The Influence of Contradicting Implication on Inference Generation in Discourse Processing: A Phantom Recollection Approach

by

Anjum Fazaluddin

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF ARTS

Department of Psychology

University of Manitoba

Winnipeg

Abstract

Many cognitive processes influence how we encode, store, and retrieve what we read. Two processes are considered to affect retrieval from memory: recollection and familiarity. Recollection is the explicit retrieval of the context; whereas familiarity is a vaguer feeling of remembering without knowing the exact context. Many factors influence how both recollection and familiarity function. In my M.A. thesis, participants read two-sentence passages, in which the second sentence either stated an action or implied it, and the first sentence either supported or contradicted the event of the second sentence. The participants received one of three types on instructions and made recognition judgments about test sentences in the context of the prior passages. Stating an action, and moreover making it distinct due to contradiction, would lead to more accurate recall of the same test sentence due to a high influence of recollection. Implying an action would result in inference generation of the stated action. This would indicate a high influence of familiarity on recognition judgments. A multinomial model was implemented in order to estimate the relative contributions of these distinct memory processes.

Acknowledgements

I would like to thank my faculty advisor, Dr. Murray Singer, for his help and guidance toward the completion of my Master's thesis. I would also like to thank my committee members, Dr. Randy Jamieson and Dr. Terry Janzen for their valuable suggestions and support.

Table of Contents

List of Tables

List of Figures

The influence of contradicting implication on inference generation in discourse processing: A phantom recollection approach

Introduction

Understanding discourse requires that the reader be able to connect the ideas from among sentences. Readers also need to relate the current ideas with their general world knowledge from previous experiences. Sometimes a text may not explicitly state the relation between two ideas and thus an inference may be necessary in order to relate one segment of text to the next. When such an inference is needed to preserve the coherence of the message, it is called a bridging inference (Haviland & Clark, 1974; Singer, Halldorson, Lear, & Andrusiak, 1992). If the inference is a prediction or elaboration based on the preceding text, but is not required for coherence of the message, it is called an elaborative inference (Keefe & McDaniel, 1993; Potts, Keenan, & Golding, 1988; Schmalhofer, McDaniel, & Keefe, 2002; Singer, 1980; Singer & Ferreira, 1983).

The processes of reading a passage, connecting the ideas, relating the ideas to world knowledge, and generating inferences and predictions can be very complicated especially when it comes to encoding all the information. The processing takes place at several levels of text representation, which include the exact words that were read as well as the more general, inferred, and inter-connected information (Singer, 1990). One factor that influences how well any given information is encoded in memory is whether that information is distinctive or bizarre in comparison with related information (McDaniel, Dornburg, & Guynn, 2005).

Recalling information from memory is also a complex process. Generally it is agreed that there are two main, opposing memory processes involved in recognizing whether an item has been previously encountered or not: recollection and familiarity. Recollection is the explicit retrieval of remembering when and where an item was encountered before; whereas familiarity is a vaguer feeling or having encountered the item without being able to recollect the exact context (Jacoby, 1991; Mandler, 1980). More recently, researchers have begun to quantify the relative contribution of recollection and familiarity to text and discourse memory. They have done this by the procedure of multinomial modelling (Brainerd, Reyna, & Mojardin, 1999; Brainerd, Wright, Reyna, & Mojardin, 2001).

The purpose of my M.A. thesis is to combine the above mentioned discourse phenomena in order to evaluate the effects of stating directly and implying an action on peoples' memory for the action. Additionally, I will manipulate whether the stated action appears distinctive or not in its passage. Finally, I will use an experimental paradigm and a multinomial model to attempt to quantify the resulting data.

I will begin with reviews of the literature of bridging and elaborative inferences, multi-level text representations, distinctiveness, and multinomial modeling in order to introduce the main findings in these fields, as well as to lay the groundwork for my own experimental design. I will report three experiments: the first two are pilot and norming experiments and the third is the critical experiment to test my hypotheses. I will also describe the multinomial models I will be implementing on the experimental data.

Bridging and Elaborative Inferences

There has been much research in the area of inferences in discourse comprehension over the last forty years. There is general consensus that bridging inferences are required for coherence of a message and are generated routinely. Also, pertinent world knowledge is often accessed and utilized in order to consolidate these inferences (Albrecht & Myers, 1995; Keefe & McDaniel, 1993; Potts et al., 1988; Singer, 1980; Singer et al., 1992). Elaborative inferences on the other hand, are minimally inferred during reading, if at all (Keefe & McDaniel, 1993; Potts et al., 1988; Schmalhofer et al., 2002; Singer, 1980; Singer & Ferreira, 1983).

Early research suggested that bridging inferences are routinely drawn during the course of reading. It was interpreted that they are required for local and global coherence of narrative passages. This was demonstrated by facilitation in response latencies and reading times for passages requiring a bridging inference for comprehension, compared with those of control passages and passages that required elaborative inferences (Keefe & McDaniel, 1993; Potts et al., 1988; Singer, 1980; Singer et al., 1992).

Some of the first research on bridging inferences found that it took longer for readers to understand pairs of sentence that required a bridging inference than for pairs in which the relation between the two sentences was explicitly stated (Haviland & Clark, 1974). When people had read passages requiring a bridging inference, they took significantly longer to indicate they understood the passage than for passages in which the relation was explicitly stated. It was concluded that when there is a gap in coherence of two sentences, readers need to conduct a memory search to find a matching antecedent for the presented information. If the required antecedent is unavailable in memory, then the reader must build a bridging structure, or inference, in order to consolidate the gap in coherence.

Target sentences containing the to-be-inferred information from passages requiring a bridging inference were found to be recognized more accurately and more confidently than similar target sentences from passages that required elaborative inferences (Singer, 1980). Also, the recognition measures for the bridging inference condition were not significantly different from the recognition measures for directly stated passages. This was suggested to strongly support the notion that bridging inferences were routinely drawn during encoding. Singer and Ferreira (1983) added that bridging inferences are necessary for coherence because they often include causal links between text ideas, which were assumed to be implicitly expressed. They suggested that if an inference has been drawn during reading, then the mental representation of the passage would most likely include the inferred concepts that were not directly expressed in the text. The results indicated that bridging inferences were drawn more quickly than elaborative inferences and that passages requiring a bridging inference showed no significant difference in reading time than explicitly stated passages.

