Title
Social Contagion of Correct and Incorrect Information in Memory

Author
Rush, Ryan Allen

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Social Contagion of Correct and Incorrect Information in Memory

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Ryan Allen Rush

August 2013

Dissertation Committee:
Dr. Steven E. Clark, Chairperson
Dr. David C. Funder
Dr. Robert Rosenthal

The Dissertation of Ryan Allen Rush is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

I would like to thank my parents, Christian and Susan Rush. You both taught me the value of education and hard work. I am forever grateful for your guidance, encouragement, and endless patience. Without both your love and support, I would not be where I am today.

I would like to thank my advisor, Steve Clark, for the many years of outstanding mentorship. Your guidance and support allowed me to grow both as a researcher and a teacher. I cannot thank you enough for teaching me that persistence is the key to a successful academic career. Because of you, I am leaving the University of California, Riverside confident in my abilities and ready to take on the next stage of my life.

I would also like to thank my dissertation committee members, David Funder and Bob Rosenthal. Without your insightful and useful feedback, my dissertation would not have come to fruition.

And lastly, I would like to thank my best friends in graduate school, Ryne Sherman and James Dias. Thank you for providing the support I often needed while in graduate school. Whether we were discussing research or just talking about life in general, the countless hours spent with both of you kept me sane and changed my life forever. I am grateful that I was fortunate enough to share this moment in time with the two of you and hope to maintain a long and prosperous friendship with you both.

ABSTRACT OF THE DISSERTATION

Social Contagion of Correct and Incorrect Information in Memory

by

Ryan Allen Rush

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, August 2013
Dr. Steven E. Clark, Chairperson

Collaborative memory research has focused on the negative effects of group remembering, specifically emphasizing how collaborative memory can be worse than individual memory. Previous research has shown that collaboration can impair memory by limiting group output through retrieval disruption (Basden, Basden, Bryner & Thomas, 1997), by altering one's memory through socially induced forgetting (Coman, Manier, & Hirst, 2009), and through the social transmission of contagious errors (Roediger, Meade, & Bergman, 2000). The current dissertation includes three experiments designed to examine the effect that discussion has on subsequent individual memory reports. Experiment 1 systematically examines the transmission and acceptance of correct versus incorrect information within the specific context of a social contagion memory paradigm developed by Roediger et al. (2000). Experiment 2 examines the differential effects of social contagion when the overall amount of recalled information varies across cued and non-cued recall tasks (Tulving & Pearlstone, 1966). Experiment 3 examines how reconstructive memory processes may produce schema-consistent memory errors in recall, within the context of a social contagion memory paradigm using Deese,

Roediger, and McDermott (DRM) stimuli (Deese, 1959; Roediger & McDermott, 1995). Consistent across all three experiments, pairs of participants recalled items from a set of stimulus materials, discussed their recall with each other, and then recalled the items again individually. The current research provides strong evidence that there is more to collaboration than just the transmission of errors. Overall, participants were exposed to more correct than incorrect information during discussion, even though exposure information was less accurate than each participant's initial recall. Participants incorporated more correct than incorrect exposure items into a later memory report, suggesting that people can distinguish correct information from incorrect information. However, there was little to no change in accuracy following discussion. Within each pair of participants who discussed their initial recall, post-discussion recall accuracy increased from initial to final recall for the initially less accurate participant and decreased for the initially more accurate participant. This suggests that during discussion there may be a redistribution of accuracy between participants.

Table of Contents

List of Figures	x
List of Tables	xii
List of Appendices	xiii
Chapter 1 – Introduction	1
Current Research	10
Chapter 2 – Experiment 1	
Introduction	12
Predictions	16
Methods	17
Results and Discussion	20
Summary of Results	21
Discussion Pairs	23
No-Discussion Pairs	25
Discussion Pairs Compared to No-Discussion Pairs	25
Accuracy Change Within Pairs	26
Chapter 3 – Experiment 2	
Introduction	30
Predictions	32
Development of Stimulus Materials	33
Methods	35
Results and Discussion	39

Summary of Results 42

Manipulation Check: Comparing Cued to Non-Cued Recall 44

Cued Recall: Discussion Pairs 45

Cued Recall: No-Discussion Pairs 46

Cued Recall: Discussion Pairs Compared to
No-Discussion Pairs 47

Accuracy Change Within Cued Recall Pairs 48

Non-Cued Recall: Discussion Pairs 50

Non-Cued Recall: No-Discussion Pairs 52

Non-Cued Recall: Discussion Pairs Compared to
No-Discussion Pairs 52

Accuracy Change Within Non-Cued Recall Pairs 52

Cued Recall Compared to Non-Cued Recall 55

Chapter 4 – Experiment 3

Introduction 62

Predictions 67

Methods 68

Results and Discussion 72

Summary of Results 75

Critical Item Not Presented: Discussion Pairs 77

Critical Item Not Presented: No-Discussion Pairs 79

Critical Item Not Presented: Discussion Pairs Compared to
No-Discussion Pairs 80
Accuracy Change Within Pairs: Critical Item Not Presented 81
Critical Item Presented: Discussion Pairs 84
Critical Item Presented: No-Discussion Pairs 85
Critical Item Presented: Discussion Pairs Compared to
No-Discussion Pairs 86
Accuracy Change Within Pairs: Critical Item Not Presented 86
Critical Item Not Presented Compared to
Critical Item Presented 90
Chapter 5 – General Discussion 96
Current State of Collaborative Memory Research 96
Summary of Experiment Results 97
Implications of the Current Research 99
Future Direction 102
References 105
Appendices 113

List of Figures

Figure #	Title	Page
2-1	General methodology for Experiments 1-3	14
2-2	Mean number of correct items recalled, incorrect items recalled, and recall accuracy during initial and final recall for discussion and no-discussion pairs	23
2-3	Mean initial and final accuracy for the better and worse participant in the discussion pairs	27
2-4	Mean initial and final accuracy for the better and worse participant in the no-discussion pairs	27
3-1	Mean number of correct items recalled, incorrect items recalled, and recall accuracy during initial and final recall for discussion and no-discussion pairs in the *cued recall* condition	44
3-2	Mean number of correct items recalled, incorrect items recalled, and recall accuracy during initial and final recall for discussion and no-discussion pairs in the *non-cued recall* condition	44
3-3	Mean initial and final accuracy for the better and worse discussion participants in the *cued recall* condition	49
3-4	Mean initial and final accuracy for the better and worse no-discussion participants in the *cued recall* condition	49
3-5	Mean initial and final accuracy for the better and worse discussion participants in the *non-cued recall* condition	54
3-6	Mean initial and final accuracy for the better and worse no-discussion participants in the *non-cued recall* condition	54
4-1	Mean number of correct items recalled, incorrect items recalled, and recall accuracy during initial and final recall for discussion and no-discussion pairs on trials when the *critical item was not* presented on the wordlist	77

4-2 Mean number of correct items recalled, incorrect items recalled, and recall accuracy during initial and final recall for discussion and no-discussion pairs on trials when the *critical item was* presented on the wordlist 77

4-3 Mean initial and final accuracy for the better and worse discussion participants when the *critical item was not* presented on the wordlist 82

4-4 Mean initial and final accuracy for the better and worse no- discussion participants when the *critical item was not* presented on the wordlist 82

4-5 Mean initial and final accuracy for the better and worse discussion participants when the *critical item was* presented on the wordlist 88

4-6 Mean initial and final accuracy for the better and worse no- discussion participants when the *critical item was* presented on the wordlist 88

List of Tables

Table #	Title	Page
2-1	Mean Number of Items Recalled and Accuracy for Initial Recall, Exposure Items, Recall of Exposure Items, and Final Recall for Discussion and No-Discussion Pairs	21
3-1	Number of Items Recalled and Accuracy for Initial Recall, Exposure Items, Recall of Exposure Items, and Final Recall for Discussion and No-Discussion Pairs in the *Cued Recall* Condition	41
3-2	Number of Items Recalled and Accuracy for Initial Recall, Exposure Items, Recall of Exposure Items, and Final Recall for Discussion and No-Discussion Pairs in the *Non-Cued Recall* Condition	41
4-1	Number of Items Recalled and Accuracy for Initial Recall, Exposure Items, Recall of Exposure Items, and Final Recall for Discussion and No-Discussion Trials when the *Critical Item was Not Presented* on the Word List	74
4-2	Number of Items Recalled and Accuracy for Initial Recall, Exposure Items, Recall of Exposure Items, and Final Recall for Discussion and No-Discussion Trials when the *Critical Item was Presented* on the Word List	74

List of Appendices

Appendix	Title	Page
A	Stimulus Materials for Experiment 1: Roediger et al. (2000) Social Contagion Stimuli	113
B	Experiment 1: Statistical Analyses Conducted on the Pair Averages	114
C	Experiment 1:Test-Retest Reliability Between Initial and Final Recall	115
D	Stimulus Materials for Experiment 2: Category Names and List Items 1 to 4	116
E	Experiment 2: Statistical Analyses Conducted on the Pair Averages	117
F	Experiment 2:Test-Retest Reliability Between Initial and Final Recall	118
G	Stimulus Materials for Experiment 3: Critical Item and List Items 1 to 15	119
H	Experiment 3: Statistical Analyses Conducted on the Pair Averages	120
I	Experiment 3:Test-Retest Reliability Between Initial and Final Recall	121

Chapter 1 - Introduction

We experience and recall many events in our lives in socially rich contexts with the people around us. We even share the events we experience by ourselves with others. We share these memories to reminisce about the past, build relationships, provide others with personal information, and to answer questions about what occurred during an event. When individuals collaborate, or share memories for events, there is opportunity to transmit information, as well as to alter and/or maintain the consistency of event details stored in memory.

Collaborative memory research has predominantly focused on the negative effects of group remembering, specifically how collaborative (social) memory can be worse than individual memory. For example, when Person A and Person B work together to remember the details of an event several things can go wrong: 1.) Person A may fail to access and share details stored in memory because hearing Person B's recall disrupts this process (Basden, Basden, Bryner & Thomas, 1997); 2.) Person A may forget previously remembered details that were not shared with Person B during discussion (Coman, Manier, & Hirst, 2009); 3.) Person A may include incorrect details into later memory reports after hearing them from Person B (Roediger, Meade, & Bergman, 2000); or 4.) because there was pressure to conform to the incorrect information provided by Person B (Reysen, 2005; 2007). The current review of collaborative memory research will begin by briefly discussing general experimental procedures, followed by a detailed discussion of how research has shown that collaborative remembering can be both harmful and beneficial to memory, and concludes with an outline of the current experimental studies.

When people collaborate to remember an event, they must work together to recall the details from a previous experience. This can happen in a variety of ways, for example each person can exhaustively tell their side of the story, group members can interrupt others when they feel they have something to add or a correction to make, or group members can take turns recalling the details of the experienced event. In collaborative memory experiments, the structure of collaborative recall is often a turn taking procedure, where each group member takes a turn recalling one item and all group members report information before anyone provides a second piece of information. The measure of collaborative group output is the total number of items produced by all group members collectively, whereas individual recall is simply the total number of items reported by a single participant working alone. Collaborative group recall is typically compared with the recall of individuals, as well as with the recall of a *nominal group*. A nominal group consists of the same number of individuals as in the collaborative group, but they do not recall together. In other words if a collaborative group includes the recall from two participants working together, a nominal comparison group will include the total recall of two individuals who have worked alone. A nominal group is therefore an artificially manufactured group where no social interaction occurred between group members. The total output for a nominal group is determined by combining the recalls from two or more independent individuals, counting overlapping items only once. Thus, the comparison between collaborative groups and nominal groups is a comparison between the total output of individuals who work together versus the total output of individuals who work alone.

Research on group memory, going back over 50 years (Perlmutter & de Montmollin, 1952), shows that groups collectively recall more correct information than the group's individual members, simply because the individuals recall non-overlapping sets of items (see also more recently, Basden, Basden, Bryner & Thomas, 1997; Blumen & Rajaram, 2008; Maki, Weingold & Arellano, 2008; Vollrath, Sheppard, Hinsz, & Davis, 1989; Weldon & Bellinger, 1997).

One might predict that collaborative groups should recall more information than nominal groups. In the course of the collaboration, individuals can cross-cue each other. That is, the items recalled by one person during a collaborative effort may serve as a retrieval cue for others' unrecalled information (Meudell, Hitch, & Kirby, 1992; Meudell, Hitch, & Boyle, 1995). Thus, Person A may recall some item that serves as a memory cue for Person B. However, when the comparison is between collaborative groups and nominal groups, increased recall due to cross cuing does not generally occur. In fact, collaborative groups recall significantly *less* information than nominal groups (Blumen & Rajaram, 2008; Weldon & Bellinger, 1997). Weldon and Bellinger called this phenomenon *collaborative inhibition*, because while collaborative groups outperform individual members, they recall less than nominal groups. In other words, *n* people working together (i.e., collaborating) tend to recall less than *n* people working alone.

Several possibilities have been hypothesized for why collaborative inhibition occurs. First, it seems reasonable to assume that when individuals work together in a collaborative group they may feel less personal responsibility or motivation to contribute. Therefore, group members may engage in *social loafing* by not exhibiting their best abilities (Latané, Williams, & Harkins, 1979). This may occur because the individuals working within the group experience a diffusion of responsibility, because individuals want to maintain equality of effort and not work harder than others, or because individuals do not feel personally accountable, as their personal contributions will not be identified (Karau & Williams, 1993). Johansson, Andersson, and Rönnberg (2000) also found that group cohesion is important. Knowing the other group member well may result in less anxiety about the collaborative task and allow partners to utilize effective strategies for remembering information. Participants who are not familiar with others in the social situation may experience increased anxiety for making mistakes; resulting in fewer contributions than if the person is working alone.

Weldon, Blair, and Huebsch (2000) examined how motivation can affect collaborative recall in a series of studies where they manipulated motivation in a variety of ways, for example (1) offering a monetary incentive to the group that produced the most correct information, (2) setting a minimum level of recall for the group, (3) increasing personal accountability, and (4) increasing group cohesiveness. In all cases, even when motivation to recall was high, collaborating groups recalled significantly less than nominal groups, suggesting that impaired group recall was not due to social loafing. The results suggest that in the context of a small group where members are not anonymous there are other processes, beyond motivational factors, that lead to an impairment of collaborative group performance (Weldon et al., 2000).

An alternative cognitive explanation proposed by Basden et al. (1997) suggests that collaborative inhibition occurs because of a disruption of individual retrieval strategies during collaboration. Basden et al. explained that each individual develops his or her own unique organizational structure in memory for the studied material. Disruption occurs during collaboration to the extent that others' recall is misaligned with the individual's own retrieval strategy and memory organization. In other words, individuals working together will recall fewer items from memory when working in groups and recalling information aloud because hearing another person's recall will disrupt their own personal retrieval strategy.

Basden et al. (1997) found support for the retrieval disruption hypothesis by having participants study one of two lists containing the same number of words. Lists were comprised of either fifteen 6-item categories or six 15-item categories. Basden et al. suggested that when participants study more lists with fewer items per list there would be little retrieval disruption because group members are likely to share similar organization and retrieval strategies. On the other hand, when participants study fewer lists with more items per list, there is a higher probability that item organization in memory will be less consistent across individual group members. Their results showed a reduction in collaborative inhibition relative to nominal group controls when participants collaboratively recalled 6-item lists, but not when participants collaboratively recalled 15-item lists. These results suggest that collaborative inhibition occurs to the extent that the information is not consistently organized across individuals. Therefore, a reduction in retrieval disruption and collaborative inhibition occurs when all participants are able to organize the to-be-remembered information the same way. A reduction in collaborative inhibition can also occur in other ways. For example providing a cue during recall makes reliance on personal retrieval strategies less important and reduces the effects of retrieval disruption (Finlay, Hitch, & Meudell, 2000). Additionally, during recognition tasks, collaboration facilitates memory when one group member makes a compelling argument for why he or she is correct and the other group members are incorrect (Clark, Hori, Putnam, & Martin, 2000). Group size also plays an important role in collaborative inhibition. The larger a group is the more opportunity there is for retrieval disruption during discussion (Thorley & Dewhurst, 2007).

While collaboration may lead to an inhibition of performance during group recall, results suggest that when individuals later recall on their own, sometimes the effects of retrieval disruption disappear and other times the effects of retrieval disruption persist beyond collaboration. Finlay et al. (2000) found that when participants performed three free recall tasks, first individually, then collaboratively, and then again individually, they showed increased memory performance between collaborative recall and final individual recall, such that they report information that was not reported during the collaborative effort (Finlay et al., 2000). This finding suggests that the effects of collaborative inhibition are temporary, because information initially accessible during an individual recall is lost during a subsequent collaborative recall, but becomes accessible again during a final individual recall task when there is no longer mutual disruption of retrieval strategies. When collaborative inhibition results from a cognitive disruption during group recall the effects seem to be temporary and disappear once the cognitive disruption is removed. Coman, Manier, & Hirst (2009) however, suggest a different pattern of results. Coman et al. used a questionnaire to initially probe individual participants about the details of their day on September 11, 2001. Following the questionnaire, participants recounted their personal experiences on September 11, 2001 with another participant. Finally, participants performed an individual recognition memory test for their original responses to the questionnaire. The results showed that when participants did not recall previously known information during collaboration, the information often remained absent or forgotten during a recognition task performed alone. Coman et al. called this phenomenon *socially induced forgetting*. Coman et al.'s results suggest that when collaborative inhibition is the result of social influence, rather than a cognitive disruption, people may exhibit prolonged effects of memory inhibition beyond the collaborative effort.

Collaboration not only leads to decreased output when compared to nominal groups, but can also reduce accuracy of the individuals' memory through the incorporation of others' errant responses into ones' own memory. Roediger et al. (2001) have called this phenomenon the *social contagion of memory*, and have shown that during collaboration where a participant experiences exposure to highly probable erroneous information, from his or her partner, he or she is likely to incorporate that information into a subsequent individual memory report.

Similarly, social pressure to conform produces deficits in individual memory performance (Reysen, 2005; 2007). When participants performed a group recognition task for a list of words with a virtual confederate, conforming to the confederate's responses was associated with a decrease in the participants' ability to later correctly recognize items during an individual recognition task (Reysen, 2005). Additionally, Reysen (2007) showed that when the implicit social pressure to say "old" to a previously unseen item is increased by adding additional virtual confederates, the likelihood that the item will incorrectly later be identified by the participant as old goes up. This suggests that the pressure to conform during collaboration, beyond mere exposure to incorrect information, increases later individual memory errors.

Taken together these results suggest that collaboration can impair memory by limiting group output through retrieval disruption (Basden et al., 1997), by altering one's memory through socially induced forgetting (Comen et al., 2009), through the social transmission of contagious errors (Roediger et al, 2000), and through group pressures to conform (Reysen, 2005; 2007). However, collaboration does not only involve the spread of errors and decreased output. We collaborate on a regular basis because it provides an opportunity to re-experience the information and possibly pick up details that we failed to properly encode and store in memory.

One benefit to collaboration is that it serves as an opportunity for re-exposure to forgotten information. That is, when group member A recalls an item that was forgotten by group member B the item serves as a recognition cue and provides an additional opportunity to further encode and store that piece of information in memory. Weldon and Bellinger (1997) found that re-exposure to information during a collaborative recall attempt enhanced later individual recall. Blumen and Rajaram (2008) found similar results suggesting that a final individual recall benefits from re-exposure during repeated group recalls (e.g. collaborative recall, followed by collaborative recall, followed by individual recall), as well as when initial individual recall is followed by a collaborative recall (e.g. individual recall, followed by collaborative recall, followed by individual recall). A social contagion paradigm would predict a similar pattern of results (Roediger et al., 2000). It is unlikely that during a collaborative task an individual would be exposed to only incorrect information from his or her partner; therefore social contagion should not only be thought of as a means to spread errors amongst group members. To the extent that individuals are exposed to correct information from their collaborating partner, they may show improvement in their post-collaborative recall (Roediger et al., 2000; Meade & Roediger, 2002). In fact, Roediger et al. noted that when participant exposure was to only correct items from a confederate, participants later recalled 43% of those correct items during an individual recall. These results suggest that re-exposure to information during collaboration can lead to net gains in memory.

In addition, collaboration can lead to memory benefits through the process of error pruning. When participants are given feedback from other group members during the discussion task they can edit erroneous information from memory, leading to a reduction in errors in later memory reports (Ross, Spencer, Blatz, & Restorick, 2008). Ross et al. found that both older and younger adult spouses benefited from collaboration. Older adult couples benefited in two ways; during collaboration older couples tended to inhibit the production of errors (e.g. specifically, items they recalled with low confidence) and when errors were expressed during collaboration older adults corrected mistakes made by their partner. Younger adults in contrast only benefited from the correction of errors produced during collaboration. Results also showed that the majority of rejected items were incorrect and that the rejection of an incorrect item followed a statement of reservation from the participant's partner 40% of the time.

It is important to note however, that this benefit may only occur when collaboration involves a free-flowing discussion. As mentioned before, collaborative memory research often involves participants taking turns reporting information to one another (Roediger at al., 2001; Basden et al., 1997). Under this very structured collaborative exchange, there is no opportunity for error correction. Only when participants have the opportunity to express reservations, correct, or challenge other group members' memories can error pruning occur. These results taken together suggest that collaboration can have a positive impact particularly for post-collaborative individual recall. Under the right conditions, collaboration affords the opportunity of re-exposure and error correction, which depending on the nature of the information one is exposed to during collaboration, might lead to an increase in memory accuracy.

