whereas free recall intrusions were fewer, and less often semantically related to studied words (79% of intrusions were related to studied words).

Given the possible neurological link between intrusions and specific brain regions, and that intrusions differ not only quantitatively between free recall and cued recall but possibly also qualitatively, it seems reasonable to suggest that they deserve attention, just as recalls do. Although intrusions and recalls do not share a relation like hits and false alarms, where the difference between them is a better measure of memorability than just recalls, they nonetheless provide important measures of memory. Where recalls can be taken as a measure of memory, intrusions can be taken as a measure of false memory. And these two estimates seem to be relatively independent.

Correct Intrusions

All of this discussion leads up to an issue that has been deferred until now—that of correct intrusions. Across every cued recall experiment, except Experiment 12, cues were found to increase false alarm rates. The present explanation of this finding was that cues lead subjects to produce unstudied candidates more fluently than if no cues had been given. This enhanced fluency is then occasionally mistakenly attributed to study and, as a result, these items are identified as "old," hence increasing false alarm rates.

One concern with this explanation is that, if it is true that cues increase the fluency with which items can be produced, and that cues can also sometimes direct subjects to free associate studied words, then occasionally subjects should produce an actually studied word fluently, due to the cue. Hence, on some trials, subjects should both produce and recognize a studied word but this will have been due to the bias of the cue and not to the fact that that word was truly remembered and recalled. In essence, subjects could occasionally produce intrusions that just happen to be studied words (i.e., correct intrusions) however, because these intrusions are studied words, they then are incorrectly counted by the researcher as recalls, not as intrusions—

113

after all, the researcher cannot tell. In all of our past experiments then, our measures of recall

may actually have been inflated by correct intrusions.

The notion of correct intrusions may seem counter-intuitive at first. How could a subject

produce a studied word and then recognize it as studied, but this be considered an intrusion, not a

recall? The answer lies in *why* the subject recognized the word as studied. If a subject produces

a studied word, and that word is fluent because it was recently studied, and the subject realizes

this and calls that word "old," this is a correct recall because it provides a measure of memory.

However, if a subject produces a studied word simply due to the cue, and that word would not

have been fluent from study, but because that word was produced from a cue it now is fluent

enough to be recognized, that should be considered to be a correct intrusion. In essence, it was

just happenstance that the subject actually produced the right study word. Had the cue led to a

different word, that word also might have been accepted as "old" due to the fluency afforded by

the cue. That the word happened to be a studied word, however, does not change the fact that it

was an intrusion, as it lacks subjective memorability and was only called "old" due to the effect

of the cue.

The notion of correct intrusions is somewhat speculative. Indeed, one might argue that

the fact that cues do increase the fluency of studied items, and cause them to be accepted where

they would not be if the cue was not present, is in fact the mnemonic effect of cues. Hence, one

could suggest that correct intrusions are not a bias, they *are* the mnemonic benefit of cueing. To

some degree then, whether correct intrusions are considered an "error" is somewhat subjective.

Nonetheless, the argument here is that correct intrusions are probably best conceived of as errors

in the sense that, on correct intrusion trials, subjects would have called any word produced from

the cue "old," and therefore subjects simply "got lucky" when they produced a studied word

through free association to the cue. Had any other word been produced, it would have been counted as an intrusion.

If correct intrusions do exist as has been argued here, then all of the recall data reported thus far have been contaminated with correct intrusions. That is, because correct intrusions are correct, they have been counted as recalls, whereas they should have been counted as intrusions. If this is the case, then the question becomes whether correct intrusions can be estimated or corrected for. Using equations similar to those of Jacoby's (1991) process-dissociation procedure—and indeed identical to Yonelinas' dual-process signal detection model (Yonelinas 1994; Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996)—it turns out that estimates of correct intrusions can readily be calculated. Thus, we can estimate to what extent recall scores may have been biased by correct intrusions. Under this model, assume that:

$$H = M + (1 - M)F \tag{E1}$$

$$FA = F \tag{E2}$$

Where H is hit rate, FA is false alarm rate, M is the probability of recognizing an item due to memory of the study episode (i.e., M is what the hit rate would be if it were based solely on memory), and F is the probability of identifying an item as studied due solely to non-study-based fluency. (It should be noted that E1 is mathematically identical to the inclusion equations used in the process-dissociation procedure, although rewritten to increase conceptual clarity for present purposes).

From these two equations, we can see that this model assumes that false alarms are a pure measure of the influence of non-study-based fluency. Although such a strict, process-pure claim about false alarms is unlikely to be true, false alarms may provide a reasonable estimate of non-

study-based fluency—indeed, the best estimate of non-study-based fluency from the available data. Thus, this assumption is accepted for present purposes.

In terms of hit rates, the model assumes that the probability of calling an item "old" at study is equal to the probability that the subject remembers that item, plus the probability that the subject does not remember that item but that the item is fluent for non-study reasons. Thus, "old" responses that arise due to M are true recalls in this model, whereas "old" responses due to $(1 - M)F$ are correct intrusions. Solving for M gives:

$$M = \frac{H - F}{1 - F} \qquad\qquad (E3)$$

Finally, based on evaluation of the equation for hit rates, we can calculate true recalls and correct intrusions as M or $(1 - M)F$ multiplied by the number of studied items produced. With these equations, M can be calculated, and this value can be used to calculate true recalls and correct intrusions for data from all of the previous experiments. In essence, it is possible to examine the degree to which correct intrusions may have inflated recall scores in the previous analyses of the experiments reported here. For Experiments 1 through 8 and 12, the calculations of M, true recall, correction intrusions, along with other relevant data, are reported in Table 9. For Experiments 9 through 11, this information is provided in Table 10.