Current directions in bridging inference research. Research on bridging inferences in the last fifteen years has focused on (a) the influence of different genres of discourse, such as narrative versus expository text (Wiley & Myers, 2003); and (b) the impact of differing degrees of world knowledge (Saldana & Frith, 2007; Shears & Chiarello, 2004) on bridging inferences. Current research also addresses the effects of brain damage on the ability to draw bridging inferences (Shears & Chiarello, 2004; Tompkins, Fassbinder, Lehman Blake, Baumgaertner, & Jayaram, 2004). Neural measures, such as event-related potentials, have been applied to determine which mechanisms may be at work when bridging inferences are drawn compared to other types of inferences or explicit text (St. George, Mannes, & Hoffman, 1997; Yang, Perfetti, & Schmalhofer, 2007).

Text genres. Causal bridging inferences were found to be generated when subjects read scientific texts, but only if the texts provided readers with all the information they needed in order to create such an inference, in the form of multiple premises (Wiley & Myers, 2003). Another requirement was that there be no significant delay between the presentation of relevant information and the onset of the inference generating task. The authors recounted previous research that had failed to find evidence of bridging inferences in scientific texts and attributed this finding to readers' lack of knowledge in the domain of many of the scientific texts used in previous studies.

Influence of world knowledge. There has been research on two types of world knowledge that influence the generation of causal bridging inferences. Shears and Chiarello (2004) distinguished between the two types and designated them as physical cause and effect knowledge (e.g., *water extinguishes fire*) versus knowledge of planning and goals (e.g., *it is customary to give a gift at a birthday party*). Physical knowledge was assumed to be at the lowest level of a type of knowledge hierarchy and therefore readily available to individuals, whereas goal and planning knowledge was thought to be at a higher level of the hierarchy, such that it was less available to individuals. They found that physical causal inferences were indeed more readily generated than those related to goals and planning. They found higher levels of accuracy and faster response times to target questions following physical inferences as opposed to planning inferences.

Brain damage. Research on brain-damaged individuals (Shears & Chiarello, 2004; Tompkins et al., 2004) showed that individuals with acquired brain injuries were able to make inferences related to physical relations just as well as healthy individuals but had more difficulty with relations of planning and goals compared to healthy individuals.

Both studies used narrative passages which required a bridging inference to connect two sentences (e.g., *Dorothy poured the water on the fire. The fire went out*) compared with control passages that did not require such a bridging inference. They measured individuals' response times and accuracy on subsequent comprehension questions related to the required bridging inferences.

Neural measures. Event-related potentials (ERPs) are brain signals that can be measured by placing electrodes in certain locations on an individual's scalp and measuring the resulting activation. Measuring brain signals is an on-line measure of discourse comprehension and therefore ERPs can been used to clarify when, during reading, inferences are generated. When reading a target word that has been explicitly stated, there is no significant change in the amplitude of certain ERP signals, indicating that the word has successfully been integrated into readers' representation of the text passage. When readers encounter difficulty integrating a target word into their forming memory representation, ERP signals show distinct changes in amplitude (St. George et al., 1997; Yang et al., 2007).

ERPs measures have been used when giving subjects typical narrative passages to read, that either required a bridging inference in order to maintain coherence, had the relation explicitly stated, or a control which consisted of two unrelated sentences (St. George et al., 1997; Yang et al., 2007). These authors found that when bridging inferences were required for comprehension of a text passage, ERP signals were similar to the explicitly stated text conditions. In contrast, control passages that did not explicitly or implicitly (with the use of a bridging inference) state a causal relation between two sentences elicited ERP signals with significant changes in amplitude as compared to both the explicit and inference conditions. These findings provided support for the widely accepted idea that bridging inferences are generated routinely during reading and aid in forming a comprehensive text representation.

Multilevel Text Representation

There is considerable evidence that language comprehension results in the encoding of multiple levels of discourse representation. These levels are the surface form, the propositional textbase, and the situation model, or mental model (van Dijk & Kintsch, 1983). There are additional levels advocated by some theorists, such as the communication level and the text genre level, but the most common distinction is the one described here (Graesser, Millis, & Zwaan, 1997).

Surface form and textbase. The surface form preserves the exact wording and syntax of the message. This is the weakest level of text representation in that the verbatim structure of sentences is quickly forgotten after the reader progresses to the next sentence. However, this weakness can be countered if an individual is instructed to read in order to remember sentence wording (Singer, 1990). This is not to say that surface form generally is completely forgotten. There is some evidence that recognition for identical sentences has an advantage over paraphrases. The advantage has been documented to last up to five days after reading a message (Kintsch & Bates, 1977).

The textbase refers to the idea units or propositions of the message. In the textbase, all the content of the text is present, but it is abstract rather than in a verbatim form and is broken down into component parts called propositions. A proposition consists of a predicate and one or more arguments (Kintsch, 1974). The predicate refers to the part of a sentence or clause which usually contains a verb or noun related to the subject of the sentence or clause. It may also be realized in the surface structure of the sentence as an adjective, adverb or sentence connective. An argument is a noun phrase within a clause related directly to the predicate. The arguments in a proposition may be concepts or else other embedded propositions, and are used to fulfill different semantic functions including agent, object, and goal. (Kintsch & van Dijk, 1978; Oxford Dictionaries Online, n.d.). Content that is explicitly stated in the message, or an inference that fills a gap in a proposition, is found in the textbase. It is suggested that the propositions are ordered in sequence and stored in semantic memory.

The difficulty in comprehension increases with the number of propositions in a text. It takes more time to understand a sentence composed of numerous propositions than one composed of few propositions, even when total paragraph length is kept equal across conditions (Kintsch & Keenan, 1973). Another factor that influences ease of comprehension at the propositional level is the number of word concepts that appear within a text. Texts with fewer word concepts, when concepts were repeated frequently, were read more quickly than texts with several different word concepts, or fewer repeated concepts. This result held even when overall paragraph length was held equal, for both short and long paragraph lengths (Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975). Another finding is that sentences are stored as connected propositions. Probes from within the same proposition as the test word are recalled faster than those that were from a different proposition (Ratcliff & McKoon, 1978).