Current Research

The current studies examine how people share, maintain, and alter their memories for events during a social collaborative memory task. Experiment 1 systematically examines the transmission and acceptance of correct versus incorrect information within the specific context of a social contagion memory paradigm developed by Roediger et al. (2001). Experiment 2 examines the differential effects of social contagion when the overall output, or amount of recalled information, varies using a cued recall (high correct and incorrect output) versus non-cued recall (low correct and incorrect output) task (Tulving & Pearlstone, 1966). And finally Experiment 3 examines how reconstructive memory processes may produce schema-consistent memory errors in recall, within the context of a social contagion memory paradigm using Deese, Roediger, and McDermott (DRM) stimuli: a list of highly associated words where one critical associate has been removed from the list (Deese, 1959; Roediger & McDermott, 1995; Stadler, Roediger, & McDermott, 1999). The current studies extend the existing social contagion literature by overcoming three current limitations. First, current research reflects an overemphasis on the transmission of error. Second, current research relies heavily on the use of confederates. Third, current research has emphasized empirical phenomena, with little theoretical guidance or development. Thus, there is little discussion regarding the mechanisms by which memories "travel" from one person to another. By addressing these limitations the current research contributes to the collaborative memory literature by examining the extent to which people are exposed to and incorporate both correct and incorrect information under different circumstances, and advances the field by providing a better theoretical framework for understanding why the social transmission of information occurs between individuals.

Chapter 2 – Experiment 1

Introduction

Experiment 1 systematically examined the transmission and acceptance of correct versus incorrect information within the specific context of a social contagion memory paradigm. Roediger et al. (2001) developed the social contagion paradigm in a study that presented participants with six slides, each depicting a common but cluttered household scene, such as a closet, a bedroom, or a bathroom. After the presentation of all six slides, participants recalled the items shown in the six scenes, together with another person. However, the other person was not another participant (unbeknownst to the one actual participant), but was a confederate of the experimenter. As the participant and confederate took turns recalling items from each slide, the carefully scripted confederate recalled two incorrect items for three of the slides. The results showed that the confederate's incorrect information was contagious; participants, when later asked to individually recall items from the slides, reported incorrect information to which the confederate had exposed them.

This basic result is robust and reliable across experimental variations (Meade & Roediger, 2002), and is consistent with results from a large number of studies showing memory conformity effects (e.g., Allan & Gabbert, 2008; Gabbert, Memon, & Allan, 2003; Gabbert, Memon, & Wright, 2006; Reysen, 2005, 2007; Schneider & Watkins, 1996; Wright, Mathews, & Skagerberg, 2005; Wright, Self, & Justice, 2000).

However, there was another aspect of the Roediger et al. results, mentioned in the last paragraph of their paper, as a caveat to their main conclusions. They noted that for three of the slides the confederate recalled only correct and no incorrect items, and that participants later recalled 43% of those correct items. These results, however, are difficult to interpret for two reasons. First, the 43% recall, which is about 8 of the 18 correct items recalled by the confederate, may include items that the participant would have recalled had the confederate not recalled them first. Because the confederate and participant took turns, each recalling one item at a time, it is impossible to know which items came exclusively from the confederate. Second, the experimenter determined the proportion of correct and incorrect items recalled by the confederate. The methodology used by Roediger et al. limits the interpretation of the results. However, the results suggest that obtaining information from another person might actually be beneficial.

In all of the current experiments, the social contagion paradigm was modified to not use a confederate during group recall, and to include an initial individual recall to determine which recall items were unique to each participant. The three current experiments all follow the same basic methodology depicted in Figure 2-1.

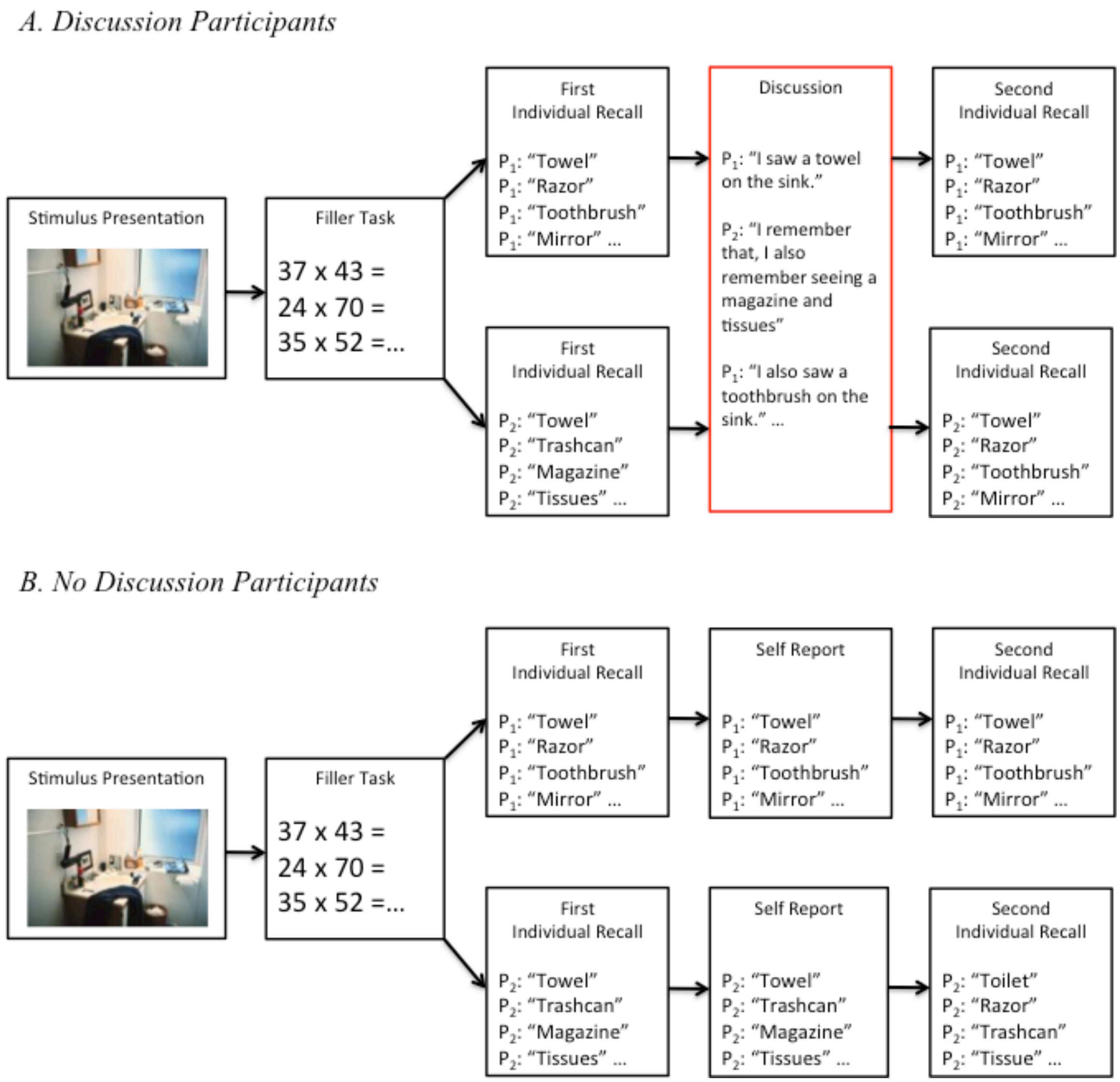

Figure 2-1. General Methodology for Experiments 1-3.

In the present experiment, pairs of participants viewed the same slides used by Roediger et al. Within each pair, participants first recalled items from the slides alone (similar to the procedure used by Chan, Thomas, and Bulevich, 2009), to establish which recalled items "belonged" to whom. Participants in the discussion condition then exchanged notes about what each had recalled, and finally, they recalled the six slides again, individually. Rather than discussing with another person, participants in the no-discussion control condition, engaged in a self-report task, where they reported their initial recall to a recording device before performing a final individual recall. Participants in the no-discussion group performed a self-report task rather than solve additional math equations to equate the number of opportunities for re-exposure to initial recall between groups. This ensures that any differences observed between the discussion and no-discussion group are a result of the social interaction, rather than the product of self-cuing resulting from re-exposure to initial recall.

Importantly, the present study does not utilize a confederate, for reasons outlined by Clark, Abbe, and Larson (2006). The use of a confederate has the advantage that researchers can precisely specify and control the information to which a participant is exposed. The disadvantage, however, is that there is perhaps too much control over the information. To the extent that the confederate follows instructions, the experimenter has control over the proportion of correct and incorrect information to which the participant is exposed. The consequences of exposure depend, of course, on what one is exposed to, and thus the net gain or loss in accuracy may depend in large part on decisions made not by the participants, but rather by the experimenter. Koriat (2012) has shown that collaboration leads to error for test items that are likely to elicit incorrect individual responses, and leads to an increase in accuracy for test items that are likely to elicit correct individual responses. Such results suggest that the net gain or loss in memory accuracy due to an interaction with another person will depend largely on the accuracy of that other person’s information.

Predictions

The key data concern the accuracy of the exposure information and the proportions of correct and incorrect exposure items incorporated into each person's subsequent recall. *Exposure items* are, by definition, items initially recalled by one member of the pair but not the other. Application of a *truth-in-numbers heuristic*, and previous results by Clark et al. (2006), suggest that items recalled by two people are more likely to be correct than items recalled by only one person. The other side of this, of course, is that items recalled by only one person, but not the other, are less likely to be correct. This suggests that the exposure items should be less accurate, on average, than each individual's initial recall.

The second question concerns the incorporation rates of correct and incorrect exposure items. Two factors suggest that correct exposure items will be recalled at a higher rate than incorrect exposure items. These two factors are embodied in dual-process theories of recognition memory (Jacoby, 1991; Yonelinas, 1994; Wixted & Mickes, 2010). The key assumption is that a person will incorporate an exposure item into his or her recall to the extent that the item is recognized as having been in the scene. Dual-process theories of recognition memory suggest that items are recognized when the information can be recollected or when it is accompanied by a strong sense of familiarity. During discussion, recollection occurs when an exposure item provides the individual with access to the previously stored memory. Thus, an individual will incorporate an exposure item if the item provides a cue for recall ("Oh, yes, now I do remember – there *was* a tea kettle sitting on the stove; it was black with a teak handle"). Even without such recollection, an individual may incorporate the exposure item because it seems familiar. During a discussion, recognition occurs due to familiarity simply because the exposure item provides the individual with a strong feeling that the item was previously experienced and belongs to the stimulus set. Importantly, when familiarity leads to recognition it does not cue an existing memory or provide access to the details about that item. To the extent that the person has information about the item in memory, recollection and familiarity should *both* be higher for correct exposure items than for incorrect exposure items, and thus correct exposure items should be incorporated into participants' post-discussion recall at a higher rate than incorrect exposure items.

Methods

Participants

Two hundred (79 male, 121 female) University of California, Riverside undergraduates participated in pairs, as partial fulfillment of a requirement for introductory psychology.

Materials

The stimulus materials were six photographs of common but cluttered household scenes, developed by Roediger et al. (2001), and used by several others (e.g. Allan & Gabbert, 2008; Meade & Roediger, 2002; Ross et al., 2008). The six scenes depicted a *toolbox, bathroom, kitchen, bedroom, closet, and desk* (See Appendix A). The number of observable items per scene ranged from 22 to 32 items with an average of 26.17 items.

Procedure

Participants came to the laboratory in pairs and were seated at individual computer stations. Participants were instructed that they would be viewing a PowerPoint presentation consisting of several photographs, and were told to pay close attention to each scene because they would be tested later for their memory of items presented in the scenes. When both participants were ready they were instructed to press the spacebar on the computer's keyboard to begin the presentation. The presentation was timed so participants were not required to press any additional keys. Each scene was presented for 15 seconds along with a descriptive title (e.g., "TOOLBOX", "BATHROOM", etc.), and following the methodology of Roediger et al. (2001), the scenes were presented in the same order for all participants. Following the presentation of all six slides participants were given a filler task (4 min of solving multiplication problems), and then were asked to recall the six scenes, one scene at a time.

The first recall was performed individually. Participants were given a packet of response sheets to record their written recall. The title of a specific scene was shown at the top of each response sheet. Recall for the scenes was performed in the same order as scene presentation during the study phase. Participants were given two minutes per scene to individually recall as many items as possible. After individual recall of all six scenes, half of the participant pairs were asked to discuss their recall with each other (discussion pairs), while the other half was asked to report their initial recall to a microphone alone (no-discussion pairs).

Discussion pairs were instructed to share their initial recall with their partner. Participants were told that they were to mention every item that they had written down during the previous individual recall. Participants were given their response sheets to aid in this process. Response sheets were placed on a table in front of the discussing pairs so that both participants could see what they and their partner had previously written down. This was done to prevent participants from editing the information they shared with their partner. Participants were also encouraged to engage in conversation about the scene rather than simply list the items they had written down or read their partner's recall. Participants were given as much time as they needed to discuss all six scenes, and spent an average of 1.20 min (*SD* = 0.48 min) discussing each scene.

Each participant in the no-discussion pairs went to a separate room and read his or her list of initially recalled items into a microphone rather than discussing it with a partner. Participants in the no-discussion pairs were given as much time as they needed to report items from each scene. Each participant was provided with his or her response sheets and was instructed to mention each item previously written down. Participants in the no-discussion group spent an average of 0.49 min (*SD* = 0.17 min) reporting items for each scene.

It is important to note that individual participants in both the discussion and no-discussion groups were equated in that they all verbally reported their initial recall, either to another person (discussion group), or to a microphone (no-discussion group). As a result, the opportunity for re-exposure to one's own recall was controlled for between groups.

After either discussing their individual recalls for all six scenes, or reading their individual recalls into a microphone for all six scenes, participants' initial response sheets were collected. All participants performed a second individual recall task. Participants were given a new set of response sheets identical to the one used during initial recall. Again, participants were instructed to write down as many items as they could remember from each scene. Participants were given two minutes per scene to individually recall as many items as possible and they performed this task individually for each scene in the same order as they were originally presented.

Results and Discussion

The data were scored using Roediger et al.'s (2001) listing of the items for each scene, with a few modifications.[1] I examined three measures of recall: the number of correct items recalled, the number of incorrect items recalled, and recall accuracy. Following Koriat and Goldsmith (1994), recall accuracy was calculated as the proportion of recalled items that were correct (Σ [Correct_i / (Correct_i + Incorrect_i)] / N). These data are shown in Table 2-1, for initial recall, exposure items, recall of exposure items, and final recall. To examine the effects of discussion, the difference scores between Recall 2 and Recall 1 (R2 – R1) were calculated for correct item recall, incorrect item recall, and recall accuracy and compared between the discussion and no-discussion groups. The results section begins with a brief summary of the main findings followed by an in depth discussion of the statistical details. The analyses are presented as planned comparisons to

[1] These modifications were minor. For example, whereas Roediger et al. (2001) listed the single item "wrench" for the toolbox slide, there were actually three kinds of wrenches in the slide, a distinction that was reflected in participants' recall. Thank you to Michelle Meade for providing the listing of items used by Roediger et al.

assess the key questions regarding the accuracy of initial recall, exposure, incorporation, and final recall, for discussion and no-discussion groups.

Table 2-1.

Mean Number of Items Recalled and Accuracy for Initial Recall, Exposure Items, Recall of Exposure Items, and Final Recall for Discussion and No-Discussion Pairs.

	Correct	Incorrect	Accuracy
	Discussion Pairs (N_{Pairs} = 50; N = 100)		
Initial Recall	4.67 (1.40)	1.34 (1.05)	.79 (.12)
Exposure Items	2.75 (1.11)	1.20 (1.02)	.71 (.16)
Recall of Exposure Items	1.30 (0.57)	0.32 (0.28)	.80 (.16)
Final Recall	6.07 (1.51)	1.65 (0.83)	.79 (.08)
	No Discussion Pairs (N_{Pairs} = 50; N = 100)		
Initial Recall	4.44 (1.30)	1.45 (0.86)	.76 (.12)
Final Recall	4.63 (1.37)	1.51 (0.88)	.76 (.11)

Note. Accuracy = Σ [$Correct_i$ / ($Correct_i$ + $Incorrect_i$)] / N; Standard deviations in parentheses. Calculating accuracy for each individual weights each participant's accuracy score by the magnitude of his or her recall. As a result, accuracy cannot be calculated using the mean correct and incorrect recall values provided in the table.

Summary of Results

The results of Experiment 1 may be summarized as follows: When people discussed their recall with each other, the additional information provided by each person, i.e., the exposure information, included far more correct information than incorrect information. Nonetheless, these exposure items were less accurate than individuals' initial recall. People were more likely to incorporate the correct exposure information than the incorrect exposure information. In other words, people were able to accurately distinguish between correct and incorrect exposure information such that correct information was more contagious than incorrect information. These results suggest that the lower accuracy of the exposure information, combined with the higher accuracy for the incorporation of exposure information, produced a near zero net gain in the accuracy of final recall compared to initial recall. Within pairs recall accuracy increased from initial to final recall for the initially less accurate participant and decreased for the initially more accurate participant, for both discussion and no-discussion pairs, with the effect being larger for discussion pairs. We refer to this result as an *accuracy redistribution effect*. Finally, discussion pairs showed a greater increase in the number of correct and incorrect items recalled during a second individual recall than no-discussion pairs, but there was virtually no difference in accuracy change between the two groups. This suggests that discussion played a greater role in gains of overall output, than self-cuing from re-exposure to one's initial recall, but played little role in overall accuracy change.

The following is a detailed report of the statistical analyses from the current experiment. Results are first presented for the discussion pairs, followed by the no-discussion pairs, and then finally as a comparison between the discussion and no-discussion pairs. Figure 2-2 highlights the changes for correct recall, incorrect recall, and accuracy between initial and final recall for both discussion and no-discussion pairs.[2]

[2] The analyses presented were conducted on the individual subjects (N = 100), which may be problematic in meeting the independence assumption. The data were also analyzed by averaging across each pair (N = 50) to meet the independence assumption. Those results can be found in Appendix B.

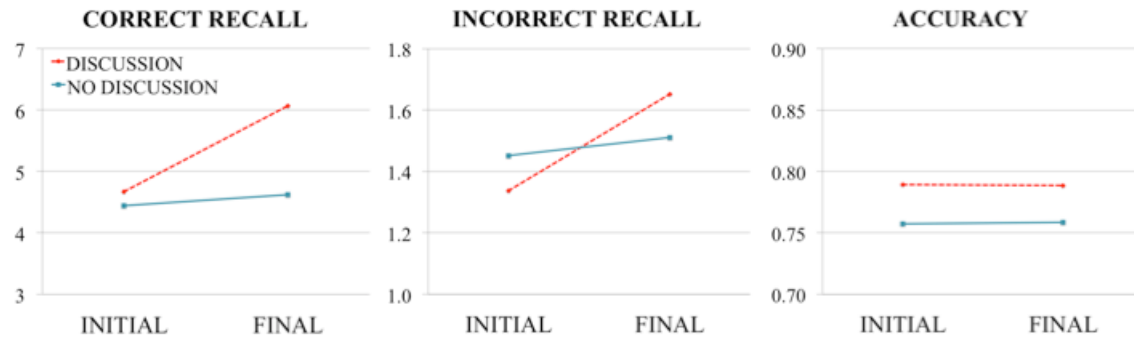

Figure 2-2. Mean number of correct items recalled, incorrect items recalled, and recall accuracy during initial and final recall for discussion and no-discussion pairs.

Discussion Pairs

Initial Recall and Exposure. During initial recall individuals recalled an average of 4.67 correct items and 1.34 incorrect items, with an overall mean accuracy of .79. When people discussed their recall with each other, the additional information provided by each person, i.e., the exposure information, included more correct items (2.75) than incorrect items (1.20), t (99) = 10.77, $p < .001$, $r = .73$. Nonetheless, the accuracy of exposure items (.71) was lower than the accuracy of participants' initial recall (.79), t (99) = -3.28, $p = .001$, $r = .31$. These results support the prediction that exposure accuracy would be lower than initial accuracy. Consistent with the truth in numbers heuristic, these results suggest that the items recalled by one participant but not the other, the exposure items, are less accurate than each individual's initial recall, which includes overlapping accurate information.

Incorporation of Exposure Items. Participants incorporated a greater proportion (R_I / R_E)[3] of correct exposure items (.49) than incorrect exposure items (.30) into their final recall, *t* (98) = 6.85, $p < .001$, *r* = .57.[4] In other words, people were able to accurately distinguish between correct and incorrect exposure information such that correct information was more contagious than incorrect information. The accuracy of the incorporated items (.80) was higher than the accuracy of the exposure items (.71), *t* (99) = 7.25, $p < .001$, *r* = .59, and only slightly higher than the accuracy of participants' initial recall (.79), *t* (99) = 0.66, *p* = .508, *r* = .07. Consistent with a dual process theory of recognition memory, these results suggest that correct items are more likely to become contagious during a discussion because information is more likely to be stored in memory for correct than incorrect items. Therefore, correct exposure items are more likely to serve as a retrieval cue or be highly familiar than incorrect information.

Change Across Recalls. The net effect of discussion was assessed by comparing initial to final recall. Participants recalled significantly more correct items, *t* (99) = 18.97, $p < .0001$, *r* = .89, and incorrect items, *t* (99) = 4.82, $p < .001$, *r* = .44, in their final recall compared to their initial recall. The accuracy of final recall (.7886) was nearly identical

---

[3] The proportion of incorporated exposure items was calculated as the number of incorporated items divided by the total number of exposure items (R_I / R_E). This calculation is dependent on exposure to at least one item and does not represent the measure of accuracy (Σ [$Correct_i$ / ($Correct_i$ + $Incorrect_i$)] / N) used in the current analyses.

[4] The analysis of the proportion of exposure items incorporated requires that a participant be exposed to at least one correct and one incorrect item during discussion. As a result, the analysis was preformed on a subset of the sample, and the proportions presented in the text cannot be calculated using the mean values of exposure and incorporation presented in Table 2-1.

to the initial recall (.7891), t (99) = -.07, p = .945, r = -.007.[5] These results suggest that the lower accuracy of the exposure information, combined with the higher accuracy for the incorporation of exposure information, produced a near zero net gain in the accuracy of final recall compared to initial recall.

No Discussion Pairs

To separate the effects of self-cuing, final and initial recall were compared for the no-discussion group. This comparison showed a statistically significant increase in the number of correct items recalled, t (99) = 3.91, p < .001, r = .37, and a non-significant increase in the number of incorrect items recalled, t (99) = 1.29, p = .199, r = .13, for final recall relative to initial recall. These results are consistent with a large literature on reminiscence and hypermnesia that suggests total output increases across multiple recall opportunities (Erdelyi & Becker, 1974; McDaniel, Moore, & Whitman, 1998; Roediger & Payne, 1982). The increases in correct and incorrect recall combined to produce equivalent levels of accuracy (.76 for initial and final recall), t (99) = .23, p = .819, r = .02. This suggests that participants working alone add a proportionally equivalent number of correct and incorrect items to their second recall.