Examining Table 9, what is immediately apparent is that correct intrusion rates were higher in cued recall than in free recall.[8] The two exceptions are Experiments 8 and 12. Before considering this further, it should be acknowledged that by and large correct intrusions did not play a great biasing role in the estimates of recall reported throughout this dissertation. That is, although these estimates do suggest that correct intrusions occur more frequently in cued recall

---

[8] Note that relevant statistical comparisons were carried out. Given the exploratory and impressionistic nature of the discussion concerning correct intrusions, tests are not reported in the main text. However, for the curious, relevant analyses are reported in the Appendix.

Table 9. Mean hit and false alarm rates of produced words, *M* estimate, number of words recalled, and number of true recalls and correct intrusions as estimated with *M* in Experiments 1 through 8 and 12. Standard errors are shown in parentheses below means.

	p("old")				True	Correct
Condition	**Old**	**New**	***M***	**Recall**	**Recall**	**Intrusions**
Free Recall						
Experiment 1	0.97	0.06	0.97	8.52	8.50	0.02
	(.01)	(.01)	(.01)	(0.74)	(0.74)	(0.01)
Experiment 2	0.92	0.08	0.90	8.23	8.15	0.08
	(.02)	(.02)	(.03)	(0.71)	(0.73)	(0.04)
Experiment 7	0.94	0.06	0.94	8.09	8.07	0.01
Delayed Recognition	(.02)	(.01)	(.02)	(0.63)	(0.63)	(0.005)
Experiment 12	1.00	0.05	1.00	9.73	9.73	0.00
Subjective Cues	-	(.01)	-	(0.68)	(0.68)	(0.00)
Cued Recall						
Experiment 1	0.78	0.13	0.75	8.31	8.03	0.28
	(.02)	(.01)	(.03)	(0.49)	(0.50)	(0.05)
Experiment 3	0.77	0.29	0.67	9.82	8.61	1.21
8 Cues/Trial	(.04)	(.03)	(.05)	(0.92)	(1.00)	(0.27)
Experiment 4	0.78	0.16	0.73	12.14	11.40	0.75
Strong Associate Cues	(.03)	(.03)	(.04)	(0.68)	(0.73)	(0.26)
Experiment 5	0.82	0.18	0.78	9.35	8.95	0.40
Status-Identified Cues	(.03)	(.03)	(.04)	(0.91)	(0.98)	(0.14)
Experiment 6	0.80	0.22	0.73	9.27	8.59	0.67
Studied Word Cues Only	(.03)	(.03)	(.05)	(0.77)	(0.86)	(0.17)
Experiment 7	0.79	0.17	0.74	8.73	8.30	0.43
Delayed Recognition	(.03)	(.03)	(.04)	(0.62)	(0.66)	(0.10)
Experiment 8	0.99	0.12	0.99	8.81	8.79	0.02
No Response Pressure	(.01)	(.03)	(.01)	(0.65)	(0.65)	(0.02)
Experiment 12	0.997	0.07	0.997	19.67	19.66	0.01
Subjective Cues	(.003)	(.02)	(.003)	(0.77)	(0.77)	(0.01)

Table 10. Mean hit and false alarm rates of produced words, *M* estimate, number of words recalled, and number of true recalls and correct intrusions as estimated with *M* in Experiments 9 through 11. Standard errors are shown in parentheses below means.

Condition	p("old")		*M*	Recall	True Recall	Correct Intrusions
	Old	New				
Free Recall						
Experiment 9	0.98	0.03	0.98	5.65	5.64	0.002
Related Pairs	(.01)	(.01)	(.01)	(0.41)	(0.41)	(0.002)
Experiment 10	0.94	0.04	0.94	5.03	5.01	0.01
Unrelated Pairs	(.02)	(.01)	(.02)	(0.39)	(0.39)	(0.01)
Cued Recall						
Experiment 9	0.95	0.08	0.95	13.60	13.53	0.07
Related Pairs	(.01)	(.01)	(.01)	(0.53)	(0.54)	(0.03)
Experiment 10	0.97	0.12	0.97	7.68	7.76	0.03
Unrelated Pairs	(.01)	(.02)	(.02)	(0.63)	(0.65)	(0.02)
Experiment 11	0.82	0.09	0.81	6.33	6.24	0.10
Weak-Cue Study/	(.04)	(.02)	(.04)	(0.46)	(0.47)	(0.03)
Strong-Cue Test						

than in free recall, in most cases the absolute level of correct intrusions is quite low. Hence, generally speaking, the previous statistical comparisons and conclusions for these experiments hold. It is interesting to note, however, that when correct intrusions are removed from recall scores, cued recall conditions that seemed to trend toward more recalls than free recall now show a much weaker trend (e.g., Experiments 3, 5, and 6). This fact supports the previous nonsignificant difference between the recall scores for these cued recall conditions against free recall, as well as the interpretation of Experiment 3, where it was suggested that presenting multiple cues at test increases the fluency of produced items, and this resulted in increased false alarm rates. If this explanation is correct, it further predicts that correct intrusions should be greater in Experiment 3 than in other experiments; this was indeed found.