Situation models. The situation model is the understander’s representation of the situation to which the message refers. This involves the content of the text being integrated with the reader’s pertinent world knowledge about that content. The situation model captures the meaning of the message taking into account additional information not explicitly stated in the passage (Graesser et al., 1997). The situation model is the most enduring level of text representation. The meaning of a message is remembered better and longer than the verbatim message (Kintsch, Welsch, Schmalhofer, & Zimny, 1990).

Situation model representations are different from both surface and textbase representations. A situation model representation incorporates the content of the specific text within a broader context, taking into account factors such as the temporal, spatial, and goal structure aspects as well as pertinent a priori knowledge that the reader already possesses to assist in overall comprehension. Although these levels are sufficiently distinct and able to be isolated, there is sometimes overlap and ambiguity between levels (Graesser et al., 1997).

Additional aspects of situation models. Many aspects of a situation can influence the construction, organization, and maintenance of a text situation model. Research on these factors has focused on spatial relations (Morrow, Greenspan, & Bower, 1987; Radvansky & Copeland, 2000), temporal relations (Rinck, Hahnel, & Becker, 2001) and causal relations including goals (Suh & Trabasso, 1993).

Spatial situation models. Morrow et al. (1987) studied spatial situation models. They reported that readers were able to create situation models based on studying maps of buildings and subsequently reading narratives describing a character’s progress through different rooms of the building. The authors reported that the spatial relation between a character’s location and a probed item had an effect. For example, at certain points throughout the narrative, probe words were presented which were comprised of items found either in the room the character was currently located in, in the room the character most recently had left, in a nearby room that was not recently mentioned in the text, or else in an unmentioned room. The response time to responding to the probe word was the dependent measure. The closer a probed item was in terms of the current room the character occupied, the more easily accessible the probe was to the reader which was reflected in faster response times.

Radvansky and Copeland (2000) found that the functionality of spatial relations in a text influences how well the relation will be integrated into the situation model. They defined functionality as the actual or likely interaction between two or more entities in a situation in which the state of one entity would be affected by a second entity in a meaningful way. The more functional the relation between two entities was, the better integrated it was in the situation model, which was observed by better recall, recognition, and faster reading times of critical sentences. Due to this functionality effect, they concluded that spatial relations are not consistently encoded and stored in the situation model. Spatial relations were better encoded and integrated when they were the sole focus of the text, in which there were no competing aspects, such as temporal and causal relations.

Temporal situation models. Rinck et al. (2001) applied the concept of the iconicity assumption to the study of temporal situation models. From this perspective readers keep track of temporal information while reading a message and have a preference for events described in chronological order. Violations of this assumption led to a decrease in comprehension. Violations included instances of events being described out of chronological order as in flashbacks, or when there were no clear time change indicators between sentences. In their study, they used an inconsistency paradigm in which chronological events were either consistent at test and matched the initial temporal sequence, or inconsistent at test and did not match the initial temporal sequence. They found that the inconsistent condition produced longer reading times for test sentences than in the consistent condition. They attributed this to inconsistency producing a conflict in readers' ability to create an integrated situation model of the temporal events, and therefore required more processing time.

Differences in rate of decay for levels of representation. The surface form level of representation is the fastest to be forgotten followed by the textbase, and the situation model is the most enduring (Kintsch et al., 1990). After people have read discourse material and subsequently received a sentence recognition test, performance is quite high at immediate testing. Participants are readily able to discriminate between sentences that were stated exactly in the text and sentences that were generated through an inference. However, after 40 minutes, the memory for verbatim text reduces by more than half, whereas memory for propositions or the general meaning of the text remains constant. After four days, the memory for verbatim text reduces to near zero, memory for textbase propositional elements reduces to half the original strength, whereas the memory for semantic meaning, or the situation model remains as high as the original strength (Kintsch et al., 1990).

Bridging inferences and levels of representation. Verbatim text is represented at all three levels of representation. Paraphrased sentences are represented in the same way as verbatim text at the textbase and situation model levels, but differ at the surface structure level of representation. Inferences, on the other hand, differ from the verbatim text at both the surface and textbase levels, but would share the same situation model as the verbatim studied text (Kintsch et al., 1990)

Schmalhofer et al. (2002) maintained that the construction of a bridging inference was supported by the activation of any relevant information from the preceding text at the surface and propositional levels. It also required substantial activation from all relevant world knowledge of the reader, at the situation model level. These activations were thought to be high and therefore maintained throughout comprehension, leading to the bridging inference being strongly encoded, with the same strength of encoding as explicitly stated causal relations. The difference in processing between explicitly stated concepts and bridging inferences was attributed to the extent to which different levels of text representation become activated between the two conditions. The production of bridging inferences was influenced more by the encoding of related world knowledge at the situation model level of text representation than by surface or textbase encoding. On the other hand, explicit concepts were supported by a higher level of encoding at the surface and propositional levels of text representation. Although the sources of activation came from different levels of representation, the strength of encoding for both conditions was thought to be fairly equal. This assumption was used to explain why the occurrence of bridging inferences facilitated comprehension to the same extent as explicit texts. The strength of encoding for both bridging and explicit texts was found to be almost equivalent, therefore both facilitated comprehension to the same degree (Schmalhofer et al., 2002).

Distinctiveness

Distinctiveness has traditionally been defined in two different ways. First, items in a given environment can appear to be distinct from their surroundings due to either salience or incongruence of the item compared to its surrounding items. Second, the item can be classified as distinct due to the increasingly poor memory for the surrounding items, due to their increased similarity (Hunt & Lamb, 2001). Research in this area has often used isolation paradigms in which one item in a list is incongruent with all the other items. Results have found that the isolated item is remembered much better than its surrounding non-isolated items. Hunt and Lamb (2001) posited that the mechanism mediating the distinctiveness effect was differential attention to the isolated and surrounding items.

A factor that influences distinctiveness is whether the item is illogical or bizarre. The more illogical or bizarre an item is from the rest of the context, the better it will be remembered. Greater memory for distinctive items compared with memory for the similar items the distinctive items appear with, has been found when bizarreness is manipulated as a within-subjects, mixed variable. This means that only mixed lists of common and bizarre items can allow a given item to appear distinct from its environment (McDaniel, Einstein, DeLoch, May, & Brady, 1995; McDaniel et al., 2005).