Discussion Pairs Compared to No-Discussion Pairs

To compare the effect of discussion relative to self-cuing, correct item, incorrect item, and accuracy difference scores were compared between the discussion and no-discussion groups. Discussion pairs showed significantly larger increases for both correct item recall, t (198) = 13.89, p < .001, r = .70, and incorrect item recall, t (198) = 3.233, p

---

[5] Test-retest reliabilities for correct recall, incorrect recall, and accuracy between initial recall and final recall are found in Appendix C for discussion and no discussion pairs.

= .007, *r* = .22, during their final recall, relative to the no-discussion group. These results suggest that the effect of discussion is significantly greater than self-cuing memory through re-exposure to one's own information, and cannot be explained by the phenomenon of hypermnesia. There was virtually no difference in the change in accuracy between initial and final recall for the discussion and no-discussion groups, t (198) = -.182, p = .856, r = -.01. The "null" result that accuracy did not change across recall attempts should not be taken to mean that there were *no* changes in accuracy from recall 1 to recall 2. Changes in accuracy may occur within pairs resulting in the appearance that accuracy remained constant across recalls.

Accuracy Change Within Pairs

Within each pair of participants, one participant came into the discussion with a higher initial accuracy. Additional analyses compared the "better" or initially more-accurate participant to the "poorer" or initially less-accurate participant. Figure 2-3 and Figure 2-4 show the initial and final mean accuracy for the better and worse pair members in the discussion group and no-discussion group respectively.

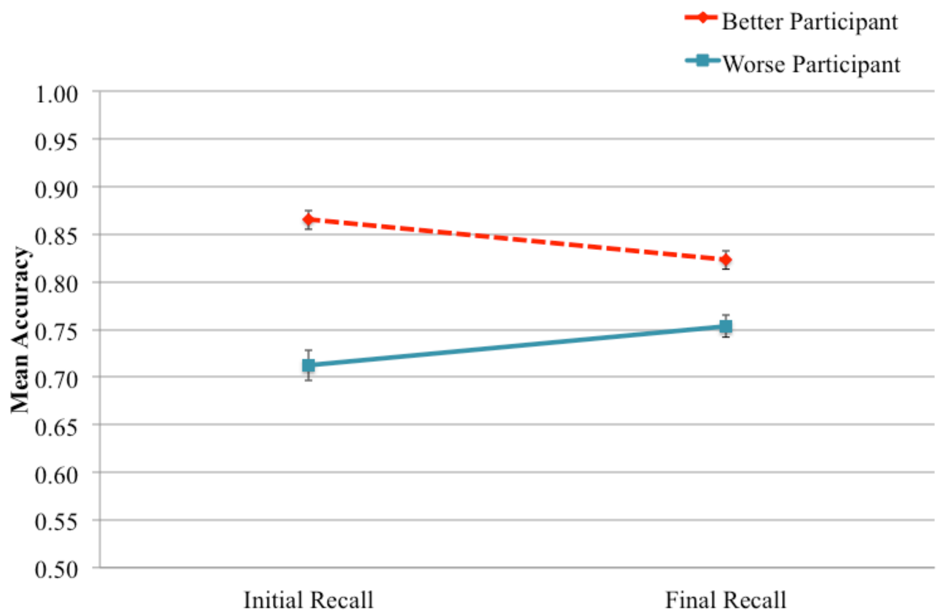

Figure 2-3. Mean initial and final accuracy for the better and worse participant in the discussion pairs.

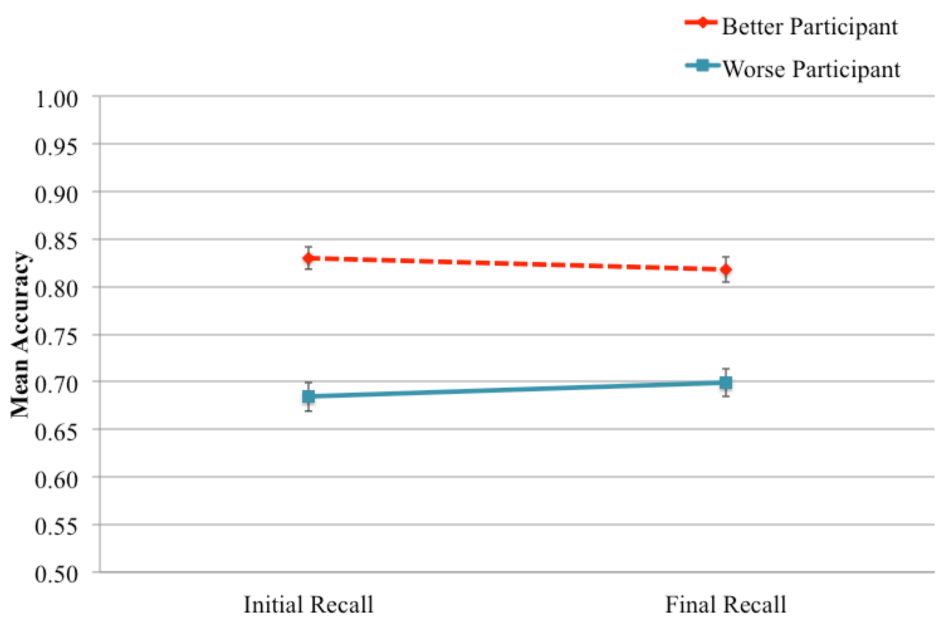

Figure 2-4. Mean initial and final accuracy for the better and worse participant in the no-discussion pairs.

From Figure 2-3 it is clear that following discussion the person who was more accurate for initial recall showed a small decline in accuracy from initial to final recall ($M = -.042$), whereas the less accurate participant showed a small increase in accuracy from initial to final recall ($M = .041$), $t(49) = 6.11$, $p < .001$, $r = .66$. In other words, the initially less accurate person became more accurate and the initially more accurate person became less accurate as a result of the interaction. I refer to this result as an *accuracy redistribution effect*. Although the gap in accuracy between the initially more accurate and less accurate participant decreased following discussion, the initially more accurate participant remained more accurate than the initially less accurate participant, $t(49) = 6.76$, $p < .001$, $r = .69$. These results suggest that the accuracy redistribution within the pair may be responsible for the apparent "null" result regarding accuracy change from initial to final recall. This occurs because the more- and less-accurate participants show proportionally equivalent decreases and increases in accuracy.

The same comparison for no-discussion pairs showed a similar but much smaller redistribution pattern with a decrease in accuracy for the initially more accurate participant ($M = -.013$) and an increase in accuracy for the initially less accurate participant ($M = .015$), $t(49) = 3.18$, $p = .003$, $r = .41$ (See Figure 2-4). For the no-discussion group, the initially more accurate participant remained more accurate than the initially less accurate participant, $t(49) = 7.58$, $p < .001$, $r = .73$. Even when not discussing their recall the initially less accurate person became more accurate and the initially more accurate person became less accurate. The redistribution effect for no-discussion groups suggests that some of the accuracy redistribution effect may arise from simply reporting one's recall, whether it is to another person or into a microphone, or could be the result of statistical regression to the mean. However, an independent samples comparison showed that the patterns of gains and losses were significantly greater for discussing pairs compared to non-discussing pairs, t (98) = 3.48, $p < .001$, $r = .33$,[6] suggesting that the accuracy redistribution effect is exacerbated by the interaction between individuals and the information they were exposed to during that discussion.

The results from Experiment 1 provide evidence that collaboration results in more than just error swapping between discussants. The current results show that both correct and incorrect information becomes contagious following exposure during discussion, and that correct information is more contagious than incorrect information. However, the increase in both correct and incorrect information following discussion produced no change in post-discussion accuracy. The effect that discussion has on post-discussion accuracy is likely to be a result of the items exposed to a person during discussion. Exactly how changes in the accuracy of initial recall influence post-discussion recall and accuracy remain to be seen. Experiment 2 examines this within a social contagion paradigm by manipulating the overall amount of correct and incorrect information people recall using cued recall and non-cued recall tests.

[6] This independent samples *t*-test was computed using difference scores calculated by subtracting the initially less accurate participant's accuracy change from the initially more accurate participant's accuracy change for the discussion and no discussion pairs.

Chapter 3 – Experiment 2

Introduction

In any collaborative memory experiment, the unique information that each participant brings to the discussion directly influences the post-discussion changes to memory. Therefore, the consequences of exposure depend on what one is exposed to, and thus the proportions of correct and incorrect items that are introduced in the course of the discussion are a critical factor. When exposure consists of predominantly correct information, the person is likely to include more items that are correct during a later memory report and may show an increase in overall accuracy. However, as the levels of incorrect exposure increase, the person is likely to include more items that are incorrect into subsequent recall and as a result may show a decline in overall accuracy. It is therefore important to consider how the overall output of initial recall influences memory following a discussion.

Experiment 2 was designed to examine the differential effects of social contagion when the overall output, or amount of recalled information, is manipulated using a cued recall versus non-cued recall task. The net gain or loss in accuracy during a collaborative memory task depends on the particular blend of correct and incorrect information to which a person is exposed, and can be manipulated by the type of free recall task given to participants.

Tulving and Pearlstone (1966) examined this by giving participants a categorized list of 48 words that contained 12 categories and 4 words per category. During the study phase participants were given a category name followed by four category members presented one at a time, this process was repeated until participants were shown all 12 categories and all 48 words. Following encoding, participants performed a free recall test. Half of the participants was given a cued recall test where they were provided with the category names as cues for recall and the other half of participants was given a non-cued recall test where they were not given the category names. These category cues produced an increase in both correct recall of list words, but also an increase in false recall of non-list words.

This result suggests that retrieval cues (i.e. the category names) may sometimes lead to the activation and inclusion of schematically or categorically related incorrect information. In other words, people may use this schematically or thematically consistent information to reconstruct memories, resulting in the occasional production of incorrect items (Brewer & Treyens, 1981).

The current study borrows the basic methodology from Tulving and Pearlstone (1966) to establish initial levels of recall within the social contagion experimental paradigm. The current study uses lists of words as stimulus materials rather than the Roediger et al. (2000) household pictures from Experiment 1 for two reasons. First, the Roediger et al. pictures are outdated and not always clear. Some of the items presented in the photographs are out-of-date (e.g. a portable Walkman) and likely to be unfamiliar to current college students. Additionally some regions of the different photographs are dark or washed out making it hard to clearly identify all of the presented items. Secondly, using complex picture stimuli allows the participant to freely gaze at items throughout the scene leaving an open question as to which items were attended to and for how long during the stimulus presentation. The use of word lists is advantageous because the experimenter can control the items presented to the participant and how long each word is visible. The same basic social contagion methodology from Experiment 1 was utilized in the present experiment. As with the previous study, Experiment 2 does not utilize a confederate.

Predictions

As in Experiment 1 the key data concern the accuracy of the exposure information and the proportions of correct and incorrect exposure items incorporated into each person's subsequent recall. Participants who perform a cued recall task should initially produce more correct information (words from the list) and more false information (words not from the list) than those performing a non-cued recall task. The truth-in-numbers heuristic suggests that two participants are less likely to produce the same incorrect items than correct items during a cued recall task. To the extent that this pattern occurs for participants performing a cued recall, there should be more errors exposed and spread between participants performing a cued recall than a non-cued recall.

On the other hand, participants performing a non-cued recall are more likely to make omission errors than those performing a cued recall task because they do not recall all of the correct information they have stored in memory. To the extent that participants in the non-cued recall condition report unique correct items there should be greater exposure and acceptance of correct information than incorrect information. In other words because participants in the non-cued recall condition recall initially fewer correct items than those in the cued recall condition they should have more opportunity for exposure and inclusion of correct items from their partner.

This suggests that participants who perform a cued recall will initially produce more overall output than those performing a non-cued recall, and that this output will include more intrusion errors. During a subsequent discussion participants in a cued recall condition are more likely to be exposed to incorrect information than those in a non-cued recall, possibly resulting in a greater spread of error and a decrease in overall accuracy. During discussion participants in the non-cued recall condition have the opportunity to be exposed to a greater number of correct items and incorporate those items into subsequent recall, possibly resulting in an increase in accuracy. To the extent that the described pattern of results is observed participants in the non-cued recall condition should show greater increases in overall accuracy following discussion relative to those in the cued recall condition.

Development of Stimulus Materials

In the current experiment the key manipulation involved the amount of over all output produced by each participant during a pre-discussion individual recall. Pilot testing was used to develop the stimulus materials to ensure that more overall output occurs during a cued recall task and less overall output occurs during a non-cued recall task. More importantly, it was necessary that the stimulus materials produced both more correct and incorrect items during a cued recall task than a non-cued recall task. This important feature allows for the examination of the effects of social contagion during increased (cued recall) and decreased (non-cued recall) exposure to both correct and incorrect information during discussion.

Participants

Forty-nine (20 male, 29 female) University of California, Riverside undergraduates participated in partial fulfillment of a requirement for introductory psychology.

Materials

The stimulus materials included categorized word lists organized into twelve categories with four words per category (48-items). Items were selected from the updated Battig and Montague (1969) category norms (Van Overschelde, Rawson, & Dunlosky, 2004), which represent lists of words ranked by their associative strength to a specific category. Van Overschelde et al. determined the associative strength of items within a given category by asking at least 600 participants to provide a list of items belonging to each category. The more frequently an item was reported for a specific category, the higher it was ranked on associative strength to that category. In the current experiment categories included *four-footed animals, kitchen utensils, fruits, furniture, sports, natural earth formations, articles of clothing, instruments, birds, vegetables, insects, and carpenter's tools*. Two sets of word lists were used during pilot testing. Following Tulving and Pearlstone (1966), the first set of word lists included the second, fourth, sixth, and eighth ranking words from each of the twelve category norms, while the second set of word lists included the third, fifth, seventh, and ninth ranking words from each of the twelve category norms.

Procedure

Participants came to the lab in pairs and were seated in front of a computer monitor. Although participants did not interact during pilot testing it was important to use pairs of participants to mimic the conditions under which participants in the current experiment would initially recall information. In other words it was important to account for the effect that the mere presence of another participant may have on one's initial recall. Participants were presented with one of the two sets of categorized word lists. Each category name was presented and followed by all four words belonging to that category. Each category name and word was presented one-at-a-time, for 3 seconds. Participants were then asked to perform either an individual cued or non-cued recall task. Items were coded for correct and incorrect responses. The word list that produced the most intrusions for each category during the cued recall task was chosen to make up the final set of word lists (see Appendix D). Although this was done to ensure that the current stimulus materials would produce intrusion errors during initial recall, it is important to note that performance on this task should not be viewed as a general or universal pattern.

Experiment 2 Methods

Participants

One hundred and sixty (50 male, 110 female) University of California, Riverside undergraduates participated in pairs, as partial fulfillment of a requirement for introductory psychology.

Materials

The stimulus materials included categorized word lists organized into twelve categories with four words per category (48-items) selected from the updated Battig and Montague (1969) category norms (Van Overschelde et al., 2004).

Procedure

The procedures and instructions given to participants through initial recall were consistent with those used by Tulving and Pearlstone (1966), with the exception that word lists in the current experiment were presented visually on a computer monitor rather than auditorily from a tape recorder. Participants came to the laboratory in pairs and were seated in front of a single computer monitor. Both participants viewed the stimulus presentation on the same monitor at the same time. This was done to ensure that the presentation of words was the same for both participants, and so that participants *knew* they had seen the same list of words. Before beginning the experiment, participant pairs were told that they would view a long list of words, and that the words were grouped into twelve different categories. Participants were also told that they would first see a category name (e.g., FOUR-FOOTED ANIMALS) followed by four words belonging to that category (e.g., HORSE, TIGER, ELEPHANT, MOUSE). They were also told that the category name was printed in blue and the category members were printed in black. Each category member was presented one-at-a-time, and remained on the screen for 3 seconds. Participants viewed a category name and all of the words from that category before viewing the next category name. The list included a total of twelve category names and forty-eight category instances. Categories were presented in the same order for all participants but the four category instances were presented in a random order for each participant pair. Participants were told that their memory for the words on the list would be tested later so they should pay close attention to each word. They were told that they did not need to remember the category name, just the category members. Participants were also told that when a word was presented on the screen, they should focus their efforts on trying to remember that word until the next word was presented.

The experiment began with a short demonstration to familiarize the participants with the word list structure. Participants were shown a single category name (e.g., METALS) printed in blue and four associated words (e.g., STEEL, BRASS, GOLD, PLATINUM) printed in black. Participants were then asked verbally if they understood the task and the structure of the word list.

Following the presentation of all word lists participants were given a filler task (2 min of solving multiplication problems), and then were asked to recall as many of category instances as they could remember. Half of the participants (N = 80) was given a cued recall task, where each category name (e.g., FOUR-FOOTED ANIMALS) was presented and followed by four blank lines, while the other half of participants (N = 80) was given a non-cued recall task, where only 48 blank lines were provided. Aside from the type of recall task (cued or non-cued) the following procedure was identical for both groups.

The first recall was performed individually. Participants were given a response sheet to record their written recall. All participants were told to write down all of the words they could recall from the previous word lists. They were also instructed that they could write down their responses in any order, and that it was very important to recall as many of the 48 words from the previous word lists as possible. Participants were given four minutes to individually recall as many words as they could. After individual recall, half of the participant pairs was asked to discuss their recall with each other (discussion pairs: 40 cued recall; 40 non-cued recall), while the other half was asked to report their initial recall to a microphone alone (no-discussion pairs: 40 cued recall; 40 non-cued recall).

The procedure used to facilitate discussion between participants was the same as outlined in Experiment 1 (See Figure 2-1). Participants were given their response sheets and instructed to engage in conversation about all of the items they previously recalled. Participants were given as much time as they needed to discuss their initial recalls. Participants in the cued recall group spent an average of 3.1 min (*SD* = 1.4 min) discussing the word lists, and participants in the non-cued recall group spent an average of 3.1 min (*SD* = 1.2 min) discussing the word lists.

The procedure used to get no-discussion participants to report their recall to a microphone was the same outlined in Experiment 1 (See Figure 2-1). Participants were separated into different rooms, given their response sheets and instructed to report each previously recalled word aloud to a microphone. Participants in the no-discussion cued recall group spent an average of 0.98 min (*SD* = 0.25 min) reporting words, and participants in the no-discussion non-cued recall group spent an average of 0.91 min (*SD* = 0.51 min) reporting words.

After either discussing their initial recalls for the word lists, or reading their initial recalls into a microphone, all participants performed a second individual recall task. Participants performed the same type of recall task as they had initially. That is, if a participant was initially given a cued recall task they were again given a second individual cued recall task, and if they were initially given a non-cued recall task they were again asked to do an individual non-cued recall task. Participants were told to write down all of the words they could recall from the original word list presentation. They were also instructed that they could write down their responses in any order, and that it was very important to recall as many of the 48 words from the original word list as possible. Participants were given four minutes to individually recall as many words as they could.

Results and Discussion

The data were scored for correct and incorrect recall. Any word recalled that was presented during the study phase was scored as a correct item, and any word recalled that was not presented during the study phase was scored as an incorrect item.[7] I examined three measures of recall: the number of correct items recalled, the number of incorrect items recalled, and the accuracy of recall (Σ [$Correct_i$ / ($Correct_i$ + $Incorrect_i$)] / N).

The analyses are quite lengthy. Thus, a brief summary of the results is presented first, followed by the details of the statistical analyses. The numbers of correct and incorrect items recalled, and recall accuracy, are given in Table 3-1 and Table 3-2, for

[7] Minor misspellings of presented words were counted as correct items. Presented items recalled with the wrong grammatical number (e.g. plural form rather than singular form) were also counted as correct.

initial recall, exposure items, recall of exposure items, and final recall for the cued recall and non-cued recall conditions respectively.

Table 3-1.
Number of Items Recalled and Accuracy for Initial Recall, Exposure Items, Recall of Exposure Items, and Final Recall for Discussion and No-Discussion Pairs in the *Cued Recall* Condition.

	Correct Recall	Incorrect Recall	Accuracy
	Discussion Pairs (N_{Pairs} = 20; N = 40)		
Initial Recall	24.00 (8.53)	2.75 (3.40)	.90 (.12)
Exposure Items	11.55 (6.74)	2.55 (2.96)	.82 (.21)
Recall of Exposure Items	6.58 (3.47)	0.93 (1.23)	.88 (.16)
Final Recall	30.73 (8.19)	3.15 (3.75)	.91 (.10)
	No Discussion Pairs (N_{Pairs} = 20; N = 40)		
Initial Recall	26.25 (7.90)	2.48 (2.96)	.91 (.10)
Final Recall	27.13 (8.16)	3.35 (3.42)	.89 (.12)

Note. Accuracy = Σ [$Correct_i$ / ($Correct_i$ + $Incorrect_i$)] / N; Standard deviations in parentheses. Calculating accuracy for each individual weights each participant's accuracy score by the magnitude of his or her recall. As a result, accuracy cannot be calculated using the mean correct and incorrect recall values provided in the table.

Table 3-2.
Number of Items Recalled and Accuracy for Initial Recall, Exposure Items, Recall of Exposure Items, and Final Recall for Discussion and No-Discussion Pairs in the *Non-Cued Recall* Condition.

	Correct Recall	Incorrect Recall	Accuracy
	Discussion Pairs (N_{Pairs} = 20; N = 40)		
Initial Recall	17.23 (5.52)	1.48 (1.97)	.92 (.11)
Exposure Items	9.73 (4.52)	1.43 (1.95)	.88 (.16)
Recall of Exposure Items	6.03 (3.11)	0.28 (0.64)	.96 (.10)
Final Recall	24.63 (6.55)	1.48 (2.23)	.94 (.09)
	No Discussion Pairs (N_{Pairs} = 20; N = 40)		
Initial Recall	18.00 (6.21)	1.30 (1.87)	.93 (.09)
Final Recall	18.98 (6.47)	1.95 (2.66)	.91 (.11)

Note. Accuracy = Σ [$Correct_i$ / ($Correct_i$ + $Incorrect_i$)] / N; Standard deviations in parentheses. Calculating accuracy for each individual weights each participant's accuracy score by the magnitude of his or her recall. As a result, accuracy cannot be calculated using the mean correct and incorrect recall values provided in the table.