Furthermore, correct intrusions are largely absent in free recall and, although they still are at relatively low levels in cued recall, they are much more frequent than in free recall. The fact that the model estimates that correct intrusions should be present occasionally in cued recall but not in free recall fits completely with the description offered here for why correct intrusions arise in the first place. Hence, there is some empirical support here for the notion that correct intrusions exist, and that they influence cued recall but not free recall.

Experiments 8 and 12 were the only cued recall conditions reported in Table 9 to show virtually no correct intrusions. This is consistent with the instructions that were given to subjects in Experiment 8, whereby subjects were instructed not to guess, which could be expected to reduce correct intrusions by forcing subjects to adopt a stricter recognition criterion. As well, in Experiment 12, in the face of highly effective subjective cues, there may have been little need or opportunity to guess at all. Similarly, examining Table 10, it is clear that when cues are encoding specific (as in Experiments 9 and 10), they produce little increase in correct intrusions.

Thus, when the influence of memory is strong, the number of observed correct intrusions drops significantly. Conversely, in Experiment 11, where cues were less able to cue the specific memory of studied items (as might be expected was the case in Experiments 1 through 8), correct intrusions increased once again.

Across all of the dissertation experiments, then, there does seem to be evidence that correct intrusions may be a real and measurable phenomenon. As well, correct intrusions in the model offered are related to both the number of actual intrusions observed and the strength of memory for the studied words. Correct intrusions should be more prevalent as false memories increase and/or as actual memory strength declines. In essence, correct intrusions appear not to be an issue when memory strength is strong. But when memory strength is weak, they contaminate measures of recall based on the propensity to make false alarms.

Of course, conceding that correct intrusions are real does not mean that they play a large role in biasing recall scores. Certainly, in all of the experiments here, correct intrusions are rare. Furthermore, correct intrusions were not measured directly but only estimated. Indeed, it may be that the only way to directly measure correct intrusions would be using neuroimaging. Using neuroimaging techniques, it may be possible to sort recalls based on whether they give rise to a signal that looks like others recalls or like intrusions. In fact, this is an avenue of future work that I am currently exploring.

In the absence of such neuroimaging data, however, to estimate correct intrusions we are forced to rely on the mathematical procedure. One limitation of the mathematical estimates of correct intrusions here is that it is unclear whether these equations provide accurate measures of correct intrusions as hit rates converge to 1. Examining Experiment 3, if hit rates were equal to 1, M would always be estimated as 1 as well, regardless of the level of non-study-based fluency

(i.e., *F*). Thus, it may be that although these equations provide some estimates of correct intrusions, they do not hold as hit rates approach 1.

By this fact alone, it could be argued that the estimates of correct intrusions were much lower in free recall than in cued recall because in free recall hit rates were closer to 1. Hence, perhaps the entire notion that correct intrusions exist and differ between free recall and cued recall may not be accurate. In the absence of further evidence for correct intrusions, all I can say here is that even with hit rates of .95 and .99, simple simulations demonstrated that estimates of correct intrusion rates will increase significantly as false alarms increase. Thus, the formulas for correct intrusions may still be able to measure relative differences between conditions even when hit rates are very high.

Furthermore, the main goal in the discussion of correct intrusions was simply to introduce this phenomenon as a potential area for future interest and consideration. As a result, these data have not been subjected to detailed statistical analyses nor can it be argued, based solely on these data, that strong evidence for correct intrusions exists. Instead, the data are suggestive with regard to the notion of correct intrusions, and demonstrate that there may be reliable methods for estimating their influence. Furthermore, although correct intrusions are likely relatively rare, the estimates suggest that they appear to be nearly completely absent in free recall and, thus, are really only relevant in cued recall.

Indeed, if the broader goal of recognized recall is to more precisely measure subjective memory during recall, development of more sophisticated methods of analyses may be a parallel goal. The present consideration of correct intrusions could be seen as a first step in such an attempt, or at the very least, as bringing to light this type of intrusion.

Recognized Recall and Generate-Recognize Models of Recall

Throughout this dissertation, I have focused on a generate-recognize model of recall, and indeed, have interpreted the recognized recall procedure and results in the context of this model. However, it is important to point out that although recognized recall and generate-recognize models have similarities (i.e., both discuss a generation/production followed by recognition), the recognized recall procedure and generate-recognize models of recall are not dependent on one another. Neither is recognized recall a procedure that is nested in the assumptions of generate-recognize models of recall. That is, although I have presented recognized recall in the context of generate-recognize models of recall, this procedure is actually atheoretical, in the sense that it is a behavioural technique that can be interpreted in virtually any theory of recall.

Recognized recall does have a close correspondence to generate-recognize models of recall in the sense that the recognized recall procedure asks subjects to produce studied and unstudied items and then to recognize those items whereas generate-recognize models of recall suggest that to recall words, subjects generate candidates and then, when they recognize a candidate word, they output it. Hence, both contain elements of "generation/production" and "recognition." However, although these concepts may seem shared between the procedure and the theory, they have different meanings.