McDaniel et al. (1995) proposed that the expectation-violation framework could be used to explain distinctiveness phenomena. The expectation-violation assumption stated that bizarre items violated the readers' expectations. This violation would activate particular encoding processes leading to more elaborate processing and better memory for the bizarre items. However, this account was not supported by their experiments.

Rather, it was suggested that readers spend more time processing or studying bizarre items than common items, leading to better encoding and recall (McDaniel et al., 1995; McDaniel et al., 2005).

Process Dissociation

Recognition can be defined as the perception that something has been previously encountered (Mandler, 1980). Mandler's (1980) work was among the earliest to make the distinction between two separate processes involved in recognition judgments: one based on familiarity and the other involving specific retrieval to identify a target. More recent research has termed the two processes as familiarity and recollection (Brainerd et al., 1999; Brainerd et al., 2001; Chan & McDermott, 2006; Jacoby, 1991; McDermott & Chan, 2006). It is proposed that recollection is a conscious, controlled process, involving the specific retrieval of having encountered a given stimulus. Recollection can be thought of as a search and retrieval process. Familiarity, on the other hand, is thought to be more automatic, utilizing unconscious or implicit memory and is based on the global similarity and relationships among items and concepts.

It has been established that these two processes cannot be mapped directly to specific tasks, because no task can be deemed "process-pure." Labeling a task as process-pure is the assumption that particular processes can be equated with particular tasks and subsequently assuming those tasks provide pure measures of the related processes (Jacoby, 1991). Recollection and familiarity work together and are involved to different extents, providing differential contributions, within a given task. In order to potentially assess the separate contributions of familiarity and recollection, Jacoby implemented a process-dissociation procedure. After reading two distinct lists of words, participants were instructed to make either inclusion judgments, in which targets from both lists were to be accepted during a recognition test; or exclusion judgments, in which only targets from one list were to be accepted. This paradigm can be used to determine whether conscious or unconscious mechanisms may be at play during recognition. The ability to discriminate in which list a given target word appeared under exclusion instructions is an example of recollection being used because the reader is able to consciously retrieve from memory which items appeared in which list. The inability to make this discrimination is an example of familiarity since the reader is unable to distinguish which list an item was encountered in and uses a feeling of familiarity as the basis for responding.

Conjoint recognition. Due to some limiting properties of Jacoby's original process dissociation paradigm, Brainerd et al. (1999) developed a new model called the conjoint recognition model. This model was based on an identity-similarity distinction which assumed that readers access separate representations of both the target's (a) surface forms and verbatim traces and (b) meanings and relational information. When both of these types of representations are available, recognition depends on retrieval cues provided by recognition probes. The verbatim, surface forms favour recollection; because when verbatim traces are activated, they are compared to the presented probe. If the verbatim trace matches the probe, it will provoke correct acceptance of the target; if it does not match, correct rejection of a distractor will occur. On the other hand, the semantic meaning, or gist retrieval, favour feelings of familiarity. If a probe word activates the semantic representation of a word that was studied during the previous experimental phase, the feeling of familiarity provided by the activated meaning will promote recognition of the probe word, even if it is only a semantically-related distractor.

The conjoint recognition (CR) model uses the same logic as Jacoby's (1991) process-dissociation analysis, but expands it to include three separate instructional conditions and uses two distinct classes of distractors. The two types of distractors used are semantically-related distractors and unrelated distractors. If category words were being studied for example, a given list could include several bird names, such as "robin", "sparrow", and "eagle". A semantically-related distractor would be a bird name that had not been presented during study, such as "hummingbird", whereas an unrelated distractor would be a word from a category different than bird names, such as the word "oxygen".

The CR model uses a single study list instead of two, in order to avoid the source-memory problem of participants recollecting a study word but being unable to remember which list it came from. Three conditions of instruction type are given: (a) accept targets and reject both types of distractors ("recognize" instructions); (b) accept related distractors and reject targets and unrelated distractors ("imply" instructions); and (c) accept both targets and related distractors, and reject unrelated distractors ("verify" instructions). After study, three types of items are presented at test: targets, semantically-related distractors, and unrelated distractors. The semantically related distractors would include synonyms, antonyms, and category names of the targets (Brainerd et al., 1999, 2001; Singer & Remillard, 2008).

Being able to recollect a word as having been presented during study will cause readers to accept the word in the recognize and verify instruction conditions only. Being able to identify a semantically-related distractor will cause acceptance in the imply and verify instruction conditions only. If a test word provokes only strong feelings of familiarity however, the reader is unable to recollect whether the word was actually presented or not. The related distractor may be treated as a target and accepted in the recognize and verify instruction conditions, causing false alarms. Alternatively, it could be treated as a distractor and accepted in the imply and verify instruction conditions. In the case of semantically-related distractors, recollection and familiarity have opposite influences on recognition judgments. If a semantically-related distractor is accompanied by strong feelings of familiarity, it will result in acceptance of the test item, or a false alarm. However, if there is a strong feeling of contrast between the distractor and the word from study, correct rejection will result due to recollection (Brainerd et al., 2001).

It is suggested that inferences cause a dual representation in memory – the text that was stated directly, plus the new information generated due to the inference (Fincher-Kiefer, 1995). This dual representation can cause uncertainty about the source of memory for the target. This is known as the source-monitoring hypothesis. When one reads a text that involves an inference, both the stated text and the inferred information become stored in the person's memory. The theoretical analyses like that of Schmalholfer et al. (2002) suggest that stated information is stored in all three levels of text representation – the surface form, the textbase, and the situation model. On the other hand, inferred information is stored only at the situational level. At time of recall, both types of information become activated and the reader must make an effort to determine which elements of the memory representation were explicitly stated and which were only implied (Fincher-Kiefer, 1995).

A limitation of the CR model is that it is unable to determine how confident readers are of their recognition judgments, and how to distinguish between false alarms and correct rejections involving semantically-related distractors. Brainerd et al. (2001)

reported certain recognition paradigms in which false alarms were followed by retrospective reporting of actually recollecting having studied the semantically-related distractors. This suggested that in some cases the distractors were being experienced as being just as real as the actual studied items. In particular, this result was found in paradigms in which several of the studied targets belonged to a specific category and shared meaning. In this case, semantically-related distractors were strong retrieval cues for the meaning of the studied items. This type of paradigm is known as the Deese-Roediger-McDermott (DRM) procedure (Brainerd et al., 2001).