Summary of Results

The results of Experiment 2 were generally consistent with those of Experiment 1. For both the cued recall and non-cued recall conditions when people discussed their recall with each another, (a) the participant-generated contagion items included more correct than incorrect items; (b) these contagion items had a lower level of accuracy compared to initial recall; (c) people were more likely to incorporate into their subsequent recall the correct information they were exposed to than the incorrect information they were exposed to; (d) the accuracy of final recall increased only slightly relative to initial recall for the discussion pairs; and (e) within pairs recall accuracy increased from initial to final recall for the initially less accurate participant and decreased for the initially more accurate participant, for discussion pairs, while for no-discussion pairs both the initially less accurate and more accurate participants showed a decrease in accuracy from initial to final recall.

A comparison between the cued recall and non-cued recall condition showed that during a discussion, (a) participants in the cued recall condition were exposed to more incorrect information but only slightly more correct information relative to participants in the non-cued recall condition; (b) participants in the cued recall condition incorporated more incorrect items but not more correct items than participants in the non-cued recall condition, and (c) following discussion participants in the cued recall and non-cued recall conditions showed a similar increase in the number of correct items recalled, increase in the number of incorrect items recalled, and the change in accuracy. An in depth report of the statistical analyses from the current experiment is presented next. Results to check the cuing manipulation are reported first, followed by the results for the cued recall condition and the non-cued recall condition. Finally a comparison is made between the cued and non-cued recall conditions. Figures 3-1 and 3-2 highlight the changes for correct recall, incorrect recall, and accuracy between initial and final recall for both discussion and no-discussion pairs in the cued recall and non-cued recall conditions respectively.[8]

[8] The analyses presented were conducted on the individual subjects (N = 40), which may be problematic in meeting the independence assumption. The data were also analyzed by averaging across each pair (N = 20) to meet the independence assumption. Those results can be found in Appendix E.

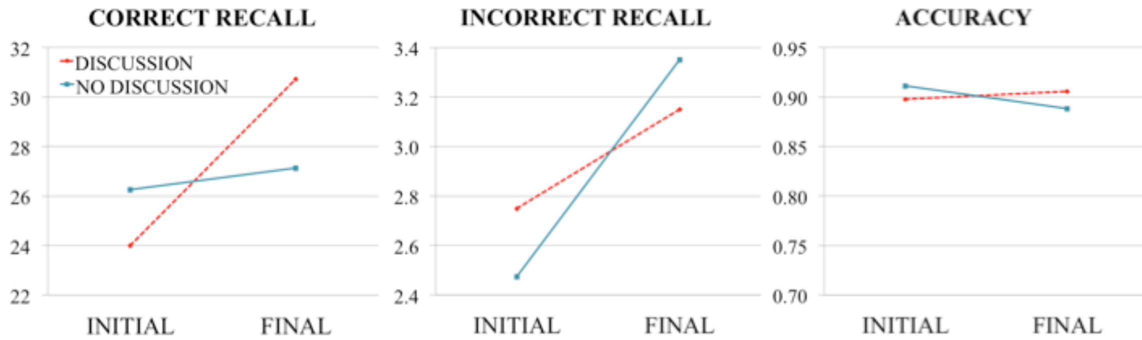

Figure 3-1. Mean number of correct items recalled, incorrect items recalled, and recall accuracy during initial and final recall for discussion and no-discussion pairs in the *cued recall* condition.

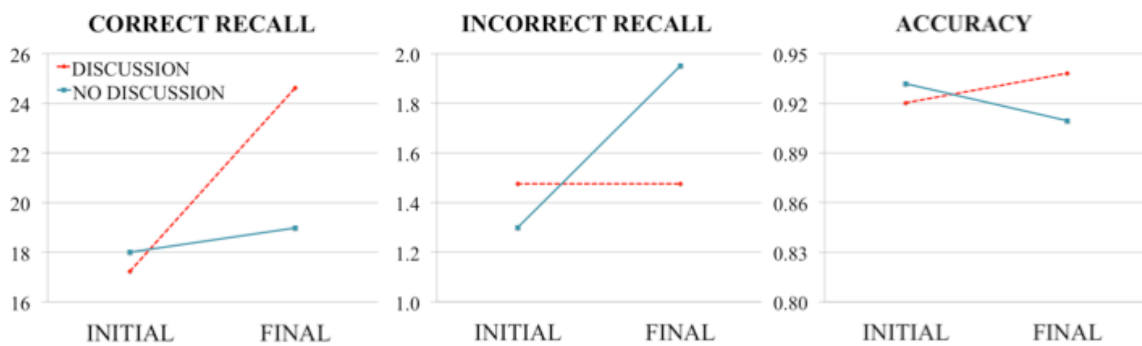

Figure 3-2. Mean number of correct items recalled, incorrect items recalled, and recall accuracy during initial and final recall for discussion and no-discussion pairs in the *non-cued recall* condition.

Manipulation Check: Comparing Cued to Non-Cued Recall

To check that participants in the cued recall condition recalled more information than participants in the non-cued recall condition (replicating the results of Tulving and Pearlstone, 1966) initial recall was compared for the cued recall and non-cued recall conditions. Cued recall produced significantly more correct items than non-cued recall, for both discussion (t (78) = 4.22, p < .001, r = .43) and no-discussion pairs (t (78) =

5.19, $p < .001$, $r = .51$). Cued recall also produced significantly more incorrect items than non-cued recall, for both the discussion (t (78) = 2.05, $p = .044$, $r = .23$) and no-discussion pairs (t (78) = 2.12, $p = .037$, $r = .23$). Although the differences were not statistically significant participants in the cued recall condition had slightly lower initial accuracy than participants in the non-cued recall condition, for both discussion pairs (t (78) = -.873, $p = .385$, $r = -.10$) and no-discussion pairs (t (78) = -.935, $p = .353$, $r = -.11$). The slightly lower accuracy observed for the cued recall condition is a result of the increased production of errors during the cued recall task.

These results provide evidence that cued recall enhanced retrieval strategies resulting in a greater output of both correct and incorrect information, whereas non-cued recall resulted in less effective retrieval strategies resulting in less overall output and more omission errors.

Cued Recall: Discussion Pairs

Exposure Items. The results from Experiment 2 are consistent with those from Experiment 1. When people discussed their recall with each other, the participant-generated exposure items included more correct than incorrect items (t (39) = 7.60, $p < .001$, $r = .77$), and these exposure items had a lower level of accuracy compared to initial recall (t (39) = -1.97, $p = .056$, $r = -.30$). These results provide further support that unique exposure items generated by participants are less accurate than the participants' initial recall.

Incorporation of Exposure Items. Participants again showed the ability to distinguish correct items from incorrect items by incorporating a greater proportion (R_I /

R_E) of the correct exposure items (.64) than the incorrect exposure items (.40) into their final recall (t (25) = 3.13, p = .004, r = .53).[9] The accuracy (Σ [Correct$_i$ / (Correct$_i$ + Incorrect$_i$)] / N) of the incorporated items (.88) was higher than the accuracy of the exposure items (.80), t (39) = 3.11, p = .004, r = .45, and slightly lower than the accuracy of participants' initial recall (.90), t (39) = -0.51, p = .615, r = -.08.

Change Across Recalls. Following the discussion, participants recalled significantly more correct items (t (39) = 12.18, p < .001, r = .89) and only slightly more incorrect items (t (39) = 0.82, p = .415, r = .13). However, despite the greater increase in correct information relative to incorrect information, the accuracy of final recall (.91) was only slightly higher than the accuracy of initial recall (.90), t (39) = .505, p = .616, r = .08.[10] These results again suggest that exposure to less accurate information coupled with the incorporation of more accurate information produces only small changes in overall accuracy.

Cued Recall: No-Discussion Pairs

When participants did not discuss their recall but reported it to a microphone, they recalled significantly more correct items (t (39) = 2.78, p = .008, r = .41) and incorrect items (t (39) = 4.45, p < .001, r = .58) during a second individual recall attempt. These results provide additional evidence that an opportunity to self-cue memory through re-

[9] The analysis of the proportion of exposure items incorporated requires that a participant be exposed to at least one correct and one incorrect item during discussion. As a result, the analysis was preformed on a subset of the sample, and the proportions presented in the text cannot be calculated using the mean values of exposure and incorporation presented in Table 3-1.

[10] Test-retest reliabilities for correct recall, incorrect recall, and accuracy between initial recall and final recall are found in Appendix F for the cued recall discussion and no discussion pairs.

exposure to one's initial recall can produce an increase in overall output. The increases of correct and incorrect recall combined to produce a small but significant decrease in accuracy (t (39) = -4.06, p < .001, r = -.55).

Cued Recall: Discussion Pairs Compared to No-Discussion Pairs

Pairs of participants who discussed their initial recall showed a significantly larger increase for correct item recall relative to the no-discussion group (t (78) = 9.20, p < .001, r = .72). These results are consistent with Experiment 1, providing additional evidence that discussion produces increases in correct output above that produced by self-cuing memory. However, inconsistent with Experiment 1, the no-discussion group showed a slightly larger increase in incorrect item recall relative to the discussion group (t (78) = .907, p = .367, r = .10). Although this effect is small and non-significant, the greater increase of incorrect output for the no-discussion pairs relative to the discussion pairs suggests that the social interaction may provide some protective benefits that reduce error incorporation. The information shared during discussion is likely to include more than just the recalled items, such as a person's confidence in recall and feedback about the accuracy of reported items. To the extent that participants can accurately assess this information during discussion, it may result in less error incorporation in post-discussion recall than when participants recall alone. The slight increase in accuracy for the discussion group (M = .008) and the slight decrease in accuracy for the no-discussion group (M = -.023) produced a marginally significant difference in accuracy change between the discussion and no-discussion groups (t (78) = 1.89, p = .062, r = .21).

Accuracy Change Within Cued Recall Pairs

The results of the current experiment suggest that discussion produces only a small change in accuracy for individuals who discussed their recall prior to the second recall. Additional analyses were conducted to examine whether the accuracy change occurred within pairs. Figure 3-3 and Figure 3-4 depict the mean accuracy for initial and final recall for the better and worse participant in each discussion pair and no-discussion pair respectively. Pairs of participants with the same initial accuracy were excluded from the following analyses.

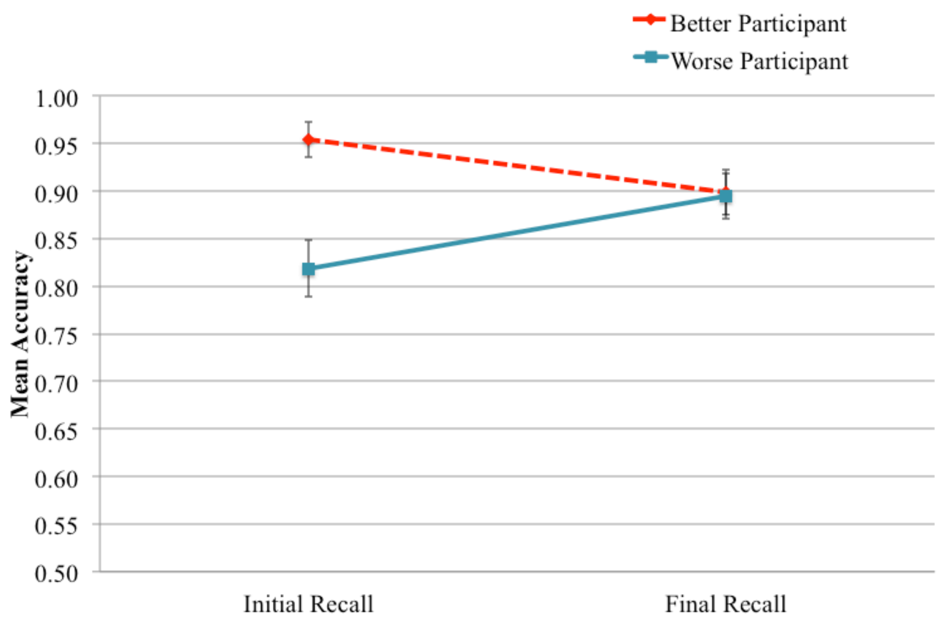

Figure 3-3. Mean initial and final accuracy for the better and worse discussion participants in the *cued recall* condition.

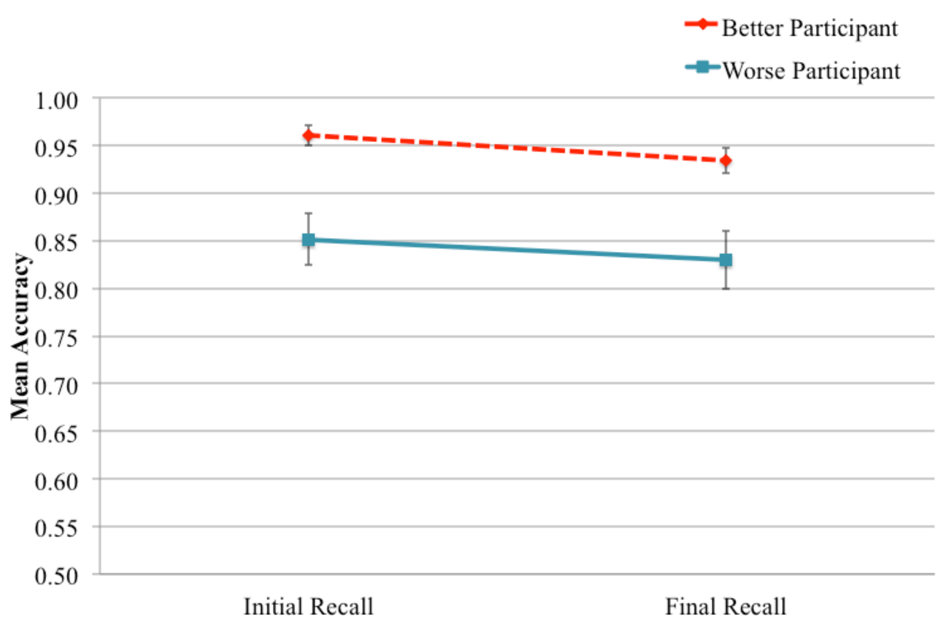

Figure 3-4. Mean initial and final accuracy for the better and worse no-discussion participants in the *cued recall* condition.

Analysis of accuracy change within discussion pairs revealed the same accuracy redistribution effect observed in Experiment 1. Figure 3-3 clearly shows that discussion led to a decrease in accuracy for the initially more accurate member (M = -.056) and an increase in accuracy for the initially less accurate member (M = .076) of each participant pair (t (17) = 5.68, p < .001, r = .81), resulting in participant pairs having virtually the same final accuracy (t (17) = .255, p = .802, r = .06).

Figure 3-4 shows a different pattern of results for no-discussion pairs, with a slight decrease in accuracy for the initially more accurate participant (M = -.027) and also a slight decrease in accuracy for the initially less accurate participant (M = -.022), t (18) = -.371, p = .715, r = -.09. For the no-discussion group, the initially more accurate participant remained more accurate than the initially less accurate participant, t (18) = 3.51, p = .003, r = .64. An independent samples comparison showed that the patterns of gains and losses were significantly greater for discussing pairs compared to non-discussing pairs, t (35) = 4.83, p < .001, r = .63.[11]

Non-Cued Recall: Discussion Pairs

Exposure Items. The analyses follow the same order as those presented for the cued recall condition and are generally consistent with those findings. When people discussed their recall with each another, the participant-generated exposure items included more correct than incorrect items (t (39) = 10.65, p < .001, r = .86), and these

[11] This independent samples t-test was computed using difference scores calculated by subtracting the initially less accurate participant's accuracy change from the initially more accurate participant's accuracy change for the discussion and no discussion pairs.

exposure items had a lower level of accuracy compared to initial recall (t (39) = -1.53, p = .135, r = -.24).

Incorporation of Exposure Items. Participants incorporated a greater proportion (R_I / R_E) of the correct items (.66) they were exposed to than the incorrect items (.17) they were exposed to (t (20) = 6.28, p < .001, r = .81).[12] The accuracy (Σ [$Correct_i$ / ($Correct_i$ + $Incorrect_i$)] / N) of the incorporated items (.96) was higher than the accuracy of the exposure items (.88), t (39) = 3.91, p < .001, r = .53, and slightly higher than the accuracy of participants' initial recall (.92), t (39) = 1.64, p = .109, r = .25. Correct exposure information once again showed a contagious advantage suggesting that these correct items may serve as retrieval cues for memory.

Change Across Recalls. Following the discussion participants recalled significantly more correct items (t (39) = 13.00, p < .001, r = .90), but there was no change in the number of incorrect items recalled post-discussion (t (39) = 0.00, p = 1.00, r = .00). Even though there was a significant gain in the amount of correct recall accompanied by no change in the amount of incorrect recall, the accuracy of final recall (.94) was only slightly higher than the accuracy of initial recall (.92), t (39) = 1.68, p = .101, r = .26.[13]

---

[12] The analysis was preformed on a subset of the sample that was exposed to at least one correct and incorrect item during discussion. The proportions presented in the text cannot be calculated using the mean values of exposure and incorporation presented in Table 3-2.

[13] Test-retest reliabilities for correct recall, incorrect recall, and accuracy between initial recall and final recall are found in Appendix F for the non-cued recall discussion and no discussion pairs.

Non-Cued Recall: No-Discussion Pairs

Participants who did not discuss their recall but reported it to a microphone recalled significantly more correct items (t (39) = 3.16, p = .003, r = .44) and incorrect items (t (39) = 3.66, p = .001, r = .51) at final recall compared to initial recall. Opportunity for self-cuing one's memory once again resulted in increased overall output. The increases in correct and incorrect recall combined to produce a small but significant decrease in accuracy (t (39) = -3.81, p < .001, r = -.52).

Non-Cued Recall: Discussion Pairs Compared to No-Discussion Pairs

Participants who discussed their initial recall showed a significantly larger increase for correct item recall relative to the no-discussion group (t (78) = 9.92, p < .001, r = .75), suggesting that discussion benefits correct recall beyond what would be gained through self-cuing. Participants in the no-discussion group showed a significantly larger increase for incorrect item recall relative to the discussion group (t (78) = 2.09, p = .04, r = .23). This result is consistent with the result observed in the cued recall condition. These results taken together suggest that discussion benefits later memory performance and possibly reduces the risk of increased error production. The slight increase in accuracy for the discussion group (M = .018) and the slight decrease in accuracy for the no-discussion group (M = -.022) produced a significant difference for accuracy change between the two groups (t (78) = 3.30, p = .001, r = .35).

Accuracy Change Within Non-Cued Recall Pairs

Accuracy change for the discussion pairs was again small and non-significant. To examine if there were greater changes in accuracy within pairs, additional analyses were conducted. Figure 3-5 and Figure 3-6 depict the mean accuracy for initial and final recall for the better and worse participant in each discussion pair and no-discussion pair respectively. Pairs of participants with the same initial accuracy were excluded from the following analyses.

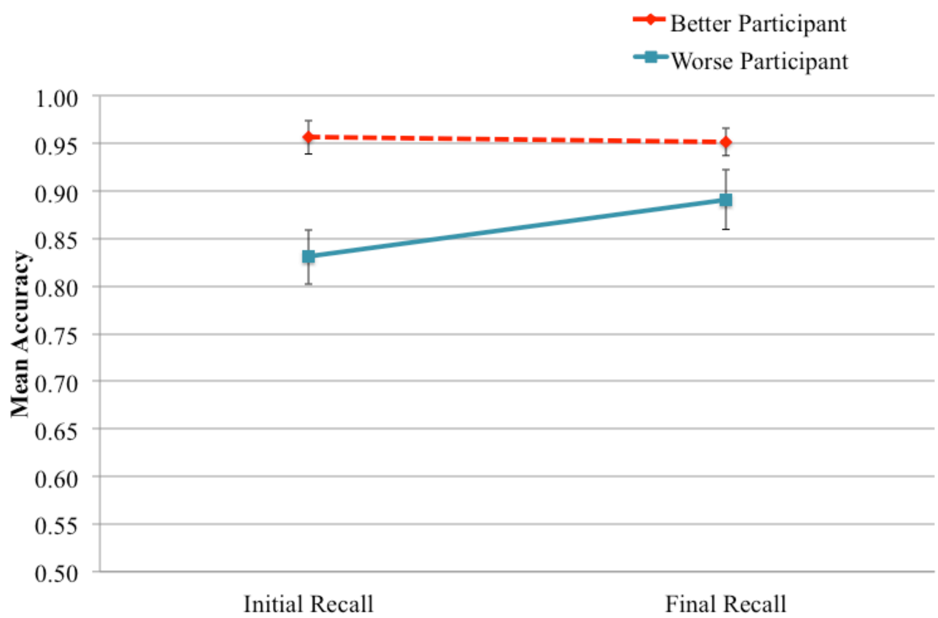

Figure 3-5. Mean initial and final accuracy for the better and worse discussion participants in the *non-cued recall* condition.