In recognized recall, production refers to words that are output by subjects when they are asked to produce a studied or unstudied word. In generate-recognize models, generation refers to the covert act of mentally retrieving or re-creating items that were possibly studied. Hence, subjects may generate many covert items before finally deciding on an item to output. Recognized recall would only consider the final output to be a production.

Similarly, in recognized recall, recognition refers to a subject's classifications of produced items as "old" or "new," whereas in generate-recognize models, recognition refers to a subject noticing that a generated candidate was studied and consequently deciding to output it as a recall. In the generate-recognize model, recognition of items occurs implicitly for all recalls. In the recognized recall paradigm, recognition refers to the overt "old"/"new" decision made on produced items. Hence, in the recognized recall paradigm, recognition refers to an overt behavioural response whereas in generate-recognize models it refers to a covert recognition action that can *give rise* to an overt recognition action, but does not necessarily need to.

Thus, the recognized recall paradigm has been presented in the context of generate-recognize models and shares the elements of "generate/production" and "recognize," but this paradigm itself is not dependent on or nested in a generate-recognize framework. Indeed, even if the dominant theory of recall were to shift away from generate-recognize models, recognized recall still should be a more accurate paradigm for measuring recall performance than traditional recall paradigms. In essence, the recognized recall paradigm provides more precise behavioural measures than do traditional recall paradigms, regardless of the theoretical framework that one chooses to use to interpret the results.

Conclusion

Cueing can help memory retrieval when cues are strong and/or specific to the study phase. Especially when cues are self-generated during the study phase, cueing can be an excellent memory aid at test. However, unless cues are self-generated (i.e., subjective), any benefit of cueing is accompanied by increased false memories. In many cases, too, it can be shown that free recall is superior to cued recall, in the sense that cues do not provide a strong mnemonic benefit but do have a measurable cost in terms of increasing intrusions beyond free

recall. Hence, although cues can sometimes be useful, there is a consistent cost of false memories—except, of course, in the case of subjective cues.

Researchers have long known that cues can lead to the production of studied items that cannot be recognized (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1971, 1973), so in some sense the findings here are not surprising. However, what is surprising is that this is the first work to demonstrate that recognition failure of recallable words is not an isolated phenomenon that requires certain learning conditions. Instead, this phenomenon may be occurring in almost all cases of cueing. Therefore, this is the first work to demonstrate that the benefit of cued recall over free recall may be overestimated in traditional recall paradigms, due to the fact that unrecognizable "recalls" are being counted in cued recall.

Finally, the current work used the recognized recall paradigm, a paradigm designed to better equate free recall and cued recall in terms of response pressures and guessing. Given the paradigm's success in measuring subjective memory during recall and disambiguating guessing from actual memory in cued recall paradigms, it would seem advisable that more researchers measuring recall should adopt this paradigm over traditional recall paradigms. Indeed, researchers can still obtain traditional measures of recall using the recognized recall paradigm. The main benefit of recognized recall, however, is that it allows further differentiation of subjective memorability, so that conclusions can be drawn from data where guessing and response pressures have been controlled.

Significant improvements in our understanding of cognition may come about only once we realize the gaps in our existing knowledge. The development of the recognized recall procedure demonstrates that the exercise of developing new, more precise methods for studying even basic phenomenon can shift our understanding and illuminate new avenues of investigation.

Indeed, as this dissertation has shown, even our most unquestioned and basic empirical effects can be hiding new truths.

<u>References</u>

Anisfield, M. & Knapp, M. (1968). Association, synonymity, and directionality in false

    recognition. <u>Journal of Experimental Psychology, 77</u>, 171-179.

Allan, K., Doyle, M. C., & Rugg, M. D. (1996). An event-related potential study of word-stem

    cued recall. <u>Cognitive Brain Research, 4(4)</u>, 251-262.

Allan, K. & Rugg, M. D. (1998). Neural correlates of cued recall with and without retrieval of

    source memory. <u>Neuroreport: An International Journal for the Rapid Communication of

    Research in Neuroscience, 9(15)</u>, 3463-3466.

Anderson, J. R. & Bower, G. H. (1972). Recognition and retrieval processes in free recall.

    <u>Psychological Review, 79(2)</u>, 97-123.

Anderson, J. R. & Bower, G. H. (1974). A propositional theory of recognition memory.

    <u>Memory & Cognition, 2(3)</u>, 406-412.

Angel, L., Fay, S., Bouazzaoui, B., Granjon, L., & Isingrini, M. (2009). Neural correlates of

    cued recall in young and older adults: An event-related potential study. <u>NeuroReport:

    For Rapid Communication of Neuroscience Research, 20(1)</u>, 75-79.

Angel, L., Fay, S., Bouazzaoui, B., & Isingrini, M. (2010). Individual differences in executive

    functioning modulate age effects on the ERP correlates of retrieval success.

    <u>Neuropsychologia, 48(12)</u>, 3540-3553.

Angel, L., Isingrini, M., Bouazzaoui, B., Taconnat, L., Allan, K., Granjon, L., & Fay, S. (2010).

    The amount of retrieval support modulates age effects on episodic memory: Evidence

    from event-related potentials. <u>Brain Research, 1335</u>, 41-52.