Phantom recollection. To address the phenomenon of high levels of retrospective illusory recollection, Brainerd et al. (2001) incorporated an additional parameter into their conjoint recognition model, namely, phantom recollection. Incorporating this additional mechanism allowed the assessment of the contribution of three distinct processes to false alarm rates involving semantically-related distractors: familiarity, target recollection, and phantom recollection. The addition of phantom recollection to the model allows quantification of its contribution to false alarm rates. Phantom recollection produces a pattern of responding to recognition judgments that is different than the responding induced by feelings of familiarity. A probe word that induced phantom recollection would be accepted only in the recognize and verify conditions but not in the imply condition. Phantom recollection of a semantically-related distractor causes it to be treated as a target is normally treated. The separate contributions of familiarity, target recollection, and phantom recollection can be measured using the proportion of acceptances for targets, related distractors, and unrelated distractors in each of the different instructional conditions, and mapping these values to a set of formulas indicating the relative contributions of each of target recollection, familiarity, and phantom recollection. The modeling to be used in this proposal will be examined further in the experimental section below.

Overview of Thesis Project

The purpose of the subsequent experiments was to evaluate the effect on readers' memory representations of implying a discourse action versus stating it explicitly. It was evaluated whether there was any effect of not-implying, or contradicting an action when it is subsequently stated, and whether creating a contradictory situation would cause distinctiveness effects on memory. The phantom recollection procedure outlined above was implemented. Traditionally, multinomial models have been used for memory of lists of words, but have recently also been applied to discourse comprehension (Singer & Remillard, 2004, 2008). Furthermore, the conjoint recognition and phantom recollection multinomial models have very seldom been applied in past research due to the complex design of the required experimental paradigm (Erdfelder, et al., 2009; Stahl & Klauer, 2008). My experiments aimed to further the research on whether the phantom recollection paradigm and model can be adequately applied to the study of discourse memory and comprehension as well as inference generation. A pilot experiment and norming experiment were conducted in order to evaluate whether the materials created for the final experiment would generate the expected pattern of results. Further details about the materials used and the subsequent hypotheses of results are included in the following sections.

Experiment 1: Pilot Experiment

Experiment 1 was a pilot experiment conducted to evaluate whether the materials created would yield the expected pattern of results. The passages created for this experiment were based on materials from Potts et al. (1988). Two-sentence passages were created in which the second sentence either stated an action ("stated"), or did not ("not-stated"). The first sentence either supported the implication of the action of the second sentence ("implied"), or did not support, or contradicted, the implication ("not-implied").

An example of one of the passages is:

(1) a. Sharon cautiously pulled on her (beach clogs/bathing cap).

b. The bather overlooked the piece of glass in the shallow water and (cut her foot/stepped) on it.

The parentheses in sentence (1a) indicate the not-implied/implied manipulation, and the parentheses in sentence (1b) indicate the stated/not-stated manipulation. The stated or to-be-inferred action in this passage is that Sharon cut her foot on the glass.

In the directly stated condition, the second sentence would read *The bather overlooked the piece of glass in the shallow water and cut her foot on it*, while in the not-stated condition it would read *The bather overlooked the piece of glass in the shallow water and stepped on it*. Although the word "cut" was not used in the not-stated condition, the inference of the bather cutting her foot was still possible when reading that she stepped on the glass. Sentence 1 was designed to either support or contradict the implication of Sharon cutting her foot. The implied condition would read as *Sharon cautiously pulled on her bathing cap*. This is classified as implied because putting on a bathing cap will not prevent Sharon from subsequently cutting her foot. On the other hand, the not-implied condition would read as *Sharon cautiously pulled on her beach clogs*. This sentence is classified as not-implied since beach clogs would normally prevent the bather from cutting her foot. This sentence should prevent the reader from making the given inference of Sharon cutting her foot due to stepping on a piece of glass.

The original materials from Potts et al. (1988) included 40 passages which I modified to fit into the two-sentence, not-implied/implied, stated/not-stated format. This implied by stated manipulation yielded four versions of each passage: implied/stated (explicit), implied/not-stated (inference), not-implied/stated (contradictory), and not-implied/not-stated (control). The full set of 40 passages is included in Appendix A. Table 1 outlines the four possible conditions.

Table 1. *Four text conditions obtained by crossing Statement X Implication.*

	Implication	
Statement	Implied	Not-Implied
Stated	Sharon cautiously pulled on her bathing cap. The bather overlooked the piece of glass in the shallow water and cut her foot on it. (explicit)	Sharon cautiously pulled on her beach clogs. The bather overlooked the piece of glass in the shallow water and cut her foot on it. (contradictory)
Not-Stated	Sharon cautiously pulled on her bathing cap. The bather overlooked the piece of glass in the shallow water and stepped on it. (inference)	Sharon cautiously pulled on her beach clogs. The bather overlooked the piece of glass in the shallow water and stepped on it. (control)

The critical condition among these four would be the not-implied/stated, or contradictory, manipulation. This condition of the above example would read as *Sharon cautiously pulled on her beach clogs. The bather overlooked the piece of glass in the shallow water and cut her foot on it.* The reasoning is that readers should find this combination of implication and statement to be quite unlikely. Using the given example, if the bather had put on beach clogs, her feet should have been protected and not subsequently cut by the glass in the water. The 40 new passages were used as the basis of the experiments reported herein.

The premise for the pilot experiment was that if participants encountered a passage in which an action was implied, but not stated, they would likely make the inference of the action occurring. Using the same example as above, after reading the inference passage *Sharon cautiously pulled on her bathing cap. The bather overlooked the piece of glass in the shallow water and stepped on it,* readers should readily draw the inference that the bather cut her foot on the glass. In a subsequent recognition test, as has been found in previous research, readers might believe they had read the stated version of the target sentence *The bather overlooked the piece of glass and cut her foot on it* as opposed to the not-stated version which they had actually read.

However, if readers had read a passage that did not imply the action, there should be no evidence of an inference being made if the not-stated version of the target sentence was presented. Yet, if the action was subsequently stated, as in the critical contradictory condition, then this should surprise readers, or be distinctive, and thus the fact that the second sentence stated an event which would not seem likely according to the first sentence will cause the second sentence to be better stored in memory, as has been found in distinctiveness research.