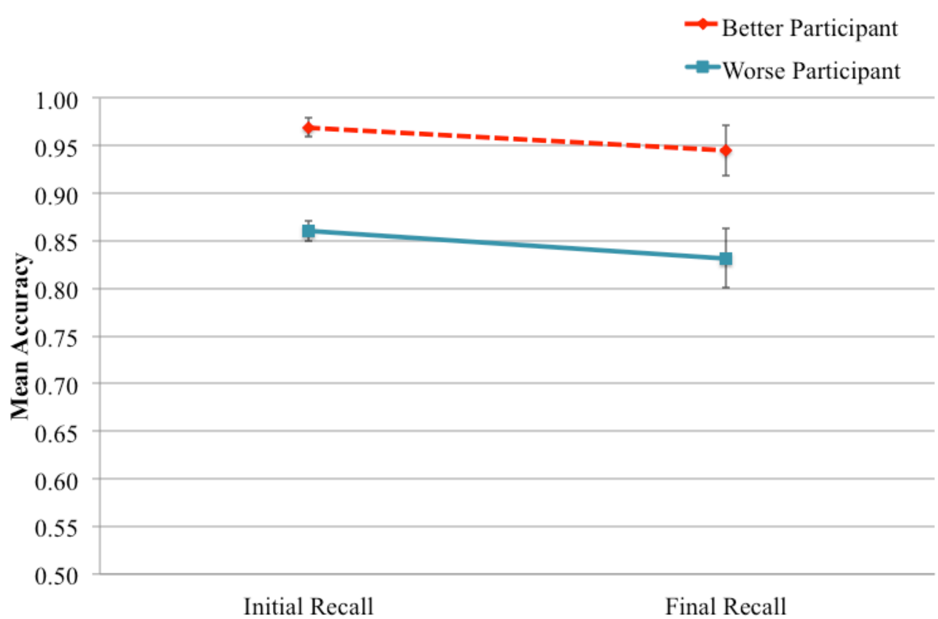

Figure 3-6. Mean initial and final accuracy for the better and worse no-discussion participants in the *non-cued recall* condition.

Analysis of accuracy change within discussion pairs did not reveal the same accuracy redistribution effect observed in the cued recall condition. Figure 3-5 shows that discussion led to a very small decrease in accuracy for the initially more accurate member (M = -.005) and an increase in accuracy for the initially less accurate member (M = .06) of each participant pair (t (14) = 2.53, p = .024, r = .56). The initially more accurate participant remained more accurate than the initially less accurate participant, t (14) = 2.72, p = .017, r = .59.

Non-cued no-discussion pairs showed the same pattern of results as the cued recall no-discussion pairs. Figure 3-6 shows a slight decrease in accuracy for the initially more accurate participant (M = -.024) and a slight decrease in accuracy for the initially less accurate participant (M = -.029), t (15) = .323, p = .751, r = .08. For the no-discussion group, the initially more accurate participant remained more accurate than the initially less accurate participant, t (15) = 3.84, p = .001, r = .70. An independent samples comparison showed that the patterns of gains and losses were significantly greater for discussing pairs compared to non-discussing pairs, t (29) = 2.40, p = .023, r = .41.[14]

Cued Recall Compared to Non-Cued Recall

The comparison of discussion pairs that performed a cued recall task and those that performed a non-cued recall task showed that discussion did not affect post-discussion recall differently for the two conditions. That is when participants discussed their initial recall, they showed similar increases in the number of correct items recalled,

[14] This independent samples t-test was computed using difference scores calculated by subtracting the initially less accurate participant's accuracy change from the initially more accurate participant's accuracy change for the discussion and no discussion pairs.

similar increases in the number of incorrect items recalled, and similar changes in accuracy during post-discussion recall, regardless of whether they performed a cued or non-cued recall. The specific statistical analyses that led to this effect are presented below.

Exposure Items. Results supported the prediction that participants in the cued recall condition would experience greater exposure to incorrect information than those in the non-cued recall condition. During discussion, individuals in the cued recall condition were exposed to significantly more incorrect items, t (78) = 2.01, p = .048, r = .22, than those in the non-cued recall condition. This result suggests that the increased output of errors in the cued recall condition relative to the non-cued recall condition translates into greater exposure to error during discussion. Because these intrusion errors are likely to be unique to each group member, almost all errors produced during an initial individual recall will become incorrect exposure items during a later discussion.

However, the data did not support the prediction that individuals in the non-cued recall condition would experience greater exposure to correct items than those in the cued recall condition because of increased omission errors. Participants in the non-cued recall condition were actually exposed to slightly fewer correct items, t (78) = -1.42, p = .159, r = -.16, than those in the cued recall condition. Providing a cue for recall increased the number of correct items reported, and although participants largely report the same additional correct items during an individual cued recall, a few of these correct items were unique to each participant. This resulted in exposure to slightly more items that were correct during discussion for participants in the cued recall condition than those in the non-cued recall condition.

The accuracy of exposure items for those in the cued recall condition (.82) was slightly lower than the accuracy of exposure items for participants in the non-cued recall condition (.88), t (78) = -1.38, p = .173, r = -.15. Participants in the cued recall condition were exposed to similar amounts of correct information but significantly more incorrect information, relative to participants in the non-cued recall condition resulting in lower exposure accuracy.

Incorporation of Exposure Items. The data supported the prediction that participants in the cued recall condition would incorporate more incorrect items than those in the non-cued recall condition. Participants in the cued recall condition incorporated a higher proportion of incorrect items (.40) than those in the non-cued recall condition (.17), t (45) = 2.65, p = .011, r = .29. As stated previously the extent to which someone can incorporate errors into their memory is completely dependent on exposure to those errors. The current results suggest that increasing the number of errors during discussion results in an increased incorporation of errors into later memory reports.

The prediction that participants in the non-cued recall condition would incorporate more correct items into recall following discussion than those in the cued recall condition was not supported by the data. While non-cued participants reported less overall output than their cued recall counterparts, both groups experienced similar exposure to correct items during discussion. This resulted in participants in the cued recall condition (.657) and the non-cued recall condition (.661) incorporating similar proportions of correct items into post-discussion recall (t (78) = -.071, p = .943, r = -.01).

The accuracy of incorporated items for those in the cued recall condition (.88) was significantly lower than the accuracy of incorporated items for participants in the non-cued recall condition (.96), t (78) = -2.47, p = .016, r = -.27. Because cued recall participants incorporated significantly more incorrect items, but incorporated similar amounts of correct items compared to non-cued recall participants, the accuracy of incorporated items for cued recall is significantly lower than the accuracy of non-cued recall.

Final Recall. During final recall discussion pairs in the cued recall condition produced significantly more correct items, t (78) = 3.68, p < .001, r = .38, and significantly more incorrect items, t (78) = 2.43, p = .017, r = .27, than those in the non-cued recall condition.

The prediction that cued recall participants would suffer a loss in accuracy following discussion because of increased exposure to incorrect items was not supported by the data. The prediction that non-cued recall participants would show increased accuracy following discussion was observed; however the effect was very small and did not reach statistical significance. Participants in both the cued and non-cued recall conditions showed the same small increase in accuracy following discussion (t (78) = -0.55, p = .585, r = -.06), and there was no significant difference in final accuracy between the cued and non-cued recall conditions, t (78) = -1.55, p = .124, r = -.17. Although discussion resulted in an increased number of correctly and incorrectly recalled items for both the cued recall and non-cued recall pairs, the number of items by which recall increased was not significantly different between the two conditions for correct information (t (78) = -.851, p = .397, r = -.10) or incorrect information (t (78) = .729, p = .468, r = .08). These results suggest that the effect of discussion is similar regardless of whether the participants initially performed a cued recall or non-cued recall. The type of information produced during a cued recall task may generate this result. The increased correct information produced during a cued rather than non-cued recall task is likely to overlap between participants. This results in very few additional correct exposure items in the cued recall condition, ultimately leading to no difference for the increased number of correct items recalled following discussion for cued rather than non-cued recall. The increased incorrect information produced during a cued recall is likely to be unique to each participant resulting in greater exposure to incorrect information during a discussion. However, the incorporation of incorrect items is dependent on the items ability to serve as a retrieval cue or be very familiar (Jacoby, 1991; Yonelinas, 1994; Wixted & Mickes, 2010). If the additional incorrect exposure items produced during cued recall do not act as retrieval cues or provide a strong sense of familiarity, than the incorporation of incorrect exposure items should be similar for both cued and non-cued participants.

Participants in the no-discussion group showed the same pattern of results. During final recall no-discussion pairs in the cued recall condition produced significantly more correct items, t (78) = 4.95, $p < .001$, r = .49, and significantly more incorrect items, t (78) = 2.04, p = .044, r = .23, than those in the non-cued recall condition. Both cued and non-cued recall participants showed small but equivalent decreases in accuracy after reporting initial recall to a microphone, however there was no difference in final accuracy between the cued and non-cued recall conditions, t (78) = -.867, p = .389, r = -.10.

The results from Experiment 2 are consistent with Experiment 1, showing that both correct and incorrect exposure information became contagious during discussion, but that correct exposure information was more contagious than incorrect exposure information. The results only partially supported the central prediction that cued recall would increase errors and decrease accuracy: Errors increased following a discussion for participants in the cued recall condition, but there was very little change in accuracy.

Although the word lists produced more intrusion errors for the cued recall than the non-cued recall condition, the error rate was low, and initial accuracy was near ceiling for both the cued (.90) and non-cued recall (.92) conditions. This may be a methodological effect, resulting from the categorized word lists themselves. Each category word list included only four items presented sequentially in a group, which may have facilitated encoding during the study phase, generating a good memory for the items. To the extent that participants had, at least, partial information regarding the list items stored in memory, they may have easily rejected incorrect exposure items because the items did not either provide a cue for recall or seem familiar.

Under the current conditions, memory was very accurate, but how will discussion influence accuracy when the design of the stimulus materials elicits specific systematic errors? Experiment 3 examines this question within a social contagion framework by presenting subjects with stimulus materials that vary on the likelihood of producing specific errors.

Chapter 4 – Experiment 3

Introduction

Many false memory and collaborative memory experiments are designed in such a way as to maximize the likelihood that participants will make errors, and the results of these experiments are taken as evidence that people can easily report errors when remembering alone or even during collaboration. However there is an issue of circularity in the relationship between the experimental design and the conclusions drawn from these studies. Specifically, people make errors in experiments designed to elicit errors. The results of these experiments are not evidence that people make errors in general, but rather that people make errors under certain conditions (e.g., in experiments that are designed to produce memory errors). This is a problem for research on the topic of social judgment and memory as posited by Funder (1986). Funder raises the concern that what constitutes an error in a limited and artificial experimental setting may not be an error when applied to a more realistic setting. Extending this idea using artificial, well-crafted procedures and stimuli to induce memory error results in an inflation of memory errors that may lead to an inaccurate overgeneralization that human memory is generally prone to of error and that collaboration compounds that error. Experiment 3 was designed to address the question: to what extent do people make errors in general when they are not presented with stimuli designed specifically to produce memory errors?

Memory is reconstructive in nature. That is, when people remember, they recreate their experiences rather than relive them. People do this by using schema-consistent information, prior knowledge, stereotypes, and beliefs to rebuild memories during retrieval (Bartlett, 1932; Brewer & Treyens, 1981, Loftus, Miller, & Burns, 1978; Roediger & McDermott, 1995; Meade, Watson, Balota, & Roediger, 2007).

Because recall is reconstructive, people will occasionally report information that is schematically or thematically consistent with, but never presented in the original stimulus. Brewer and Treyens (1981) showed this clearly when they had participants wait briefly in a student office while the experimenter finished "setting up" the experiment. After 35 seconds the participant was led to another room by the experimenter who asked them to recall everything that they could remember seeing in the previous room. Not surprisingly almost all participants accurately reported common items from the room (e.g. desk, chair, walls); however, 30% of participants also incorrectly reported having seen books. Although books are likely to be found in a graduate student's office, they had all been removed from the office that constituted the stimulus material for the experiment. Thus, recall of books was an intrusion error. Participants in this experiment were presumably using schematic knowledge to fill in gaps in their memory with expected information, which on occasion can lead to errors (e.g. reporting books in a bookless office). These results suggest that during memory reconstruction, people tend to use schema consistent or high-expectancy information to rebuild their memory.

Roediger and McDermott (1995) also provided evidence that demonstrates memories for things that did not occur, but are highly related to the study material, can easily happen. They had participants study 12-item lists (commonly referred to as the Deese, Roediger, and McDermott or DRM Lists) made up of words such as *bed, rest, awake,* and *pillow*, words highly associated with the word *sleep*. Importantly, *sleep*, was not presented on the list, but rather served as a *critical lure* word that was highly related to all of the other items presented on the list. During an immediate free recall, 40% of participants incorrectly recalled *sleep* from the list. In a second study Roediger and McDermott reported that 55% of participants reported the critical lure during a free recall task, and that during a subsequent recognition memory task 81% of participants made a false alarm for the critical word. This suggests that falsely remembering the critical lure during recall strengthened the false memories of the critical word, leading to a higher false recognition rate. It seems that these meaning-based false memories occur because people use thematic information in the study lists during efforts to remember earlier experienced information (Meade, et al., 2007).

Both of these studies as well as the work by Roediger et al. (2001) share one thing in common, each study relies on a stimulus set or experimental design that was carefully crafted to make memory errors more likely to occur. The frequency of false recall then is completely contingent on the likelihood that the stimulus will produce error in the first place. This is easily observed with a stimulus set like the DRM lists that have been analyzed and coded for which lists produce the highest incidences of false recall (Stadler, Roediger, & McDermott, 1999). The results from a study using DRM lists will be dependent on the specific subset of lists chosen for the particular experiment.

How do such high-error stimulus materials affect collaborative memory? Basden, Basden, Thomas, and Souphasith (1998) reported that using DRM lists during a collaborative recall task resulted in collaborating groups reporting the same number of critical lures as nominal group comparisons. Thorley and Dewhurst (2007) however, found that when using DRM lists collaborative groups recalled more critical lures than their nominal group comparison. How could two experiments using the same stimulus set produce different results? The answer lies in the specific subset of DRM lists used by Basden et al. and Thorley and Dewhurst. More specifically Basden et al. used a set of DRM lists that had a mean likelihood of inducing false recall of the critical lures 44% of the time, while Thorley and Dewhurst used a subset of lists that had a mean likelihood of inducing the critical lure 53% of the time. This means that in the Thorley and Dewhurst study individual group members were more likely to recall the critical lure during collaboration where there was pressure to participate than in the Basden et al. study.

Roediger et al. (2001) also reported that participants were more likely to incorporate incorrect high-expectancy contagion items (41%) (i.e. items found to be highly associated with the scene but not part of the original presentation) than incorrect low-expectancy contagion items (17%) into post-collaborative individual recall. They also reported that for scenes where participants were not exposed to any of the specified incorrect contagion items more participants spontaneously reported the high-expectancy (11%) than low-expectancy (3%) items. These results taken together suggest that people are more likely to augment their memory for an experience missing critical details with high-expectancy information than low-expectancy information both individually and following collaboration. Because Roediger et al. used a confederate to introduce incorrect information during a turn taking collaborative memory task, it is impossible to know the frequency of correct and incorrect recall and exposure had the task been a free flowing discussion between two participants. It is also important to mention that a high likelihood of error at the individual level does not necessarily mean that there will be a high level of error transmission at the group level. If Person A and Person B both recall the false critical lure, then there is no error to transmit because both participants have already produced the error on their own. On the other hand, if the likelihood of error is extremely low such that neither Person A nor Person B recalls the critical lure, then there is also no error to transmit, because neither person produced the error on their own. The question remains, to what extent do people make errors in general when they are *not* presented with a stimulus or an experimental procedure designed specifically to produce memory errors?

Experiment 3 used the same basic social contagion methodology as Experiment 1 and 2 with two changes. First, pairs of participants saw a set of DRM word lists, pioneered by Deese (1959) and Roediger and McDermott (1995) that either included or did not include a high expectancy critical item.[15] Second, the current study employed a within-subjects design rather than a between-subjects design. A within-subjects design was selected to ensure that subjects would not figure out the nature of the stimulus materials. The concern was that using a between-subjects design might result in participants recognizing that the critical item was not presented. That is if people are given four repeated trails, where they are presented with a list of items that all seem to be related to a particular concept, but that key concept is not presented, they might notice this across trials and become resistant to producing the critical item during recall. As with the previous studies, Experiment 3 did not utilize a confederate.

---

[15] When the critical item is presented as part of the study list it is not considered a lure. For consistency purposes the term *critical item* is used throughout the rest of the paper.

Predictions

The key data here concern not only the accuracy of the exposure information and the proportions of correct and incorrect exposure items incorporated into each person's subsequent recall, but importantly, the exposure and later incorporation of the high-expectancy critical item. Previous results by Brewer and Treyens (1981), Roediger and McDermott (1995), and Roediger et al. (2001), as well as, theories of recognition memory suggest that in a free recall task, people tend to make very few intrusion errors, unless the materials are conducive to such errors. If participant pairs are exposed to a DRM word list where a high-expectancy item has been removed rather than a list where the item is included, it is more likely that at least one member will initially report the errant piece of information possibly causing the other member to make a contagion error by incorporating the item following discussion. In other words, contagion errors will only occur to the extent that there are some intrusion errors to begin with. This should affect accuracy in two ways. First, initial accuracy should be lower when participants recall items from a DRM list where the critical item is absent than where the critical item is not absent. This will occur because the design of DRM lists in their original form (critical item not presented) produce specific false memories. When the critical item is included as part of the list, very few intrusions should occur because there is no longer a highly associated item missing from the list. Second, because participants are more likely to experience exposure to the critical item during discussion for trials where the critical item is not present during the study phase, participants should show a greater decrease in accuracy following discussion on these trials compared to trials where the critical item is initially included. This will occur to the extent that only one of the participants initially reports the critical item during trials when the critical item is not present in the word list.

When exposure to an item occurs during discussion, people can only use the item as a cue for recollection to the extent that there is information about the item stored in memory. If there is no information for the item stored in memory, participants will only recognize the item to the extent that the item is highly familiar. During the study phase of an experiment when people view lists of semantically related concepts like those in the DRM lists, they may create a false memory for the non-presented high expectancy critical item. To the extent that this occurs, incorporation of the critical item following exposure should be high on trials when the critical item was not included in the word list. For correct information that is partially stored in memory, exposure should serve as a cue for recollection and familiarity, resulting in a greater proportion of correct exposure items incorporated into subsequent recall than incorrect exposure items.

Methods

Participants

One hundred and sixty (62 male, 98 female) University of California, Riverside undergraduates participated in pairs, as partial fulfillment of a requirement for introductory psychology.

Materials

Stimulus materials included a set of four DRM word lists (Roediger & McDermott, 1995; Stadler, Roediger, & McDermott, 1999; Deese, 1959). Word lists included *Sweet, Chair, Smoke, and Rough* (See Appendix G). Word lists included 15-

items (e.g. SOUR, CANDY, SUGAR, etc.) associated with a critical item (e.g. SWEET). These items were rank-ordered from the item most associated with the critical item concept to least associated item. For half of the trials the critical item was included in the list, and was not included in the other half. To maintain consistency of list length across those two conditions the least associated word in the standard DRM list was removed for these trials. The DRM lists selected for the current experiment produced the false recall of the critical item about 50% of the time (*M* = 52.75%; Stadler et al., 1999). This was done to ensure that the critical item would be falsely recalled during the experiment, but not so often that both participants always recalled the critical item. This would create a situation where the critical item could serve as an incorrect exposure item during discussion trials, because one but not both participants incorrectly recalled the critical item during initial recall.

Procedure

Participants came to the laboratory in pairs and were seated in front a single computer monitor. Both participants viewed the stimulus presentation on the same monitor at the same time. Before the experiment began participants were told that they would be viewing several lists of related words. Participants were also told that they would first see a title screen indicating which list they were viewing (i.e. LIST 1). The list name was printed in blue and the related words were printed in black. Each word was presented one-at-a-time and remained on the screen for 1 sec. The word lists were presented in one of two orders for all participants but the 15 items within each list were presented in a random order for each participant pair. Participants were also told that following the presentation of a word list they would be asked to complete a series of math equations for several minutes (2 min) followed by an individual recall task where they would be asked to write down as many items from the word list as they could remember. They were also instructed that for each trial they may or may not be asked to discuss their recall with their partner.

To familiarize participants with the task, the experiment began with a short demonstration. Participants were shown the list title "EXAMPLE LIST" printed in blue, which was followed by the presentation of the 15-item DRM list for the critical item BREAD, printed in black. For this demonstration the critical item was always presented among words in the list and the fifteenth ranked associated word (toast) was removed. Participants were then asked to perform a 2-minute recall for all of the items they could remember from the example word list. Participants were then asked verbally if they understood the task and the structure of the word list.

The experiment included four test trials in a 2 (critical item included or not included in word list) x 2 (discussion or no-discussion of initial recall) within-subjects design. The inclusion of the critical item and whether the participants discussed their recall was counterbalanced such that all four combinations (item included – discussion, item included – no-discussion, item not included – discussion, item not included – no-discussion) occurred equally often. Although there were only four test trials participants were told that they would be completing six trials in an effort to prevent them from trying to figure out the order of discussion and no-discussion trials. During a test trial participants viewed a list of words, completed an individual recall task, discussed their recall or reported it to a microphone, and completed a second individual recall before moving on to the next test trial. This process was repeated four times during the course of the experiment.

The procedures were the same regardless of whether the critical item was included in the word list. Following the presentation of a word list, the first recall was always performed individually. Participants were given a response sheet to record their written recall. The response sheet was labeled with the list name (e.g. List 1). Participants were told to write down all of the words they could recall from the presented word list. They were also instructed that they could write down their responses in any order, but to list them in the order that the words came to mind. They were also told that it was very important to recall as many of the words from the previous word list as possible. Participants were given two minutes to individually recall as many words as they could. After individual recall, the participant pairs were asked to either discuss their recall with each other (on 2 trials), or to report their initial recall to a microphone (on 2 trials).

The procedure used to facilitate discussion between participants was the same outlined in Experiments 1 and 2. Participants were given their response sheets and instructed to engage in conversation about all of the items they previously recalled. Participants were given as much time as they needed to discuss their initial recalls. When the critical item was included participants in the group spent an average of 1.39 min (*SD* = 0.54 min) discussing their recall, and when the critical item was not included participants in the group spent an average of 1.39 min (*SD* = 0.63 min) discussing their recall.

The procedure used to get participants to report their recall to a microphone was also the same as outlined in Experiments 1 and 2. Participants were separated into different rooms, given their response sheets and instructed to report each previously recalled word aloud to a microphone. When the critical item was included, participants spent an average of 0.48 min (*SD* = 0.31 min) reporting words, and when the critical item was not included participants in the group spent an average of 0.43 min (*SD* = 0.20 min) reporting words.

After either discussing their individual recalls, or reading their individual recalls into a microphone, participants' initial response sheets were collected. All participants performed a second individual recall task. Participants were given a new response sheet identical to the one used during initial recall. Participants were given the same instructions as for initial recall. Participants were given two minutes to individually recall as many words as they could.