Bahrick, H. P. (1969). Measurement of memory by prompted recall. <u>Journal of Experimental

    Psychology, 79(2, Pt.1)</u>, 213-219.

Bahrick, H. P. (1970). Two-phase model for prompted recall. Psychological Review, 77(3), 215-222.

Bahrick, H. P. (1971). Accessibility and availability of retrieval cues in the retention of a categorized list. Journal of Experimental Psychology, 89(1), 117-125.

Basden, D. R. & Basden, B. H. (1995). Some tests of the strategy disruption interpretation of part-list cuing inhibition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(6), 1656-1669.

Basden, B. H., Basden, D. R., Bryner, S., Thomas, R. L. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(5), 1176-1189.

Basden, B. H., Basden, D. R., & Henry, S. (2000). Costs and benefits of collaborative remembering. Applied Cognitive Psychology, 14(6), 497-507.

Basden, D. R., Basden, B. H., & Galloway, B. C. (1977). Inhibition with part-list cuing: Some tests of the item strength hypothesis. Journal of Experimental Psychology: Human Learning and Memory, 3(1), 100-108.

Bilodeau, E. A. (1967). Experiment interference with primary associates and their subsequent recovery with rest. Journal of Experimental Psychology, 73(3), 328-332.

Bilodeau, E. A. & Blick, K. A. (1965). Courses of misrecall over long-term retention intervals as related to strength of pre-experimental habits of word association. Psychological Reports, 16(3, Pt. 2), 1173-1192.

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(4), 657-668.

Bodner, G. E., Masson, M. E. J., & Caldwell, J. I. (2000). Evidence for a generate–recognize model of episodic influences on word-stem completion. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(2), 267-293.

Bower, G. H. (1970). Organizational factors in memory. Cognitive Psychology, 1(1), 18-46.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. Journal of Verbal Learning & Verbal Behavior, 8(3), 323-343.

Bregman, A. S. (1968). Forgetting curves with semantic, phonetic, graphic, and contiguity cues. Journal of Experimental Psychology, 78(4, Pt.1), 539-546.

Brown, R. & McNeill, D. (1966). The "tip of the tongue" phenomenon. Journal of Verbal Learning & Verbal Behavior, 5(4), 325-337.

Buschke, H. (1984). Cued recall in amnesia. Journal of Clinical Neuropsychology, 6(4), 433-440.

Buschke, H. & Lenon, M. (1969). Encoding homophones and synonyms for verbal discrimination and recognition. Psychonomic Science, 14, 269-270.

Clark, H. H. & Stafford, R. A. (1969). Memory for semantic features in the verb. Journal of Experimental Psychology, 80(2, Pt.1), 326-334.

Cohen, B. H. (1966). Some-or-none characteristics of coding behavior. Journal of Verbal Learning & Verbal Behavior, 5(2), 182-187.

Collins, A. M. & Quillan, 1969, M. R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8(2), 240-247.

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. Philosophical Transactions of the Royal Society, London, Series B, 302, 341-359.

Craik, F. I. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning & Verbal Behavior, 11(6), 671-684.

Craik, F. I. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 104(3), 268-294.

Dalla Barba, G., Parlato, V., Iavarone, A., & Boller, F. (1995). Anosognosia, intrusions and 'frontal' functions in Alzheimer's disease and depression. Neuropsychologia, 33(2), 247-259.

Dalla Barba, G., & Wong, C. (1995). Encoding specificity and intrusion in Alzheimer's disease and amnesia. Brain and Cognition, 27(1), 1-16.

Deweer, B., Lehéricy, S., Pillon, B., & Baulac, M. (1995). Memory disorders in probable Alzheimer's disease: The role of hippocampal atrophy as shown with MRI. Journal of Neurology, Neurosurgery & Psychiatry, 58(5), 590-597.

Dong, T. (1972). Cued partial recall of categorized words. Journal of Experimental Psychology, 93(1), 123-129.

Dong, T. & Kintsch, W. (1972). Subjective retrieval cues in free recall. Journal of Verbal Learning & Verbal Behavior, 7(4), 813-816.

Earhard, M. (1967). Cued recall and free recall as a function of the number of items per cue. Journal of Verbal Learning & Verbal Behavior, 6(2), 257-263.

Fay, S., Isingrini, M., Ragot, R., & Pouthas, V. (2005). The effect of encoding manipulation on word-stem cued recall: An event-related potential study. Cognitive Brain Research, 24(3), 615-626.

Fox, P. W., Blick, K. A., & Bilodeau, E. A. (1964). Stimulation and prediction of verbal recall and misrecall. Journal of Experimental Psychology, 68(3), 321-322.

Fox, P. W. & Dahl, P. R. (1971). Aided retrieval of previously unrecalled information. Journal of Experimental Psychology, 88(3), 349-353.

Fox, L. S., Olin, J. T., Erblich, J., Ippen, C. G., & Schneider, L. S. (1998). Severity of cognitive impairment in Alzheimer's disease affects list learning using the California verbal learning test (CVLT). International Journal of Geriatric Psychiatry, 13, 544-549.

Freund, J. S. & Underwood, B. J. (1970). Restricted associates as cues in free recall. Journal of Verbal Learning & Verbal Behavior, 9(1), 136-141.