Using this reasoning, it was hypothesized that when presented with a target sentence during test which stated the action from a given passage, readers would judge that they had encountered the sentence exactly as shown in both the explicit and inference conditions. The explicit condition would lead to a recognition judgment of “yes” due to the fact that the action was indeed stated during study. The inference condition would lead to a recognition judgment of “yes” as well because of the inference generated at time of study that the action had occurred.

In terms of both of the not-implied conditions, there were two different expected results. In the control condition, the subsequent target sentence was expected to be judged as not having occurred during study due to the fact that the studied passage did not imply or state the given action. In the contradictory condition, however, the target sentence would possibly be recollected as having been presented in study due to the fact that the study passage was distinctive in that the readers would have been surprised to read that a stated action occurred when the preceding sentence did not imply it.

Of the three conjoint recognition instructions, only the recognize instructions were used for the pilot experiment. Participants were instructed to only accept target sentences at test if they had appeared word-for-word during the study phase. The objective was to see whether subjects were able to discriminate between the stated and implied versions of the test sentences, so using the simple recognition instructions was sufficient for this pilot experiment.

I expected that the results would show an effect of statement, such that test sentences from both of the stated conditions would be accepted significantly more often than sentences from either of the not-stated conditions. Within the not-stated conditions, I expected to find an effect of implication, such that inference items (implied, not-stated) would be accepted significantly more often than control condition sentences (not-implied, not-stated). This hypothesis was based on the findings that participants often accept inference items as having been studied in an explicit form, when in actuality these items were only implied. Additionally, I expected there to be a Statement x Implication interaction, such that the effect of statement would be greater in the not-implied conditions than in the implied conditions. Participants should have accepted target sentences from the inference condition significantly more often than in the control condition.

Method

Participants. Sixty-four male and female Introductory Psychology students at the University of Manitoba were recruited as participants from the Introductory Psychology subject pool. All participants were native English speakers. They received partial course credit for their participation.

Materials. Sixteen of the 40 passages I created were chosen at random. These passages were then divided into four groups of four passages each. Each of these groups was assigned to a distinct "verbal set." Using these verbal sets, list 1 was created, with each verbal set randomly assigned to one of the four conditions outlined above (i.e. explicit, inference, contradictory, control). The passages were then randomly assigned to a list position, with the restriction that two passages from each condition appear in each half of the list. The three other study lists were created by cycling the verbal sets across conditions using a Latin-square pattern. This resulted in every passage appearing in each of the four conditions across the four lists. List order remained constant and the condition each passage appeared in was varied across each list. Three filler passages preceded the 16 experimental passages, and three filler passages followed the experimental passages. The filler passages were unchanged across the four lists and served as buffers to avoid primacy and recency effects. Participants encountered 22 passages in total.

The test list was generated from the stated versions of each of the second sentences from the 16 experimental passages used. These test probes appeared in a fixed random order and were preceded by six filler probes which corresponded to the filler passages from study. Four of the filler probes mimicked each of the four experimental conditions, while the other two filler probes were unrelated to their corresponding passage.

Procedure. Participants were tested in groups of two to four. Each participant was seated at a station with a computer screen and keyboard on a desk in front of them. The computer screen was positioned 22 cm away from the near edge of the table. Before beginning the experiment, participants read and signed an informed consent form, and also read a one-sheet, double-sided page of instructions. The instructions outlined the procedure of the study phase and the following recognition test. There was an example passage with four sample test sentences provided. The correct answers and rationale were provided for each of the sample test sentences. The experimenter asked all participants if they understood the instructions and answered any questions if necessary before starting the experimental program on the computer.

In the study phase, before each passage was presented, the word “READY” was displayed in the centre of the screen at which time participants were required to press the space bar to display the passage. Participants were first presented with a fixation “X” for 500-ms, left-justified halfway down the computer screen to orient their gaze to the correct location on the screen, after which the first sentence of the two-sentence passage appeared at the same location. They read at their own pace and pressed the space bar to proceed to the next sentence. Upon pressing the key, the first sentence was removed from the screen and the second sentence immediately appeared. Once both sentences had been read and understood, participants were required to make a 4-point rating for each passage judging the degree of activity conveyed by the whole passage. They were reminded of this task by having the phrase "Reply 1, 2, 3, 4" appear at the top of the screen when the second sentence of each passage was presented. A rating of "1" indicated "very passive" while a rating of "4" indicated "very active." Participants responded with the 1 to 4 keys indicating their rating of degree of activity. After the response was registered, the "READY" screen was displayed again after a brief interval of 3000-ms to prompt the start of the next trial. Participants proceeded through all trials at their own pace.

After the study phase was completed, a message appeared in the centre of the screen instructing participants to switch their response keys to the marked NO and YES keys on the keyboard. The "x" and "." keys were labeled "NO" and "YES", respectively. The message also reminded participants to respond YES only to sentences that had appeared word-for-word in the study phase, and to respond NO to all other sentences. Participants pressed the READY key to start the recognition test. A fixation X appeared in the same location as in the study phase for 500-ms before each test sentence was displayed. Once the test sentence appeared, participants responded "yes" or "no". After the response was registered, the sentence was removed from the screen. Following a 1000-ms interval, the next trial automatically began, again with a fixation X preceding the test sentence. There was no time limit on responding "yes" or "no" to each test probe, and response times were automatically recorded by the experimental software.

In summary, single sentences were presented one at a time, and participants were instructed to respond "yes" if the test sentence had appeared word-for-word in any of the previous passages and “no” if it had not. It was emphasized in the instructions that even if the sentence had the same meaning as one from the first phase, but had different wording, a “no” response should be given. After each response, the next sentence was presented. Again, participants were not timed and proceeded at their own pace.

Results

The results for this experiment were analyzed by alternately treating subjects (F1) and items as a random effect (F2; Clark, 1973). Mean values for proportion yes values are shown in Table 2.

Table 2. *Mean values for proportion “yes” responding in Experiment 1.*

Condition	Proportion “yes”
Explicit (Implied/Stated)	0.74 (0.24)
Inference (Implied/Not-Stated)	0.35 (0.26)
Contradictory (Not-Implied/Stated)	0.80 (0.21)
Control (Not-Implied/Not-Stated)	0.29 (0.26)

Note. Standard deviation shown in parentheses.