Results and Discussion

The data were scored for correct and incorrect recall. Any word recalled that was presented during the study phase was scored as a correct item, and any word recalled that was not presented during the study phase was scored as an incorrect item.[16] The number of correct items recalled, the number of incorrect items recalled, and the accuracy of recall (Σ [(Correct$_i$ + Incorrect$_i$)] / N) were all measured.

[16] Minor misspellings of presented words were counted as correct items. Presented items recalled with the wrong grammatical number (e.g. plural form rather than singular form) were also counted as correct.

A brief summary of the main findings is presented first followed by the comprehensive statistical analyses for the discussion and no-discussion trials when the critical item was not presented, for the discussion and no-discussion trials when the critical item was presented in the word list, and finally as a comparison between critical item present and absent trials. The numbers of correct and incorrect items recalled, and recall accuracy, are given in Table 4-1 and Table 4-2, for initial recall, exposure items, recall of exposure items, and final recall for the critical item not presented and critical item presented trials respectively.

Table 4-1.
Number of Items Recalled and Accuracy for Initial Recall, Exposure Items, Recall of Exposure Items, and Final Recall for Discussion and No-Discussion Trials when the *Critical Item was Not Presented* on the Word List.

	Correct	Incorrect	Accuracy
	Discussion Trial (N_{Pairs} = 80; N = 160)		
Initial Recall	8.04 (2.12)	1.21 (1.14)	.87 (.12)
Exposure Items	3.11 (1.51)	0.76 (1.08)	.82 (.24)
Recall of Exposure Items	2.23 (1.27)	0.34 (0.61)	.88 (.22)
Final Recall	9.69 (2.16)	1.45 (1.01)	.87 (.09)
	No Discussion Trial (N_{Pairs} = 80; N = 160)		
Initial Recall	8.21 (1.83)	1.12 (0.89)	.88 (.09)
Final Recall	7.88 (1.82)	1.24 (0.99)	.87 (.11)

Note. Accuracy = Σ [$Correct_i$ / ($Correct_i$ + $Incorrect_i$)] / N; Standard deviations in parentheses. Calculating accuracy for each individual weights each participant's accuracy score by the magnitude of his or her recall. As a result, accuracy cannot be calculated using the mean correct and incorrect recall values provided in the table.

Table 4-2.
Number of Items Recalled and Accuracy for Initial Recall, Exposure Items, Recall of Exposure Items, and Final Recall for Discussion and No-Discussion Trials when the *Critical Item was Presented* on the Word List.

	Correct	Incorrect	Accuracy
	Discussion Trial (N_{Pairs} = 80; N = 160)		
Initial Recall	8.74 (2.04)	0.47 (0.97)	.95 (.09)
Exposure Items	3.09 (1.68)	0.46 (0.96)	.88 (.23)
Recall of Exposure Items	2.16 (1.30)	0.23 (0.54)	.91 (.21)
Final Recall	10.24 (1.76)	0.73 (1.01)	.94 (.08)
	No Discussion Trial (N_{Pairs} = 80; N = 160)		
Initial Recall	8.86 (1.71)	0.54 (0.84)	.95 (.08)
Final Recall	8.52 (1.75)	0.66 (0.96)	.93 (.10)

Note. Accuracy = Σ [$Correct_i$ / ($Correct_i$ + $Incorrect_i$)] / N; Standard deviations in parentheses. Calculating accuracy for each individual weights each participant's accuracy score by the magnitude of his or her recall. As a result, accuracy cannot be calculated using the mean correct and incorrect recall values provided in the table.

Summary of Results

The results of the current experiment are consistent with those of Experiments 1 and 2. Regardless of whether the word list included the critical item during discussion trials, (a) the participant-generated contagion items included more correct than incorrect items; (b) these contagion items had a lower level of accuracy compared to initial recall; (c) people were more likely to incorporate the correct exposure information than the incorrect exposure information into post-discussion recall; (d) the accuracy of final recall changed only slightly relative to initial recall for the discussion pairs; and (e) within discussion pairs the initially less accurate and more accurate participants showed an accuracy redistribution effect, while for no-discussion pairs both the initially less accurate and more accurate participants showed a decrease in accuracy from initial to final recall.

A comparison between trials that did not include the critical item and that did include the critical item showed that during discussion, (a) participants experienced exposure to more incorrect information but not more correct information when the critical item was not included in the word list than when it was included; (b) participants incorporated similar amounts of correct and incorrect exposure information regardless of whether the critical item was included in the word list, and (c) following discussion both when the critical item was included in the list and not included in the list, participants showed a similar increase in the number of correct items recalled, the number of incorrect items recalled, and the change in accuracy. An in depth report of the statistical analyses from the current experiment is presented next. Results are reported first for the trials when the critical item was not presented, followed by the trials when the critical item was included in the word list, and then finally as a comparison between trials when the critical item was not included in the word list and trials when the critical item was included in the word list. Figures 4-1 and 4-2 highlight the changes for correct recall, incorrect recall, and accuracy between initial and final recall for both discussion and no-discussion pairs during trials when the critical item was not presented and trials when the critical item was presented respectively.[17]

---

[17] The analyses presented were conducted on the individual subjects (N = 160), which may be problematic in meeting the independence assumption. The data were also analyzed by averaging across each pair (N = 80) to meet the independence assumption. Those results can be found in Appendix H.

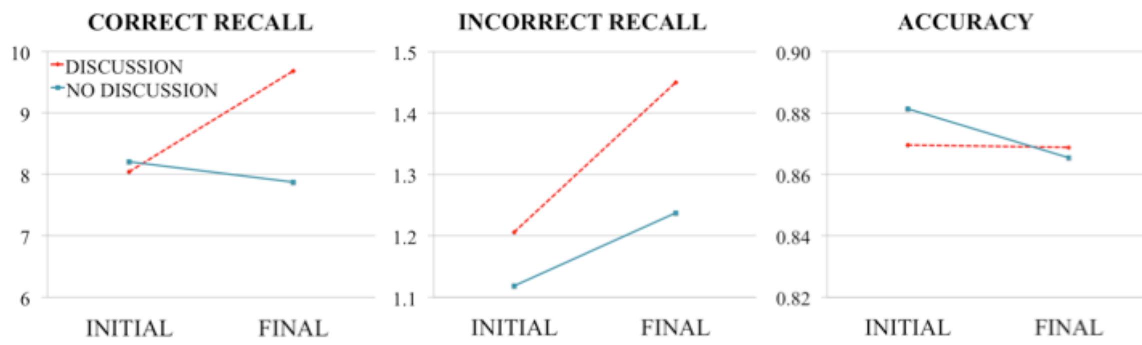

Figure 4-1. Mean number of correct items recalled, incorrect items recalled, and recall accuracy during initial and final recall for discussion and no-discussion pairs on trials when the *critical item was not present* on the word list.

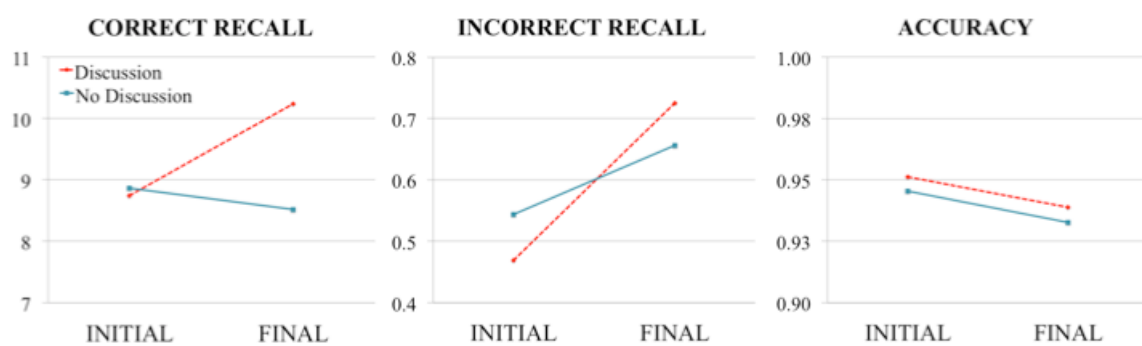

Figure 4-2. Mean number of correct items recalled, incorrect items recalled, and recall accuracy during initial and final recall for discussion and no-discussion pairs on trials when the *critical item was present* on the word list.

Critical Item Not Presented: Discussion Pairs

Exposure Items. The results from Experiment 3 are consistent with those from Experiment 1 and Experiment 2. During discussion, individuals heard more correct exposure items than incorrect exposure items (t (159) = 15.20, $p < .001$, $r = .77$). The accuracy of these exposure items was lower than the accuracy of participants' initial recall (t (159) = -2.14, $p = .034$, $r = -.17$).

Incorporation of Exposure Items. Participants incorporated a higher proportion (R_I / R_E) of correct exposure items (.72) than incorrect exposure items (.52) into their final recall (t (68) = 3.13, p = .003, r = .35).[18] The accuracy (Σ [Correct$_i$ / (Correct$_i$ + Incorrect$_i$)] / N) of the incorporated items (.88) was higher than the accuracy of the exposure items (.84), t (149) = 3.29, p = .001, r = .26, and only slightly higher than the accuracy of participants' initial recall (.87), t (149) = 0.56, p = .580, r = .05. These results suggest that during discussion participants were able to accurately make a distinction between the correct and incorrect exposure information.

A critical item can only be an exposure item if one person recalled it, but not both. This occurred in 41 of the 80 (51.3%) participant pairs. Sixty-six percent of the 41 individuals exposed to the incorrect critical item by their partner included it in their final recall following discussion. This result supports the prediction that the critical item would become contagious under conditions of exposure, and suggests that highly associated errors (e.g., the critical item) are very likely to become contagious during a discussion.

Change Across Recalls. When the critical item was not presented in the study list, participants recalled significantly more correct items, t (159) = 14.33, p < .001, r = .75, and incorrect items, t (159) = 3.25, p = .001, r = .25, in their final recall compared to their initial recall. The accuracy of final recall (.8688) was nearly identical to the accuracy of

---

[18] The analysis of the proportion of exposure items incorporated requires that a participant be exposed to at least one correct and one incorrect item during discussion. As a result, the analysis was preformed on a subset of the sample, and the proportions presented in the text cannot be calculated using the mean values of exposure and incorporation presented in Table 4-1.

initial recall (.8696), t (159) = -.102, p = .919, r = -.008.[19] This near zero change in accuracy is consistent with the results from the previous experiments suggesting that the exposure to less accurate information and incorporation of higher accuracy information results in a near zero change in accuracy.

Critical Item Not Presented: No-Discussion Pairs

When participants did not discuss their recall but reported it to a microphone, they recalled significantly fewer correct items (t (159) = -5.65, p < .001, r = -.41) during final recall relative to initial recall. This result is not consistent with the results of Experiments 1 and 2 which both showed increases in correct recall from initial to final recall. It is possible that when participants reported their initial recall aloud, not in a social context, they focused their attention on the correct items they were more confident in and allocated less attention to correct items they were uncertain of. To the extent that this occurs, participants may spend more time rehearsing correct information that they are certain of and less time rehearsing information they are uncertain of, resulting in the latter items being forgotten during subsequent memory tests. Additionally, each word list in Experiment 3 was longer (15-items) than each category list used in Experiment 2 (4-items). Even though Experiment 2 presented participants with 48 words initially, rather than just 15, the short category lists may have produced better organization and storage in memory than the longer word lists in Experiment 3. To the extent that information was less well organized in memory for Experiment 3, participants may have been less

[19] Test-retest reliabilities for correct recall, incorrect recall, and accuracy between initial recall and final recall are found in Appendix I for the critical item not presented discussion and no discussion trials.

successful at using previously recalled items as cues for additional recall of correct information.

Consistent with the previous experiments, no-discussion participants showed a significant increase in the number of incorrect items recalled, t (159) = 2.39, p = .018, r = .19, during final recall relative to initial recall. This result suggests that the effect of self-cuing still operated for incorrect information. The decrease for correct recall and increase for incorrect recall combined to produce a small but significant decrease in accuracy, t (159) = -3.13, p = .002, r = -.24.

Critical Item Not Presented: Discussion Pairs Compared to No-Discussion Pairs

A comparison of difference scores between the discussion and no-discussion groups on trials when the word list did not include the critical item examined the effect of discussion relative to self-cuing. On discussion trials participants showed a significantly greater increase for correct item recall (M = 1.64), t (159) = 15.45, p < .001, r = .77, relative to the no-discussion trials (M = -0.33), providing evidence that discussion benefits correct recall beyond the opportunity to self-cue memory. The discussion group showed only a slightly greater increase of incorrect information compared to the no-discussion group (t (159) = 1.47, p = .143, r = .12). Although the difference is small and non-significant it is in the predicted direction. That is discussion groups should show a greater increase of incorrect information as a result of exposure during discussion. The very small decrease in accuracy for the discussion trial (M = -.0007) was marginally less than the slight decrease in accuracy for the no-discussion trial (M = -.0160), t (159) = 1.75, p = .082, r = .14.

Accuracy Change Within Pairs: Critical Item Not Presented

The results of the current experiment are consistent with Experiment 1 and Experiment 2 in providing evidence that suggests discussion produces a near zero change in accuracy. Additional analyses examined the extent to which accuracy change occurred within pairs. Figure 4-3 and Figure 4-4 depict the mean accuracy for initial and final recall for the better and worse participant in each discussion pair and no-discussion pair respectively. Pairs of participants with the same initial accuracy were excluded from the following analyses.

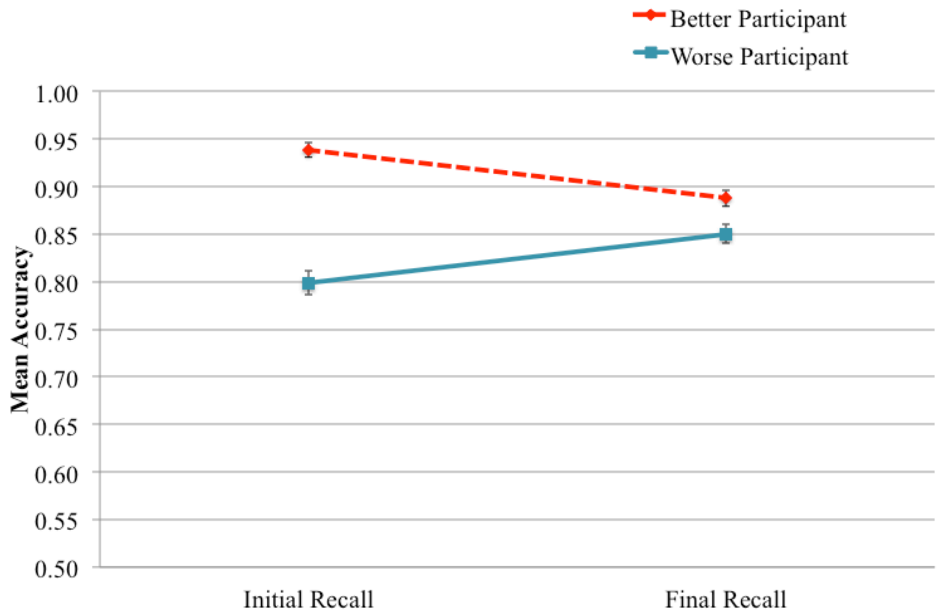

Figure 4-3. Mean initial and final accuracy for the better and worse discussion participants when the *critical item was not presented* on the word list.

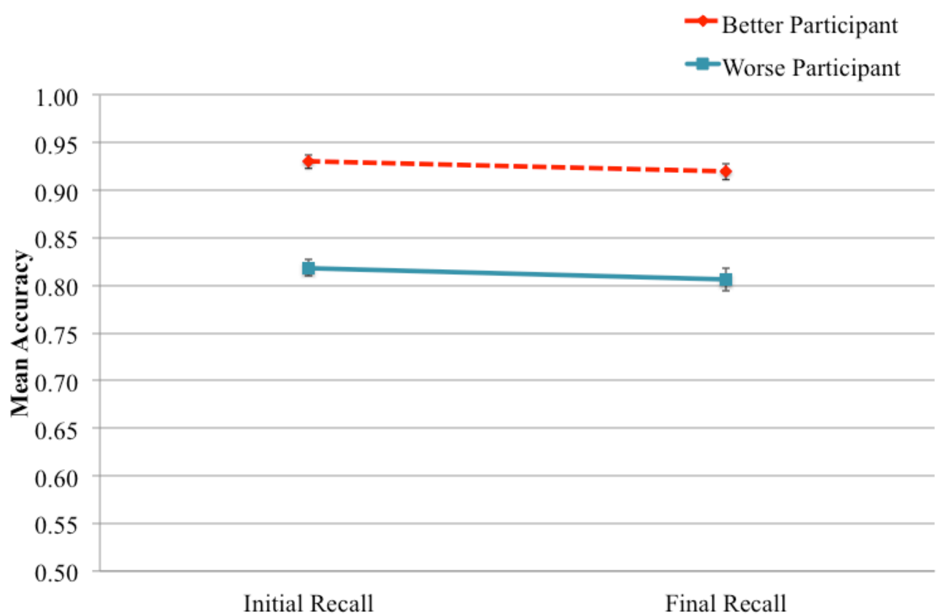

Figure 4-4. Mean initial and final accuracy for the better and worse no-discussion participants when the *critical item was not presented* on the word list.

Analysis of accuracy change within discussion pairs revealed the same accuracy redistribution effect observed in Experiments 1 and 2. Figure 4-3 shows that within each pair that discussed their recall when the critical item was not presented, the more accurate participant showed a decline in accuracy from initial to final recall (M = -.051), whereas the less accurate participant showed an increase in accuracy from initial to final recall (M = .052), t (76) = 8.24, p < .001, r = .69. Although the gap in accuracy between the initially more accurate and less accurate participant decreased following discussion, the initially more accurate participant remained more accurate than the initially less accurate participant, t (76) = 4.07, p < .001, r = .42. This evidence provides an additional caution for individuals working together to remember information, that is, be careful with whom you chose to collaborate. While discussion may increase the accuracy for the initially less accurate pair member, it may result in decreased accuracy for the initially more accurate participant.

The same comparison for the no-discussion trial showed a different pattern of results consistent with Experiment 2. Figure 4-4 shows that both the initially more accurate participant (M = -.010) and initially less accurate participant (M = -.012) showed decreases in accuracy (t (72) = .196, p = .845, r = .02). For the no-discussion trial, the initially more accurate participant remained more accurate than the initially less accurate participant, t (72) = 8.22, p < .001, r = .70. A paired samples comparison showed that the patterns of gains and losses were significantly greater for discussing trials compared to non-discussing trials when the critical item was not presented, t (70) = 6.71, $p < .001$, r = .63.[20]

Critical Item Presented: Discussion Pairs

Exposure Items. The analyses follow the same order as those presented for the cued recall condition and are generally consistent with those findings. When people discussed their recall with each another, the participant-generated exposure items included more correct than incorrect items (t (159) = 15.78, $p < .001$, r = .78), and these exposure items had a lower level of accuracy compared to initial recall (t (158) = -3.44, p = .001, r = -.26).

Incorporation of Exposure Items. These same participants incorporated a higher proportion (R_I / R_E) of correct exposure items (.74) than incorrect exposure items (.55) into their final recall (t (39) = 2.54, p = .015, r = .38).[21] The accuracy (Σ [$Correct_i$ / ($Correct_i$ + $Incorrect_i$)] / N) of the incorporated items (.91) was higher than the accuracy of the exposure items (.88), t (150) = 2.44, p = .016, r = .20, and marginally lower than the accuracy of participants' initial recall (.95), t (150) = -1.96, p = .052, r = -.16. These results are generally consistent with the pattern of results observed for the trial when the critical item was not presented. Because initial accuracy was very high for participants, it is not surprising that the accuracy of incorporated items was slightly lower. This pattern

---

[20] This paired samples *t*-test was computed using difference scores calculated by subtracting the initially less accurate participant's accuracy change from the initially more accurate participant's accuracy change for the discussion and no discussion pairs.
[21] The analysis was preformed on a subset of the sample that was exposed to at least one correct and incorrect item during discussion. The proportions presented in the text cannot be calculated using the mean values of exposure and incorporation presented in Table 4-2.

was not observed for the condition where the critical item was not presented; however in that condition initial accuracy was much lower.

Change Across Recalls. Following discussion, participants recalled significantly more correct items, t (159) = 13.16, $p < .001$, $r = .72$, and incorrect items, t (159) = 3.82, $p = .001$, $r = .29$, in their final recall compared to their initial recall. The increase in correct and incorrect information following discussion resulted in a small but significant decrease from initial recall (.951) to final recall (.939) (t (159) = -2.22, $p = .028$, $r = -.17$).[22]

Critical Item Presented: No-Discussion Pairs

Consistent with the results from the trial when the critical item was not presented participants showed a statistically significant decrease in the number of correct items recalled, t (159) = -5.54, $p < .001$, $r = -.40$, and a significant increase in the number of incorrect items recalled, t (159) = 3.17, $p = .002$, $r = .24$, for final recall relative to initial recall. These results provide additional evidence of the negative effects of reporting ones' initial recall aloud to a non-social source (e.g., a microphone). In the current experiment when participants report their initial recall aloud they may second guess some of the items initially recalled, they may focus attention and rehearsal strategies on specific well-remembered items, and this process may bring to mind new highly familiar, but incorrect pieces of information. The decrease of correct item recall and increase of incorrect item recall combined to produce a small but significant decrease in accuracy, t

[22] Test-retest reliabilities for correct recall, incorrect recall, and accuracy between initial recall and final recall are found in Appendix I for the critical item presented discussion and no discussion trials.

(159) = -3.68, *p* < .001, *r* = -.28. These results along with those obtained in Experiment 2 for no-discussion pairs suggest that self-cuing through re-exposure to one's previous recall decreases accuracy.