Fuld, P. A., Katzman, R., Davies, P., & Terry, R. D. (1982). Intrusions as a sign of Alzheimer dementia: Chemical and pathological verification. Annals of Neurology, 11, 155-159.

Funkhouser, G. R. (1968). Effects of differential encoding on recall. Journal of Verbal Learning and Verbal Behavior, 7, 1016-1023.

Gardiner, J. M. (1988). Recognition failures and free-recall failures: Implications for the relation between recall and recognition. Memory & Cognition, 16(5), 446-451.

Greene, R. L. (2004). Recognition memory for pseudowords. Journal of Memory and Language, 50, 259-267.

Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(4), 691-702.

Hollingworth, H. L. (1928). Psychology: Its facts and principles. New York: Appleton.

Hoppe, R. A. (1962). Memorizing by individuals and groups: A test of the pooling-of-ability model. The Journal of Abnormal and Social Psychology, 65(1), 64-67.

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.

Hudson, R. L. & Austin, J. B. (1970). Effect of context and category name on the recall of categorized word lists. Journal of Experimental Psychology, 86(1), 43-47.

Humphreys, M. S. & Galbraith, R. C. (1975). Forward and backward associations in cued recall: Predictions from the encoding specificity principle. Journal of Experimental Psychology: Human Learning and Memory, 1(6), 702-710.

Isaac, C. L. & Mayes, A. R. (1999a). Rate of forgetting in amnesia: I. Recall and recognition of prose. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(4), 942-962.

Isaac, C. L. & Mayes, A. R. (1999b). Rate of forgetting in amnesia: II. Recall and recognition of word lists at different levels of organization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(4), 963-977.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. Journal of Memory and Language, 30(5), 513-541.

Jacoby, L. L. & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. Journal of Memory and Language, 29(4), 433-454.

Jacoby, L. L. & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. Journal of Experimental Psychology: General, 118(2), 126-135.

Joordens, S., Ozubko, J. D., & Niewiadomski, M. W. (2008). Featuring old/new recognition: The two faces of the pseudoword effect. Journal of Memory and Language, 58(2), 380-392.

Kahana, M. J. (1996). Associate retrieval processes in free recall. <u>Memory & Cognition, 24(1)</u>, 103-109.

Kapur, N. & Coughlan, A. K. (1980). Confabulation and frontal lobe dysfunction. <u>Journal of Neurology, Neurosurgery, and Psychiatry, 43</u>, 461-463.

Katz, J. J. & Fodor, J. A. (1963). The structure of a semantic theory. <u>Language, 39</u>, 170-210.

Kern, R. S., Van Gorp, W. G., Cummings, J. L., Brown, W. S., & Osato, S. S. (1992). Confabulation in Alzheimer's disease. <u>Brain and Cognition, 19</u>, 172-182.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), <u>Models of human memory</u>. New York: Academic Press.

Kintsch, W. (1974). <u>The representation of meaning in memory</u>. Oxford, England: Lawrence Erlbaum.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. <u>Annual Review of Psychology, 51</u>, 481-537.

Lauer, P. A. & Battig, W. F. (1972). Free recall of taxonomically and alphabetically organized word lists as a function of storage and retrieval cues. <u>Journal of Verbal Learning & Verbal Behavior, 11(3)</u>, 333-342.

Lewis, M. Q. (1971). Categorized lists and cued recall. <u>Journal of Experimental Psychology, 87(1)</u>, 129-131.

Lewis, M. Q. (1974). Cue effectiveness in cued recall. <u>Journal of Experimental Psychology, 102(4)</u>, 737-739.

Light, L. L. (1969). <u>Effects of pretraining and cueing on recall and recognition</u>. Unpublished doctoral dissertation, Stanford University, 1969.

Light, L. L. (1972). Homonyms and synonyms as retrieval cues. Journal of Experimental Psychology, 96(2), 255-262.

Light, L. L. & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. Journal of Verbal Learning & Verbal Behavior, 9(1), 1-11.

Loess, H. & Harris, R. (1968). Short-term memory for individual verbal items as a function of method of recall. Journal of Experimental Psychology, 78(1), 64-69.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. Journal of Experimental Psychology: Human Learning and Memory, 4(1), 19-31.

Loftus, E. F. & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. Journal of Verbal Learning & Verbal Behavior, 13(5), 585-589.

Lowenstein, D. A., Habib, R., & Tulving, E. (1998). Hippocampal PET activations of memory encoding and retrieval: The HIPER model. Hippocampus, 8, 313-322.

Luck, S. J. (2005). An introduction to the Event-Related Potential Technique. Cambridge, MA: The MIT Press.

Manning, S. K., Greenhut-Wertz, J., & Mackell, J. A. (1996). Intrusions in Alzheimer's disease in immediate and delayed memory as a function of presentation modality. Experimental Aging Research, 22(4), 343-361.

Martin, E. (1975). Generation—recognition theory and the encoding specificity principle. Psychological Review, 82, 150–153.

McDaniel, M. A. & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. Psychonomic Bulletin & Review, 15(2), 237-255.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory.

    Journal of Verbal Learning and Verbal Behavior, 2, 1-21.

Meudell, P. R., Hitch, G. J., & Boyle, M. M. (1995). Collaboration in recall: Do pairs of people

    cross-cue each other to produce new memories? The Quarterly Journal of Experimental

    Psychology A: Human Experimental Psychology, 48A(1), 141-152.