In the subjects-random analysis a 2 x 2 x 4 mixed ANOVA was conducted. Implication (implied/not-implied) and statement (not-stated/stated) were the within-subjects variables, and the four counterbalanced lists constituted a between-subjects variable. The items-random analysis used a mixed 2 x 2 x 4 design as well. Statement and implication were the two within-items variables, while verbal set was the between-items variable. All results were analyzed using $\alpha = 0.05$.

There was no significant effect found for list or verbal set, all $Fs < 1$. Therefore the results were collapsed across lists and verbal sets and analyzed together. The effect of statement was significant, $F1(1, 58) = 179.01$, $MS_e = 1.11$, $F2(1, 12) = 49.84$, $MS_e = 0.07$, such that stated items were accepted more often than non-stated ones. The implication main effect was not significant, Fs < 1. The Statement x Implication interaction was significant, $F1(1, 58) = 4.92$, $MS_e = 0.58$, $F2(1, 12) = 5.26$, $MS_e = 0.01$, reflecting that there was a greater effect of statement in the not-implied conditions than in the implied conditions. Tests of simple effects were also conducted in the subjects-random analysis. There was no significant effect of statement in the "implied" conditions (i.e. "implied/stated" and "implied/not-stated"), $F < 2$. There was also no significant effect of statement in the "not-implied" conditions, $F < 2$.

Discussion

These results supported the initial predictions. There was an effect found for statement, as well as a significant interaction between statement and implication. Test items from the contradictory condition were accepted slightly more often than test items from the explicit condition, suggesting that these contradictory passages had the benefit of better encoding due to being distinctive. The Statement x Implication interaction lent support to this hypothesis. Inference condition items however, were not accepted at as high a rate as was predicted. This could have been due to the short delay between study and test, causing participants' verbatim text representations to still be strong. In other words, the verbatim text representations were still accessible to participants at test, so they were able to accurately remember having read the implied passage rather than a stated version.

Experiment 2: Norming Experiment

Experiment 2 was conducted as a norming experiment to determine whether the materials created for this proposal were indeed effective in eliciting the desired responses from participants. Although the materials used were the same as in the pilot experiment, the logic for the norming experiment was completely different than for the pilot. In the norming experiment, participants were asked to rate each of the two-sentence passages on how possible they felt the events conveyed in the passage were. There was no test phase in which participants would make recognitions judgments.

In both the explicit and inference conditions, participants were expected to rate the passages as being fairly to highly possible. Table 1, presented earlier, shows the setup of the passages in each condition. Using the example from Table 1, the explicit and inference conditions should generate fairly high ratings of possibility by readers when reading that the bather overlooked a piece of glass and either cut her foot, or stepped on it. The contradictory condition however was expected to generate ratings of fairly to highly impossible, due to the contradictory nature of the non-implication preceding the stated sentence. For example, readers would find it impossible for the bather to cut her foot on the glass once she had put on her beach clogs. The not-implied/not-stated version served as a control, where the rating of possibility was expected to be fairly neutral. Using the same example, readers would have a neutral rating of possibility if the bather had put on beach clogs and stepped on the glass, without mention of her foot being cut.

It was expected that there would be a main effect for both implication and statement. Also, an interaction effect was also expected, such that there would be a larger difference between both stated conditions than between both not-stated conditions. This would show that the explicit passages were rated as more possible than the contradictory passages, which should be rated as being impossible. The difference between these two conditions would be greater than the difference between the inference and control passages.

Method

Participants. Seventy-six male and female participants were recruited from the Introductory Psychology subject pool at the University of Manitoba. They received partial course credit for their participation. Again, participants were required to have English as their first language. No participant that had been a part of the pilot experiment was allowed to participate in the norming experiment.

Materials. The 40 passages derived from the Potts et al. (1988) materials were used. The full set of materials is provided in Appendix A. Four counterbalanced lists were created in which 10 of the 40 passages appeared in each of the four conditions: explicit, inference, contradictory, and control. Four practice passages preceded the experimental passages. Also, 16 filler passages were interspersed within the 40 experimental passages. The filler items were used to generate either relatively low or high ratings so that participants would use the full range of responses available. The order of the experimental and filler passages was random, with the constraint that no three consecutive items could be from the same condition or fillers. In total each list contained 60 passages.

Procedure. Participants were tested in groups of two to four. Each participant was seated at a station with a computer screen and keyboard on a desk in front of them. The stations were set up in the same way as in Experiment 1. Participants were told that they would be reading two-sentence passages and then making a rating based on how possible they felt the event from sentence two would be given that the event in sentence one had occurred. The ratings were on a 7-point Likert scale, with “1” referring to "extremely impossible", "4" referring to "50-50", and "7" referring to "extremely possible." The passages were presented on the computer screen one at a time. The participants read at their own pace and were not timed.

Before each passage was presented, "READY" was displayed on the screen. The participants then pressed the space bar on the keyboard to display the passage. A fixation "X" appeared on the left side of the screen for a brief moment before the passage appeared. The two sentences of the passage were displayed left-justified, halfway down the screen, with a single blank line separating the first sentence from the second sentence. Participants were instructed to read and understand the passage, and then to use the 1 to 7 keys on the keyboard to indicate their rating of how possible they believed the events in the passage were. After the participants had made a response, the "READY" screen would be presented again before the next passage. Halfway through the list there was a 30-second rest period. The session took approximately 20 minutes to complete.

Results

Table 3. *Mean ratings of degree of possibility for items in Experiment 2.*

Condition	Mean rating
Explicit (Implied/Stated)	4.99 (0.67)
Inference (Implied/Not-Stated)	4.76 (0.75)
Contradictory (Not-Implied/Stated)	3.87 (0.96)
Control (Not-Implied/Not-Stated)	4.30 (0.89)

Note. Standard deviation shown in parentheses.