Critical Item Presented: Discussion Pairs Compared to No-Discussion Pairs

Consistent with Experiments 1 and 2, during discussion trials when the critical item was presented participants showed a significantly larger increase for correct item recall (*M* = 1.49), relative to the no-discussion trials (*M* = -0.34), *t* (159) = 14.96, *p* < .001, *r* = .76. There was a marginally significant difference for the increase of incorrect information between the discussion (*M* = 0.25) and no-discussion trials (*M* = 0.11), *t* (159) = 1.90, *p* = .059, *r* = .15. This larger increase for incorrect information observed for the discussion pairs relative to the no-discussion pairs is consistent with the pattern of results observed in Experiment 1. Taken together these results provide additional evidence that the effect of discussion on later memory reports is greater than self-cuing opportunities alone. The decrease in accuracy for the discussion group (*M* = -.0125) was not statistically different from the decrease in accuracy for the no-discussion group (*M* = -.0127), *t* (159) = .026, *p* = .979, *r* = .002. This result is consistent with the previously described results suggesting that the social facilitation of memory involved in discussion does not facilitate a greater change in overall accuracy than the opportunity to self-cue one's memory through re-exposure to initial recall.

Accuracy Change Within Pairs: Critical Item Presented

The decrease in accuracy for the discussion group and no-discussion group was small and only significant for the no-discussion group. Consistent with the previous experiments, additional analyses examined whether there were greater changes in accuracy within pairs. Figure 4-5 and Figure 4-6 depict the mean accuracy for initial and final recall for the better and worse participant in each discussion pair and no-discussion pair respectively. Pairs of participants with the same initial accuracy were excluded from the following analyses.

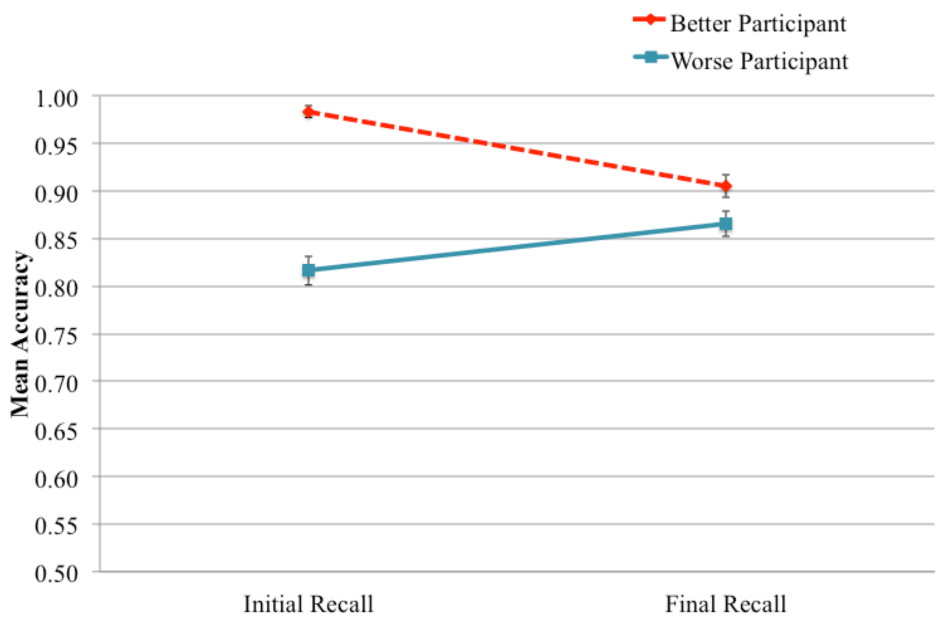

Figure 4-5. Mean initial and final accuracy for the better and worse discussion participants when the *critical item was presented*.

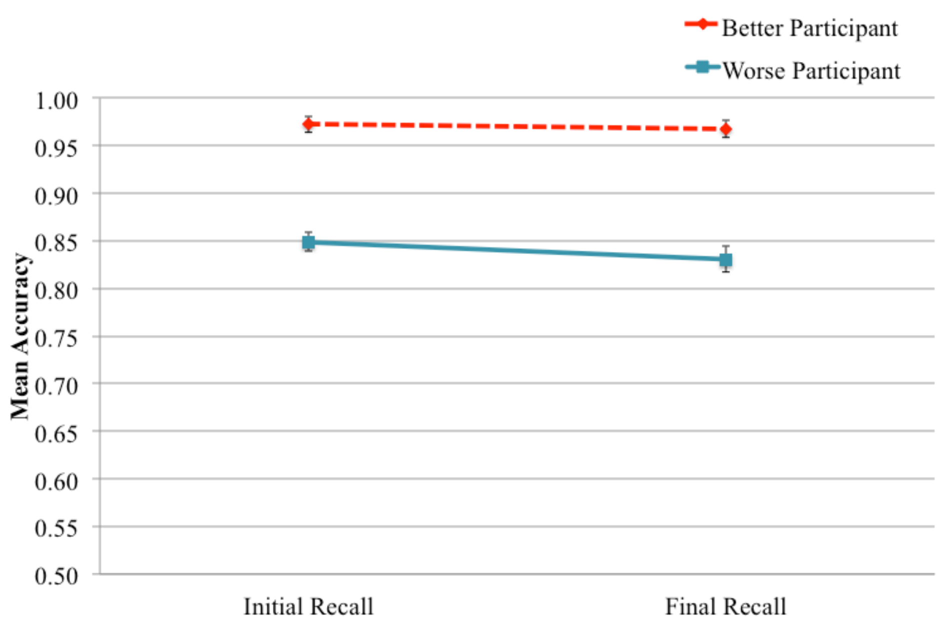

Figure 4-6. Mean initial and final accuracy for the better and worse no-discussion participants when the *critical item was presented*.

Analysis of accuracy change within discussion pairs revealed the same accuracy redistribution effect observed in Experiment 1, Experiment 2, and the critical item not presented trial. Figure 4-5 clearly shows that within each pair that discussed their recall when the critical item was presented, the more accurate participant showed a decline in accuracy from initial to final recall (M = -.078), whereas the less accurate participant showed an increase in accuracy from initial to final recall (M = .049), t (38) = 8.24, p < .001, r = .80. Although the gap in accuracy between the initially more accurate and less accurate participant decreased following discussion, the initially more accurate participant remained more accurate than the initially less accurate participant, t (38) = 2.77, p = .009, r = .41.

The same comparison for the no-discussion trial when the critical item was presented showed a pattern of results consistent with Experiment 2 and the critical item not presented trial. Figure 4-6 shows there was a very small observed decrease in accuracy for the initially more accurate participant (M = -.005) and a small decrease in accuracy for the initially less accurate participant (M = -.018), t (48) = 1.42, p = .161, r = .20. Because the decrease in accuracy for the better participant was so small, the initially more accurate participant on the no-discussion trial remained more accurate than the initially less accurate participant, t (48) = 10.59, p < .001, r = .48. The patterns of gains and losses were significantly greater for discussion trials compared to no-discussion trial when the critical item was presented, t (25) = 7.21, p < .001, r = .82.[23]

[23] This paired samples *t*-test was computed using difference scores calculated by subtracting the initially less accurate participant's accuracy change from the initially more accurate participant's accuracy change for the discussion and no discussion pairs.

Critical Item Not Presented compared to Critical Item Presented

Consistent with Experiment 2 the comparison of discussion trials when the word list did not include the critical item and when the word list did include the critical item showed that discussion did not affect post-discussion recall differently for the two conditions. That is when participants discussed their initial recall, they showed similar increases in the number of correct items recalled, similar increases in the number of incorrect items recalled, and similar changes in accuracy during post-discussion recall, regardless of whether the word list included the critical item. The specific statistical analyses that lead to this effect are presented below.

Initial Recall. During the discussion trials when the critical item was not presented participants initially recalled significantly more incorrect items, t (159) = 8.10, $p < .001$, $r = .54$, than when the critical item was presented. Additionally, when participants were not presented with the critical item they initially recalled significantly fewer correct items, t (159) = -3.69, $p < .001$, $r = -.28$, than when the critical item was presented. Initial accuracy was significantly lower during discussion trials when the critical item was not presented (.87), t (159) = -8.00, $p < .001$, $r = -.54$, than when the critical item was presented (.95), suggesting that not providing the critical item during study results in a significant decrease in recall accuracy.

The current results are driven by the production of the critical item and are completely dependent on the specific design of the stimulus materials. When the word list increased the production of errors by not including the critical item, 111 out of 160 (69.4%) participants produced the critical item as an incorrect item. However, on trials when the DRM list included the critical item, 151 out of 160 (94.4%) participants correctly produced the critical item. This suggests that the initial production of the critical item is common, likely because each item in the word list is highly associated to the critical item. Because the production of the critical item is common during recall, regardless of presentation, the final result of initial recall depends on whether or not the stimulus was designed to intentionally produce errors. In other words, for trials where the critical item was not presented participants produce more incorrect information, less correct information, and have lower recall accuracy than when the critical item was presented.[24]

The no-discussion group showed the same pattern of results. During initial recall for the no-discussion trials when the critical item was not presented participants recalled significantly more incorrect items, t (159) = 7.14, $p < .001$, $r = .49$, and significantly fewer correct items, t (159) = -4.01, $p < .001$, $r = -.30$, than when the critical item was presented. Initial accuracy was significantly lower during no-discussion trials when the critical item was not presented (.88), t (159) = -7.59, $p < .001$, $r = -.52$, than when the critical item was presented (.95). Additionally, when the word list increased the production of errors by not including the critical item, 105 out of 160 (65.6%) participants produced the critical item as an incorrect item. However, on trials when the

[24] When the participants' inclusion of the critical item was removed from analyses for both trials where the critical item was initially absent and trials where it was present, there were no significant differences between condition in the amount of correct or incorrect information produced, and participants showed nearly identical initial recall accuracy.

DRM list included the critical, 138 out of 160 (86.3%) participants correctly produced the critical item.

Exposure Items. The prediction that discussion participants would hear more incorrect exposure information during trials when the critical item was not presented was supported by the current results. During discussion, individuals not presented with the critical item were exposed to significantly more incorrect items, t (159) = 3.49, p = .001, r = .27, than trials when the critical item was presented. This result suggests that when the critical item was not presented, at least one of the pair members was likely to incorrectly include the critical item in his or her initial recall and later expose his or her partner to the incorrect detail during discussion. As previously mentioned, for trials where the critical item was not presented the initial inclusion of the critical item as an error was high. In many cases, both participants recalled the critical item incorrectly, however, there were 41 instances where only one member of the pair initially recalled the incorrect critical item and later exposed his or her partner to that item in addition to any other errors that may have been produced. This suggests that although the inclusion of the critical item as an error was initially high, there was still some opportunity for it to serve as an incorrect exposure item.

Participants experienced an equivalent amount of correct exposure information regardless of whether the critical item was presented (t (159) = -.077, p = .939, r = -.006). Although participants initially recalled more items that were correct when the critical item was presented, they were not exposed to more items that were correct during discussion. This occurred because when the critical item was presented nearly all participants (151 out of 160) initially reported the critical item. In other words because the increased correct output during initial recall for the trials when the critical item was presented was a direct result of the production of the critical item, and because almost all participants reported the critical item initially, they did not stand to benefit from increased exposure to the additional correct information during discussion.

The accuracy of exposure items when the critical item was not presented (.82) was significantly lower than the accuracy of exposure items when the critical item was presented (.88), t (159) = -2.65, p = .009, r = -.21.

Incorporation of Exposure Items. The prediction that participants would incorporate significantly more incorrect information when the critical item was not presented was not supported. Participants incorporated a slightly smaller proportion (R_I / R_E) of incorrect exposure items when the critical item was not presented (.47) than when the critical item was presented (.51), t (26) = -.322, p = .750, r = -.06.

Participants incorporated a slightly smaller proportion of correct exposure items when the critical item was not presented (.72) than when the critical item was presented (.74), t (150) = -.547, p = .585, r = -.04. This result suggests that the similar levels of correct exposure information resulted in nearly equivalent incorporation rates for correct information. The accuracy of addition items when the critical item was not presented (.88) was only slightly lower than the accuracy of addition items when the critical item was presented (.91), t (143) = -1.21, p = .228, r = -.10.

Final Recall. During final recall for the discussion trial when the critical item was not presented participants recalled significantly fewer correct items, t (159) = -2.91, p =

.004, $r = -.22$, and significantly more incorrect items, $t\ (159) = 8.10, p < .001, r = .54$, than when the critical item was presented. It is important to note that discussion did not impact the trials when the critical item was not presented and when the critical item was presented differentially in terms of increased output. That is, the increased number of items recalled following discussion was not significantly different when the critical item was not presented and when it was presented for correct information ($t\ (159) = .983, p = .327, r = .08$) or incorrect information ($t\ (159) = -.149, p = .882, r = -.01$).

The prediction that accuracy would decrease more on discussion trials when the critical item was not presented than when it was presented was not supported by the current data. Following discussion participants showed a slightly smaller decrease in accuracy when the critical item was not presented than when the critical item was presented ($t\ (159) = 1.39, p = .168, r = .11$). Final accuracy however, was significantly lower during discussion trials when the critical item was not presented (.87), $t\ (159) = -8.81, p < .001, r = -.57$, than when the critical item was presented (.94).

The same pattern of results emerged for participants during the no-discussion trials. During final recall for the no-discussion trials when the critical item was not presented, participants recalled significantly fewer correct items, $t\ (159) = -3.80, p < .001, r = -.29$, but significantly more incorrect items, $t\ (159) = 6.21, p < .001, r = .44$, than when the critical item was presented. Final accuracy was significantly lower during discussion trials when the critical item was not presented (.87), $t\ (159) = -6.56, p < .001, r = -.46$, than when the critical item was presented (.93).

The results from Experiment 3 are consistent with the results from Experiments 1 and 2: Exposure and incorporation of correct information were greater than incorrect information across conditions. The central prediction that errors would increase and accuracy would decrease following discussion for a word list that did not include the critical item was only partially supported: Errors increased following a discussion, but there was virtually no change in accuracy. These results of course depend on the type of errors made by each participant during initial recall.

Errors are more likely to be propagated by group processes to the extent that they are more likely to occur at the individual level (Koriat, 2012; Thorley & Dewhurst, 2007). The social propagation of error, however, may be curvilinear: On the one hand, if the error rates at the individual level are extremely low, there are no errors to propagate. On the other hand, if the error rates are extremely high, then all of the individuals within a group may make a particular error, in which case no individual can be viewed as having acquired it from any other individual. If however, the error rates are moderate and one participant, but not the other produces the error during pre-discussion recall, the social propagation of error is likely to occur because exposure to error occurs during the discussion. In other words the likelihood that the stimulus materials produce the error for one participant, but not the other during pre-discussion recall determines the post-discussion changes of incorrect recall and accuracy.

Chapter 5 – General Discussion

The general discussion is divided into four sections: the current state of collaborative memory research, a summary of the results from the current experiments, the implications these results, and future research directions.

Current State of Collaborative Memory Research

Collaborative memory research has predominantly focused on the negative effects of group remembering, specifically how collaborative (social) memory can be less accurate than individual memory. Previous research shows collaboration can impair memory by limiting group output through retrieval disruption (Basden et al., 1997), by altering one's memory through socially induced forgetting (Comen et al., 2009), through the social transmission of contagious errors (Roediger et al, 2000), and through group pressures to conform (Reysen, 2005; 2007). Some research however, shows collaboration can benefit later memory performance by providing re-exposure to previously studied information (Weldon & Bellinger, 1997; Blumen & Rajaram, 2008) and through error pruning during natural conversation (Ross et al., 2008). These results suggest that under certain conditions collaboration can have a negative impact on memory, while under other conditions collaboration can benefit memory performance.

Much collaborative memory research places an emphasis on the transmission of error, relying heavily on the use of experimental designs, many using confederates, to introduce error during collaborative remembering tasks. Additionally, current research has emphasized empirical phenomena, with little theoretical guidance or development. The present studies were designed to overcome these limitations by examining the transmission of both correct and incorrect information between people and to examine changes in accuracy following discussion. Importantly, the present studies did not utilize confederates to introduce error, instead examining the frequency with which people naturally transmit correct and incorrect information during discussion. The present studies also discuss the theoretical mechanisms by which memories "travel" from one person to another.

Summary of Experimental Results

The present studies addressed four important questions: 1. How accurate is the information to which people are exposed when remembering an event with another person? 2. To what extent do people incorporate correct, versus incorrect, information in their own recall following that exposure? 3. What is the net change in accuracy in final, post-discussion recall? 4. How does accuracy of initial recall predict post-discussion recall?

Regarding the first question, when people discuss their individual recall with one another, the additional information provided by each person (i.e., the exposure items) includes far more correct than incorrect information. However, the accuracy of exposure items is lower than the accuracy of individuals' initial recall. This pattern of results is consistent across all three experiments. Even when the recall task was designed to produce more intrusion errors (e.g., using a cued recall test rather than a non-cued recall) or when the stimulus materials were designed to produce intrusion errors (e.g., using DRM wordlists where a critical item highly associated with all of the list items was not presented) participants were exposed to more correct than incorrect information.

Regarding the second question, people are more likely to incorporate the correct information than the incorrect information they were exposed to during discussion with a partner. This effect was consistent across all three experiments and conditions within Experiment 2 (cue vs. non-cued recall) and Experiment 3 (critical item presented vs. not presented).

These results suggest that the lower accuracy of the exposure information, combined with the higher accuracy for the incorporation of exposure information, produced a zero net gain or loss in the accuracy of final recall compared to initial recall.

Regarding the third question, the net effect of the combination of lower accuracy of exposure items and higher accuracy in the incorporation of exposure items into final recall results in a general increase in the number of correct and incorrect items recalled, with a net change in the accuracy of recall that was at or near zero. Although the decrease in accuracy for the discussion group in Experiment 3 (when the critical item was presented) was significant, a Chi-square test for heterogeneity (Rosenthal, 1991) showed that the variation in effect sizes was not statistically significant across experiments, [*r*'s = -.007 (Exp 1), .081 (Exp 2 cued recall), .260 (Exp 2 non-cued recall), -.008 (Exp 3 critical item not presented), -.173 (Exp 3 critical item presented)], $\chi^2(4) = 7.12$, $p = .130$.

Although accuracy changed very little following discussion (comparing final to initial recall), there were important changes within each pair of participants. Specifically, an accuracy redistribution effect was observed in all three experiments, such that the more accurate person became less accurate and the less accurate person became more accurate as a result of their information exchange. There is an important practical lesson in these results: Be careful of who you collaborate with. Collaborating with someone whose memory is less accurate can undermine the accuracy of one's own memory.

Regarding the fourth question, increasing or decreasing initial recall accuracy, through recall procedures (Experiment 2) or stimulus materials (Experiment 3) had little effect on the changes to post-discussion recall. Regardless of whether initial recall accuracy was high or low, post-discussion increase for correct item recall, increase for incorrect item recall, and change in recall accuracy were the same.

The pattern of results obtained in all three experiments – the lower accuracy of exposure items, and the relatively higher accuracy of those items that were incorporated into individuals' final recall – are consistent with basic memory mechanisms. According to a truth in numbers heuristic, exposure items should be less accurate than initial recall given that items recalled by only one person and not the other are less accurate than items recalled by both individuals. Individuals should be able to discriminate between correct and incorrect exposure items to the extent that correct items can cue memory and are more familiar than incorrect exposure items as predicted by dual-process theories of recognition memory (Jacoby, 1991; Yonelinas, 1994; Wixted & Mickes, 2010). These basic memory mechanisms provide a basic theoretical framework for understanding how memory can be shaped through the transmission of information from one person to another during a discussion.

Implications of the Current Research

The current research provides strong evidence that there is more to collaboration than just the transmission of errors. When two people discuss a memory, they incorporate not only incorrect information from their partner but also correct information. This result was extremely robust, occurring in all three experiments and under a variety of conditions. In all three experiments, correct information was more contagious than incorrect information. Given that overall accuracy showed very little change following discussion, these results suggest that the benefits of collaboration may outweigh the costs of collaboration. However, it is important to be cautious when interpreting these results because the overall cost/benefit of collaboration may depend on the consequences of the specific commission errors (i.e., incorporating incorrect exposure information) and omission errors (i.e., failing to incorporate correct exposure information) made.

Although Person A may only incorporate one incorrect detail from Person B, that incorrect detail may have a greater impact than the numerous correct details Person A obtained from Person B. Imagine that two people witness a crime and discuss their memory for the event before the police arrive. Witness B may incorrectly remember and tell Witness A that the assailant was wearing a blue baseball cap. If both witnesses seem to independently provide corroborating information regarding the blue baseball cap, the police may use this incorrect piece of information when looking for the assailant. Although Witness A may have a generally accurate memory for the assailant and may have acquired several correct details from Witness B this one incorrect detail may be weighted more heavily during the police investigation.

On the other hand Person A may fail to incorporate a correct detail from Person B that may have greater impact on memory than the incorrect details incorporated from Person B. Imagine that two people engage in discussion following the interview of several possible job candidates. During the discussion, Interviewer B may correctly describe one of the candidates as being more skilled at the tasks necessary for the position than the other candidates. If Interviewer A fails to incorporate this very important piece of information he or she may recommend a job candidate that is less qualified for the position than someone else. Although Interviewer A may have generally accurate memories of the people interviewing for the position, failing to incorporate this one critical detail may be more detrimental to the hiring process than the incorporation of minor incorrect details.

The current results suggest that following discussion people's memories are no less trustworthy. Therefore, when two people collaborate and then share information with a third party, it is up to this third party to determine how much weight to give the memory reports of Person A and Person B. This burden on the third party is important. The specific commission and omission errors made by the collaborating partners may negatively influence important decisions that rely on this information.

Additionally, when the examination of accuracy change was within pairs, the initially more accurate person decreased in accuracy following discussion and the initially less accurate person increased in accuracy following discussion. This suggests that within a pair, collaboration has a simultaneously positive and negative effect. This should suggest a measure of caution to those groups likely to work collaboratively (e.g., students). To the extent that a student is more prepared than his or her study partner(s), he or she may suffer the negative effects of collaboration, demonstrating decreased accuracy following exposure to less accurate information from his or her partner(s). However, the less prepared student(s) may benefit from collaboration through exposure to information that is more accurate. The important lesson here is that one should always be careful when choosing with whom to collaborate.