Meudell, P. R., Hitch, G. J., & Kirby, P. (1992). Are two heads better than one? Experimental

    investigations of the social facilitation of memory. Applied Cognitive Psychology.

    Special Issue: Memory in everyday settings, 6(6), 525-543.

Mondani, M. S., Pellegrino, J. W., & Battig, W. F. (1973). Free and cued recall as a function of

    different levels of word processing. Journal of Experimental Psychology, 101(2), 324-

    329.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer

    appropriate processing. Journal of Verbal Learning & Verbal Behavior, 16(5), 519-533.

Müller, G. E. (1913). Zur Analyze der Gedächtnistätigkeit und des Vorstellungsverlaufe s. III.

    Teil. Zeitschrift fur Psychologie, Ergänzungsband 8.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free

    association, rhyme, and word fragment norms. Behavior Research Methods, Instruments

    & Computers. Special Issue: Web-based archive of norms, stimuli, and data: Part 1,

    36(3), 402-407.

Nickerson, R. S. (1984). Retrieval inhibition from part-set cuing: A persisting enigma in

    memory research. Memory & Cognition, 12(6), 531-552.

Nilsson, L. & Gardiner, J. M. (1993). Identifying exceptions in a database of recognition failure

    studies from 1973 to 1992. Memory & Cognition, 21(3), 397-410.

Nobel, P. A. & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall.

   Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(2), 384-413.

Norman, D. A. (1968). Toward a theory of memory and attention. Psychological Review, 75(6),

   522-536.

Ozubko, J. D. & Joordens, S. (2007). The mixed truth about frequency effects on free recall:

   Effects of study list composition. Psychonomic Bulletin & Review, 14(5), 871-876.

Ozubko, J. D. & Joordens, S. (2008). Super Memory Bros: Going from mirror patterns to

   concordant patterns via similarity enhancements. Memory & Cognition, 36(8), 1391-

   1402.

Ozubko, J. D. & Joordens, S. (2011). The similarities (and familiarities) of pseudowords and

   extremely high frequency words: Examining a familiarity-based explanation of the

   pseudoword effect. Journal of Experimental Psychology: Learning, Memory, and

   Cognition, 37(1), 123-139.

Ozubko, J. D. & Yonelinas, A. P. (under review). A familiar finding: Evidence for a familiarity-

   based account of the pseudoword effect. Journal of Memory and Language.

Paivio, A. (1969). Mental imagery in associative learning and memory. Psychological Review,

   76(3), 241-263.

Paivio, A. (1971). Imagery and deep structure in the recall of English nominalizations. Journal

   of Verbal Learning & Verbal Behavior, 10(1), 1-12.

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and why?

   Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(5), 1196-

   1204.

Perlmutter, M. (1979). Age differences in adults' free recall, cued recall, and recognition.

    Journal of Gerontology, 34(4), 533-539.

Perlmutter, H. V. & De Montmollin, G. (1952). Group learning of nonsense syllables. The

    Journal of Abnormal and Social Psychology, 47(4), 762-769.

Perry, A. R. & Wingfield, A. (1994). Contextual encoding by young and elderly adults as

    revealed by cued and free recall. Aging & Cognition, 1(2), 120-139.

Postman, L., Adams, P. A., & Phillips, L. W. (1955). Studies in incidental learning II. The

    effects of association value and of the method of testing. Journal of Experimental

    Psychology, 49(1), 1-10.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past.

    Memory & Cognition, Vol 21(1), 89-102.

Ramponi, C., Richardson-Klavehn, A., & Gardiner, J. M. (2007). Component processes of

    conceptual priming and associative cued recall: The roles of preexisting representation

    and depth of processing. Journal of Experimental Psychology: Learning, Memory, and

    Cognition, 33(5), 843-862.

Reder, L. M., Anderson, J. R., & Bjork, R. A. (1974). A semantic interpretation of encoding

    specificity. Journal of Experimental Psychology, 102(4), 648-656.

Roediger, H. L. (1973). Inhibition in recall from cueing with recall targets. Journal of Verbal

    Learning & Verbal Behavior, 12(6), 644-657.

Roediger, H. L. (1974). Inhibiting effects of recall. Memory & Cognition, 2(2), 261-269.

Roediger, H. L. & Payne, D. G. (1982). Hypermnesia: The role of repeated testing. Journal of

    Experimental Psychology: Learning, Memory, and Cognition, 8(1), 66-72.

Roediger, H. L., Wheeler, M. A., & Rajaram, S. (1993). Remembering, knowing, and

    reconstructing the past. In Douglas L. Medin (Ed.), <u>The psychology of learning and</u>

    <u>motivation.</u> San Diego, CA: Academic Press.

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998).

    Dissociation of the neural correlates of implicit and explicit memory. <u>Nature, 392(6676),</u>

    595-598.

Santa, J. L. & Lamwers, L. L. (1974). Encoding specificity: Fact or artifact. <u>Journal of Verbal</u>

    <u>Learning & Verbal Behavior, 13(4)</u>, 412-423.

Santa, J. L. & Lamwers, L. L. (1976). Where does the confusion lie? Comments on the

    Wiseman and Tulving paper. <u>Journal of Verbal Learning & Verbal Behavior, 15(1)</u>, 53-

    57.

Schott, B., Richardson-Klavehn, A., Heinze, H., & Düzel, E. (2002). Perceptual priming versus

    explicit memory: <u>Dissociable neural correlates at encoding. Journal of Cognitive</u>

    <u>Neuroscience, 14(4)</u>, 578-592.

Shiffrin, R. M. & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory.

    <u>Psychological Review, 76(2)</u>, 179-193.

Slamecka, N. J. (1972). The question of associative growth in the learning of categorized

    material. <u>Journal of Verbal Learning & Verbal Behavior, 11(3)</u>, 324-332.

Slamecka, N. J. (1968). An examination of trace storage in free recall. <u>Journal of Experimental</u>

    <u>Psychology, 76(4, Pt.1)</u>, 504-513.

Slamecka, N. J. (1969). Testing for associative storage in multitrial free recall. <u>Journal of</u>

    <u>Experimental Psychology, 81(3)</u>, 557-560.

Slamecka, N. J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. <u>Journal of Experimental Psychology: Human Learning and Memory, 4(6)</u>, 592-604.

Stephenson, G. M., Abrams, D., Wagner, W., & Wade, Gillian (1986). Partners in recall: Collaborative order in the recall of a police interrogation. <u>British Journal of Social Psychology, 25(4)</u>, 341-343.

Taconnat, L., Froger, C., Sacher, M., & Isingrini, M. (2008). Generation and associative encoding in young and old adults: The effect of the strength of association between cues and targets on a cued recall task. <u>Experimental Psychology, 55(1)</u>, 23-30.

Thomson, D. M. & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. <u>Journal of Experimental Psychology, 86(2)</u>, 255-262.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. <u>Journal of Verbal Learning & Verbal Behavior, 6(2)</u>, 175-184.

Tulving, E. (1985). Memory and consciousness. <u>Canadian Psychology, 26</u>, 1-12.

Tulving, E. & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. <u>Journal of Experimental Psychology, 77(4)</u>, 593-601.

Tulving, E. & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. <u>Journal of Verbal Learning & Verbal Behavior, 5(4)</u>, 381-391.

Tulving, E. & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. <u>Journal of Experimental Psychology, 87(1)</u>, 1-8.

Tulving, E. & Thomson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. <u>Journal of Experimental Psychology, 87(1)</u>, 116-124.

Tulving, E. & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80(5), 352-373.

Underwood, B. J. (1972). Are we overloaging memory? In A. W. Melton & E. Martin, (Eds.), Coding processes in human memory. Washington, D. C.: Winston.

Watkins, M. J. & Tulving, E. (1975). Episodic memory: When recognition fails. Journal of Experimental Psychology: General, 104(1), 5-29.

Weiss, E. (1967). Stimulus category cue and list difficulty as determinants of the amount of transfer. Journal of Experimental Psychology, 73(3), 446-449.

Weldon, M. S. & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(5), 1160-1175.

Wheeler, M. A. & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. Psychological Science, 3(4), 240-245.

Whittlesea, B. W. A. & Williams, L. D. (2001a). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(1), 3-13.

Whittlesea, B. W. A. & Williams, L. D. (2001b). The discrepancy-attribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(1), 14-33.

Winograd, E. & Conn, C. P. (1971). Evidence from recognition memory for specific encoding of unmodified homographs. Journal of Verbal Learning & Verbal Behavior, 10(6), 702-706.

Wood, G. (1967). Category names as cues for the recall of category instances. Psychonomic Science, 323-324.

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.

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.

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. Consciousness and Cognition: An International Journal, 5(4), 418-441.

Yuker, H. E. (1955). Group atmosphere and memory. The Journal of Abnormal and Social Psychology, 51(1), 17-23.

Appendix. Summary of *t*-values, degrees of freedom, *p*-values, and Cohen's *d* values for comparisons of correct intrusions between Conditions A and B.

Condition A	Condition B	*t*	*df*	*p*	*d*
Exp 1. Cued Recall	Exp 1. Free Recall	3.71	84	< .01	1.04
Exp 3. 8 Cues/Trial	Exp 1. Free Recall	4.12	51	< .01	1.17
Exp 4. Strong Associate Cues	Exp 1. Free Recall	3.05	44	< .01	0.86
Exp 5. Status-Identified Cues	Exp 1. Free Recall	3.01	43	< .01	0.85
Exp 6. Studied Word Cues Only	Exp 1. Free Recall	3.84	49	< .01	1.09
Exp 7. Delayed Cued Recall	Exp 7. Delayed Free Recall	4.90	59	< .01	1.17
Exp 8. No Pressure Cued Recall	Exp 1. Free Recall	0.15	54	.88	0.04
Exp 9. Related Pairs Cued Recall	Exp 9. Related Pairs Free Recall	1.95	67	.06	0.47
Exp 10. Unrelated Pairs Cued Recall	Exp 10. Unrelated Pairs Free Recall	0.96	65	.34	0.23
Exp 11. Weak-Cue Study/ Strong-Cue Test Cued Recall	Exp 10. Unrelated Pairs Free Recall	3.43	59	< .01	0.84
Exp 12. Subjective Cues	Exp 12. Free Recall	1.00	28	.33	0.36