Future Directions

The net gain or loss in accuracy depends on the particular blend of correct and incorrect information to which a person is exposed, as well as the various factors that determine whether information is incorporated into a person's subsequent report. There are a number of factors likely to alter the mix of correct and incorrect information, and hence overall accuracy. There is no reason to think that the zero (or near-zero) gain result obtained here is carved in stone. Therefore, it is important for future research to explore the boundaries of the current results suggesting discussion in a social contagion paradigm produces little to no change in net accuracy. For example, previous studies have shown that errors increase when individuals within a group are required to recall items through turn-taking (Thorley & Dewhurst, 2007). However, other studies have shown that errors decrease when individuals within a group can engage in a free-flowing discussion to resolve disagreements (Barber, Rajaram, & Aron, 2010; Ross, et al., 2008; Thorley & Dewhurst, 2007), or are required to reach a consensus (Harris, Barnier, & Sutton, 2012). Future research should systematically examine these different styles of collaboration effort to gain a better understanding of the specific situations that may produce a net gain or loss in post-discussion recall accuracy.

It is also important to note that there are social factors not directly tested in the current experiment that may influence the incorporation of exposure information. When two people discuss their memories' they provide additional information beyond item recall. This information may be in the form of head nods, hand gestures, or the expression of certainty in their voice. It is therefore, important for future research to examine the influence of socially relevant information during a discussion, because this information may facilitate the transfer of information between participants. If participants simply read another participants' recall rather than discussing the information with him or her, and the observed results persist, it would suggest the effect is driven by the mere exposure to another participants' recall, rather than the social interaction. If however the facilitation of memory was smaller when the social interaction is removed, it would provide evidence that the social information expressed during discussion is an important factor influencing the incorporation of information from another person. Assuming that the social interaction is important beyond simply exposure to information, the confidence of recall expressed during vocalization could cause the listener to incorporate items expressed with certainty and reject items expressed with hesitation by his or her partner.

People tend to express more confidence in their correct than their incorrect recall, and listeners may consider such information when deciding whether or not to incorporate information from another person into their own memory recall (Robinson & Johnson, 1996). Future research is necessary to examine the relationship between expressed confidence and memory incorporation during discussion. To the extent that confidence is important when obtaining information from another individual, participants should incorporate items expressed with higher confidence and reject items expressed with lower confidence. Given that people are generally more confident in their correct than incorrect recall people should incorporate more correct than incorrect exposure items into post-discussion recall.

The relationship between social interaction and memory accuracy is complex and likely to vary with a number of factors including the recall task, the stimulus materials, the structure of discussion, and confidence. While the current studies provide evidence that collaboration is more that just people swapping errors (people also swap correct information), future studies are necessary to examine the situations that produce an overall net gain or loss in post-discussion accuracy.

References

Allan, K., & Gabbert, F. (2008). I still think it was a banana: Memorable ‘lies’ and forgettable ‘truths’. *Acta Psychologica, 127*, 299-308. doi: 10.1016/j.actpsy.2007.06.001

Barber, S. J., Rajaram, S., & Aron, A. (2010). When two is too many: Collaborative encoding impairs memory. *Memory & Cognition, 38*, 255-266. doi: 10.3758/MC.38.3255

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, UK: Cambridge University Press.

Basden, B. H., Basden, D. R., Bryner, S., & Thomas, R. L. III (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23*, 1176-1189. doi: 10.1037/0278-7393.23.5.1176

Basden, B. H., Basden, D. R., Thomas, R. L. III, Souphasith, S. (1998). Memory distortion in group recall. *Current Psychology, 16*, 225-246.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut norms. *Journal of Experimental Psychology, 80*, 1-46.

Blumen, H. M., & Rajaram, S. (2008). Influence of re-exposure and retrieval disruption during group collaboration on later individual recall. *Memory, 16*, 231-244. doi: 10.1080/09658210701804495

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*, 207-230.

Chan, J. C. K., Thomas, A. K., & Bulevich, J.B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing event. *Psychological Science*, *20*, 66-73. doi:10.1111/j.1467-9280.2008.02245.x

Clark, S. E., Abbe, A., Larson, R. P. (2006). Collaboration in associative recognition memory: Using recalled information to defend "new" judgments. *Journal of Experimental Psychology, 32*, 1266-1273. doi: 10.1037/0278-7393.32.6.1266

Clark, S. E., Hori, A., Putnam, A., & Martin, T. P. (2000). Group collaboration in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 1578-1588.

Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the unforgettable through conversation: Socially shared retrieval-induced forgetting of September 11 memories. *Psychological Science, 20*, 627-633.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17-22.

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not for words in multiple recall trials. *Cognitive Psychology, 6*, 159-171. doi: 10.1016/0010-0285(74)90008-5

Finlay, F., Hitch, G. J., & Meudell, P. R. (2000). Mutual inhibition in collaborative recall: Evidence for a retrieval-based account. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 1556-1567.

Funder, D. C. (1986). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101*, 75-90.

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology, 17*, 533-543. doi: 10.1002/acp.885

Gabbert, F., Memon, A., Wright, D. B. (2006). Memory conformity: Disentangling the steps toward influence during a discussion. *Psychonomic Bulletin & Review, 13*, 480-485.

Harris, C. B., Barnier, A. J., & Sutton, J. (2012). Consensus collaboration enhances group and individual recall accuracy. *The Quarterly Journal of Experimental Psychology, 65*, 179-194.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*, 681-706.

Koriat, A. (2012). When are two heads better than one and why? *Science, 336*, 360-362. doi: 10.1126/science.1216549

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*, 297-315.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541. doi: 10.1016/0749-596X(91)90025-F

Johansson, O., Andersson, J., & Rönnberg, J. (2000). Do elderly couples have a better prospective memory than other elderly people when they collaborate? *Applied Cognitive Psychology, 14*, 121-133.

Latané, B., Williams, K., Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 822-832.

Loftus, E. F., & Greene, E. (1980). Warning: Even memory for faces may be contagious. *Law and Human Behavior, 4*, 323-334. doi: 10.1007/BF01040624

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, 19-31. doi:10.1037/0278-7393.4.1.19

Maki, R. H., Weigold, A., & Arellano, A. (2008). False memory associated word lists in individuals and collaborating groups. *Memory & Contagion, 36*, 598-603. doi: 10.3758/MC.36.3.598

McDaniel, M. A., Moore, B. A., & Whiteman, H. L. (1998). Dynamic changes in hypermnesia across early and late tests: A relational/item-specific account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 173-185. doi: 10.1037/0278-7393.24.1.173

Meade, M. L., & Roediger, H. L., III. (2002). Explorations in the social contagion of memory. *Memory & Cognition, 30*, 995-1009.

Meade, M. L., Watson, J. M., Balota, D. A., & Roediger, H. L. III (2007). The role of spreading activation and retrieval mode in producing false recognition in the DRM paradigm. *Journal of Memory and Language, 56*, 305-320.

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, 48A*, 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, 6*, 525-543.

Perlmutter, H. V., & deMontmollin, G. (1952). Group learning of nonsense syllables. *The Journal of Abnormal Social Psychology, 47*, 762-769.

Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory, and the eyewitness confidence-accuracy correlation. *Journal of Applied Psychology*, *81*, 587-594. doi: 10.1037/0021-9010.81.5.587

Roediger, H. L. III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803-814.

Roediger, H. L., III, Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin & Review, 3*, 481-485.

Roediger, H. L., III, Payne, D. G. (1982). Hypermenisa: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 66-72. doi: 10.1037/0278-7393.8.1.66

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Ross, M., Spencer, S. J., Blatz, C. W., & Restorick, E. (2008). Collaboration reduces the frequency of false memories in older and younger adults. *Psychology and Aging*, *23*, 85-92. doi: 10.1037/0882-7974.23.1.85

Reysen, M. B., (2005). The effects of conformity on recognition judgments. *Memory, 13*, 87-94. doi: 10.1080/09658210344000602

Reysen, M. B., (2007). The effects of social pressure on false memory. *Memory & Cognition, 35*, 59-65.

Schneider, D. M., & Watkins, M. J. (1996). Response conformity in recognition testing. *Psychonomic Bulletin & Review, 3*, 481-485.

Stadler, M. A., Roediger, H. L. III, & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*, 494-500.

Thorley, C., & Dewhurst, S. A. (2007). Collaborative false recall in the DRM procedure: Effects of group size and group pressure. *European Journal of Cognitive Psychology, 19*, 867-881. doi: 10.1080/09541440600872068

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381-391.

Van Overschelde, J. P., Rawson, K. A., Donlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*, 289-335.

Vollrath, D. A., Sheppard, B. H., Hinsz, V. B., & Davis, J. J. (1989). Memory performance by decision-making groups and individuals. *Organizational Behavior and Human Decision Processes*, *43*, 289-300. doi: 10.1016/0749-5978(89)90040-X

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1160-1175. doi: 10.1037/0278-7393.23.5.1160

Weldon, M. S., Blair, C., & Huebsch, P. D. (2000). Group remembering: Does social loafing underlie collaborative inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1568-1577.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*, 1025-1054. doi: 10.1037/a0020874

Wright, D. B., Mathews, S. A., & Skagerberg, E. M. (2005). Social recognition memory: The effect of other people's responses for previously seen and unseen items. *Journal of Experimental Psychology: Applied, 11*, 200-209. doi: 10.1037/1076-898X.11.3.200

Wright, D. B., Self, G., & Justice, C. (2000). Memory conformity: Exploring misinformation effects when presented by another person. *British Journal of Psychology, 91*, 189-202. doi: 10.1348/000712600161781

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*, 1341-1354. doi: 10.1037/0278-7393.20.6.1341

Appendix A

Stimulus Materials for Experiment 1:
Roediger et al. (2000) Social Contagion Stimuli

Toolbox

Bathroom

Kitchen

Bedroom

Closet

Desk

Appendix B

Experiment 1: Statistical Analyses Conducted on the Pair Averages

Test	N (*df*)	*t*	*p*	*r*
Discussion Pairs				
Correct Exposure - Incorrect Exposure	50 (49)	16.08	.000	.92
Exposure Accuracy - Initial Accuracy	50 (49)	-13.93	.000	-.89
Proportion of Exposure Items Incorporated:				
Correct - Incorrect	49 (48)	6.18	.000	.67
Incorporation Accuracy - Exposure Accuracy	50 (49)	7.34	.000	.72
Incorporation Accuracy - Initial Accuracy	50 (49)	1.12	.267	.17
Correct Items: Final Recall - Initial Recall	50 (49)	18.60	.000	.94
Incorrect Items: Final Recall - Initial Recall	50 (49)	5.59	.000	.62
Accuracy: Final Recall - Initial Recall	50 (49)	-0.09	.932	-.01
No Discussion Pairs				
Correct Items: Final Recall - Initial Recall	50 (49)	3.70	.001	.47
Incorrect Items: Final Recall - Initial Recall	50 (49)	1.26	.216	.18
Accuracy: Final Recall - Initial Recall	50 (49)	0.22	.830	.03
Discussion Compared to No Discussion Pairs				
Correct Item Difference Score*	100 (98)	13.48	.000	.81
Incorrect Item Difference Score*	100 (98)	3.51	.001	.34
Accuracy Difference Score*	100 (98)	-0.21	.837	-.02

Note. (*) Indicates an independent samples *t*-test. All other comparisons are paired samples *t*-tests. Differences in N are the result of missing data. Analyses are listed in the same order as presented in the manuscript.

Appendix C

Experiment 1:Test-Retest Reliability Between Initial and Final Recall

Variable	N	*r*	*p*
Discussion Pairs			
Correct Recall	100	.875	.000
Incorrect Recall	100	.782	.000
Accuracy	100	.782	.000
No Discussion Pairs			
Correct Recall	100	.940	.000
Incorrect Recall	100	.866	.000
Accuracy	100	.911	.000

Appendix D

Stimulus Materials for Experiment 2: Category Names and List Items 1 to 4

Category Name	**List Items**
Four Footed Animals	HORSE, TIGER, ELEPHANT, MOUSE
Kitchen Utensils	FORK, SPATULA, POT, BLENDER
Fruits	ORANGE, GRAPES, PEACH, KIWI
Furniture	COUCH, DESK, DRESSER, COFFEE TABLE
Sports	SOCCER, TENNIS, SWIMMING, VOLLEYBALL
Earth Formations	OCEAN, LAKE, HILL, CANYON
Articles of Clothing	SOCKS, HAT, SHORTS, JACKET
Instruments	GUITAR, PIANO, CLARINET, VIOLIN
Birds	BLUE JAY, HAWK, CROW, PARROT
Vegetables	BROCCOLI, ONION, CORN, CELERY
Insects	SPIDER, MOSQUITO, LADYBUG, BUTTERFLY
Carpenter's Tools	NAILS, SCREWDRIVER, WRENCH, LEVEL

Note. Categories and items obtained from Battig and Montague (1969) Category Norms updated by Van Overschelde et al. (2004). Categories are listed in the order they were presented to participants. List items are presented from highest to lowest associative strength to the category.

Appendix E

Experiment 2: Statistical Analyses Conducted on the Pair Averages

Test	N (*df*)	*t*	*p*	*r*
Cued Recall: Discussion Pairs				
Correct Exposure - Incorrect Exposure	20 (19)	14.45	.000	.96
Exposure Accuracy - Initial Accuracy	20 (19)	-5.13	.000	-.76
Proportion of Exposure Items Incorporated:				
Correct - Incorrect	8 (7)	2.85	.025	.71
Incorporation Accuracy - Exposure Accuracy	20 (19)	3.83	.001	.66
Incorporation Accuracy - Initial Accuracy	20 (19)	-1.03	.317	-.27
Correct Items: Final Recall - Initial Recall	20 (19)	11.09	.000	.93
Incorrect Items: Final Recall - Initial Recall	20 (19)	0.91	.374	.21
Accuracy: Final Recall - Initial Recall	20 (19)	0.61	.550	.14
Cued Recall: No Discussion Pairs				
Correct Items: Final Recall - Initial Recall	20 (19)	2.79	.012	.54
Incorrect Items: Final Recall - Initial Recall	20 (19)	4.57	.000	.72
Accuracy: Final Recall - Initial Recall	20 (19)	-4.40	.000	-.71
Cued Recall: Discussion Compared to No Discussion Pairs				
Correct Item Difference Score*	40 (38)	8.57	.000	.81
Incorrect Item Difference Score*	40 (38)	0.99	.328	.16
Accuracy Difference Score*	40 (38)	2.25	.030	.34
Non-Cued Recall: Discussion Pairs				
Correct Exposure - Incorrect Exposure	20 (19)	14.35	.000	.96
Exposure Accuracy - Initial Accuracy	20 (19)	-4.09	.001	-.68
Proportion of Exposure Items Incorporated:				
Correct - Incorrect	6 (5)	5.25	.003	.92
Incorporation Accuracy - Exposure Accuracy	20 (19)	3.90	.001	.67
Incorporation Accuracy - Initial Accuracy	20 (19)	-2.36	.029	-.47
Correct Items: Final Recall - Initial Recall	20 (19)	11.19	.000	.93
Incorrect Items: Final Recall - Initial Recall	20 (19)	0.00	1.000	.00
Accuracy: Final Recall - Initial Recall	20 (19)	1.81	.086	.38
Non-Cued Recall: No Discussion Pairs				
Correct Items: Final Recall - Initial Recall	20 (19)	2.72	.014	.53
Incorrect Items: Final Recall - Initial Recall	20 (19)	3.51	.002	.63
Accuracy: Final Recall - Initial Recall	20 (19)	-3.80	.001	-.66
Non-Cued Recall: Discussion Compared to No Discussion Pairs				
Correct Item Difference Score*	40 (38)	8.54	.000	.81
Incorrect Item Difference Score*	40 (38)	2.08	.044	.32
Accuracy Difference Score*	40 (38)	3.49	.001	.49
Cued Recall Compared to Non-Cued Recall: Discussion Pairs **				
Correct Item Difference Score*	40 (38)	-0.75	.457	-.12
Incorrect Item Difference Score*	40 (38)	0.79	.434	.13
Accuracy Difference Score*	40 (38)	-0.64	.529	.10

Note. (*) Indicates an independent samples *t*-test. All other comparisons are paired samples *t*-tests. Differences in N are the result of missing data. Analyses are listed in the same order as presented in the manuscript. (**) Not all of the analyses comparing Cued Recall to Non-Cued Recall are presented, however all statistical comparisons made with the pair averages resulted in a larger effect size.

Appendix F

Experiment 2: Test-Retest Reliability Between Initial and Final Recall

Test	N	*r*	*p*
Cued Recall Discussion Pairs			
Correct Recall	40	.913	.000
Incorrect Recall	40	.635	.000
Accuracy	40	.641	.000
Cued Recall No Discussion Pairs			
Correct Recall	40	.970	.000
Incorrect Recall	40	.934	.000
Accuracy	40	.951	.000
Non-Cued Recall Discussion Pairs			
Correct Recall	40	.835	.000
Incorrect Recall	40	.712	.000
Accuracy	40	.777	.000
Non-Cued Recall No Discussion Pairs			
Correct Recall	40	.953	.000
Incorrect Recall	40	.936	.000
Accuracy	40	.940	.000

Appendix G

Stimulus Materials for Experiment 3: Critical Item and List Items 1 to 15

SWEET: sour, candy, sugar, bitter, good, taste, tooth, nice, honey, soda, chocolate, heart, cake, tart, **pie**

CHAIR: table, sit, legs, seat, couch, desk, recliner, sofa, wood, cushion, swivel, stool, sitting, rocking, **bench**

SMOKE: cigarette, puff, blaze, billows, pollution, ashes, cigar, chimney, fire, tobacco, stink, pipe, lungs, flames, **stain**

ROUGH: smooth, bumpy, road, tough, sandpaper, jagged, ready, coarse, uneven, riders, rugged, sand, boards, ground, **gravel**

Note. Word lists obtained from Roediger and McDermott (1995) and Stadler et al. (2005). List items are presented in order of highest to lowest associative strength to the critical item. For trials when the critical item was presented item 15 (bolded) was not included as part of the list.

Appendix H

Experiment 3: Statistical Analyses Conducted on the Pair Averages

Test	N (*df*)	*t*	*p*	*r*
Critical Item Not Presented: Discussion Pairs				
Correct Exposure - Incorrect Exposure	80 (79)	20.24	.000	.92
Exposure Accuracy - Initial Accuracy	80 (79)	3.84	.000	-.40
Proportion of Exposure Items Incorporated:				
Correct - Incorrect	6 (5)	3.25	.023	.82
Incorporation Accuracy - Exposure Accuracy	70 (69)	3.28	.002	.37
Incorporation Accuracy - Initial Accuracy	70 (69)	0.33	.742	.04
Correct Items: Final Recall - Initial Recall	80 (79)	15.66	.000	.87
Incorrect Items: Final Recall - Initial Recall	80 (79)	3.54	.001	.37
Accuracy: Final Recall - Initial Recall	80 (79)	0.12	.903	.014
Critical Item Not Presented: No Discussion Pairs				
Correct Items: Final Recall - Initial Recall	80 (79)	-5.56	.000	-.53
Incorrect Items: Final Recall - Initial Recall	80 (79)	2.08	.041	.23
Accuracy: Final Recall - Initial Recall	80 (79)	-2.73	.008	-.29
Critical Item Not Presented: Discussion Compared to No Discussion Pairs				
Correct Item Difference Score	80 (79)	18.26	.000	.90
Incorrect Item Difference Score	80 (79)	1.57	.121	.17
Accuracy Difference Score	80 (79)	1.65	.054	.22
Critical Item Presented: Discussion Pairs				
Correct Exposure - Incorrect Exposure	80 (79)	19.90	.000	.91
Exposure Accuracy - Initial Accuracy	79 (78)	-5.88	.000	-.55
Proportion of Exposure Items Incorporated:				
Correct - Incorrect	5 (4)	1.67	.170	.64
Incorporation Accuracy - Exposure Accuracy	71 (70)	2.37	.021	.27
Incorporation Accuracy - Initial Accuracy	71 (70)	-3.14	.002	-.35
Correct Items: Final Recall - Initial Recall	80 (79)	14.38	.000	.85
Incorrect Items: Final Recall - Initial Recall	80 (79)	3.92	.000	.40
Accuracy: Final Recall - Initial Recall	80 (79)	2.60	.011	.28
Critical Item Presented: No Discussion Pairs				
Correct Items: Final Recall - Initial Recall	80 (79)	-5.43	.000	-.52
Incorrect Items: Final Recall - Initial Recall	80 (79)	3.17	.002	.71
Accuracy: Final Recall - Initial Recall	80 (79)	-3.66	.000	-.82
Critical Item Presented: Discussion Compared to No Discussion Pairs				
Correct Item Difference Score	80 (79)	15.49	.000	.87
Incorrect Item Difference Score	80 (79)	2.00	.049	.22
Accuracy Difference Score	80 (79)	0.03	.975	.00
Critical Item Not Presented Compared to Critical Item Presented: Discussion Pairs **				
Correct Item Difference Score	80 (79)	1.01	.315	.11
Incorrect Item Difference Score	80 (79)	-.172	.864	-.02
Accuracy Difference Score	80 (79)	1.75	.085	.19

Note. All comparisons are paired samples *t*-tests. Differences in N are the result of missing data. Analyses are listed in the same order as presented in the manuscript. (**) Not all of the analyses comparing Cued Recall to Non-Cued Recall are presented, however all statistical comparisons made with the pair averages resulted in a larger effect size.

Appendix I

Experiment 3: Test-Retest Reliability Between Initial and Final Recall

Test	N	*r*	*p*
Critical Item Not Presented Discussion Pairs			
Correct Recall	160	.770	.000
Incorrect Recall	160	.619	.000
Accuracy	160	.634	.000
Critical Item Not Presented No Discussion Pairs			
Correct Recall	160	.917	.000
Incorrect Recall	160	.780	.000
Accuracy	160	.791	.000
Critical Item Presented Discussion Pairs			
Correct Recall	160	.723	.000
Incorrect Recall	160	.633	.000
Accuracy	160	.665	.000
Critical Item Presented No Discussion Pairs			
Correct Recall	160	.897	.000
Incorrect Recall	160	.884	.000
Accuracy	160	.889	.000