Mean ratings for each of the four experimental conditions are shown in Table 3. A 2 x 2 x 4 mixed-design ANOVA was applied to these scores. Statement and implication were the within-items variables, while list and verbal set were the between-items variable. Again, both subjects (*F1*) and items (*F2*) were analyzed separately as random effects. There was no significant effect found for list or verbal set, *F*s < 2. There was an effect of implication, $F1(1, 60) = 100.54$, $MS_e = 4{,}041.54$, $F2(1, 36) = 53.38$, $MS_e = 4{,}782.31$, such that items in both implied conditions received significantly higher ratings than items in either of the not-implied conditions. There was no significant effect found for statement, $F1(1, 60) = 2.94$, $p = 0.09$, $F2(1, 36) = 1.02$, $p = 0.32$. The Statement x Implication interaction was found to be significant, $F1(1, 60) = 22.72$, $MS_e = 3{,}318.40$, $F2(1. 36) = 21.57$, $MS_e = 2{,}190.04$.

For the items-random analysis, a contrast was conducted. The effect of statement was not significant in the “implied” conditions (i.e. explicit and inference conditions), $F2(1, 36) = 3.45$, $p = 0.07$. However, the effect of statement was significant in the “not-implied” conditions, $F2(1, 36) = 10.08$. Additionally, the subjects-random analysis showed a significant effect for the Statement x List interaction, $F1(3, 60) = 8.26$, $MS_e = 2{,}831.31$, as well as for the Statement x Implication x List three-way interaction, $F1(3, 60) = 12.25$, $MS_e = 3{,}318.40$.

Discussion

The ANOVA results generally supported the initial hypotheses. Surprisingly, there was no significant effect for statement, but there was an effect for implication. This was opposite to the results of the pilot experiment. Because the purpose and task of the norming experiment were different than those for the pilot experiment, this is not worrisome. More importantly, there was a significant interaction effect found, which showed that there was a significantly larger difference between both stated conditions, than both not-stated conditions. This supports the prediction that reader’s should find the contradictory (Not-implied/stated) condition to be highly unlikely, whereas both not-stated conditions generated more neutral ratings.

Evaluating the mean item ratings helped determine that our materials were indeed generating the appropriate reactions from participants and helped identify those items which did not function as desired. Passages that showed either higher ratings in the contradictory condition than the explicit condition, or showed minimal rating differences across the four conditions were excluded from the subsequent experiment. Items 12, 15, 23, and 35 were excluded.

Experiment 3: Phantom Recollection Paradigm

The main experiment I conducted for my M.A. thesis utilized the above-described materials that were created for this set of experiments. I applied the full phantom recollection paradigm (Brainerd et al., 2001) to discourse passages. The result of the experiment was predicted to lend support to Singer and Remillard's (2008) finding that the phantom recollection paradigm can be applied more generally to a wider variety of stimuli which better represent the types of discourse people encounter in their everyday lives, as opposed to lists of words.

Like in the pilot experiment, it was hypothesized that during the recognition test, a target sentence which stated the action from a given passage would lead readers to judge that they had encountered the sentence exactly as shown in both the explicit and inference conditions. The explicit condition would lead to a recognition judgment of "yes" due to the retrieval of memory of the action being explicitly stated during study. The inference condition could also lead to a recognition judgment of "yes" because of the inference generated at time of study that the action had occurred. The generation of an inference could cause phantom recollection, where participants would possibly have illusory recollection of reading an explicitly stated sentence during test when they actually had not. If however, the participant was able to recollect that the studied passage had not explicitly stated the test sentence, a “no” response would be expected.

Traditionally response times (RTs) have been analyzed in order to extract patterns of responding under certain experimental conditions. It was thought to possibly be insightful to examine the RT data from this experiment as well. Therefore, I will also examine the resulting RTs and conducting statistical analysis on those data as well.

For both not-implied conditions, there were two distinct results expected. In the control condition, the subsequent target sentence should have been judged as not having occurred during study because the studied passage did not imply or state the given action. However, there is a chance that control passages would be judged as having been encountered due to participants remembering some of the content words, and subsequently basing their judgment purely on a sense of familiarity. In the critical contradictory condition, however, the target sentence should be judged as having been presented in study. During study, readers would likely have been surprised to read that a stated action occurred when the preceding sentence did not imply it. The surprise, or distinctiveness, would lead to better encoding and memory of the studied material.

I used five conditions within the materials: the four outlined above (explicit, inference, contradictory, control) and an additional “absent” condition. The absent condition provided test sentences that had not been presented during the study phase. More specifically, a subset of passages from the original set of materials was assigned to the absent condition. The passages were not presented during study. However, the test sentences corresponding to the absent passages were presented during the memory test. These sentences served as unrelated distractors. Any acceptances of the absent sentences were indicative of guessing.

Participants were randomly assigned to one of three conjoint recognition instruction conditions: recognize (only accept target sentences if they appeared word-for-word during study); verify (accept targets if they either appeared word-for-word or were implied by the passage during study); and imply (only accept targets if they were implied and not stated during study). Thus, in total there were 15 conditions for responding at test (5 text conditions x 3 instructional conditions). The proportion of “yes” and “no” responses in each of the 15 conditions was used to evaluate the phantom recollection model that was implemented for this experiment.

Multinomial Modelling

The phantom recollection model was adapted from Singer and Remillard’s (2008) model of discourse retrieval. The full set of fifteen equations is provided in Appendix B. Six of the more relevant equations are outlined below. The equations show the probability, p_{xy}, of responding “yes” to the corresponding stated test sentence for a studied passage. The subscript *x* corresponds to the text condition: explicit (*e*), inference (*i*), contradictory (*u*), control (*c*), absent (*g*); the subscript *y* corresponds to the instructional condition: recognize (*r*), verify (*v*), imply (*i*). The parameters *R* (recollection), *F* (familiarity), *P* (phantom recollection), *E* (erroneous recollection of related targets), and *G* (guessing) indicate the contribution of each of the respective parameters to the total probability of responding “yes” in each condition. Singer and Remillard (2008) defined the “erroneous recollection” parameter as the erroneous recollection of other targets due to similarity to the probe. Because the text passages used for this experiment were unrelated to each other, erroneous recollection was given a constant value of 0.05 throughout the modelling. Also, each parameter estimate, with the exception of *G*, has a single letter subscript denoting which text condition is represented by the given parameter (i.e., text conditions *e*, *i*, *u*, and *c*). The parameter *G* is the probability of accepting a target sentence given that it fails to elicit any type of recollection or sufficient familiarity. This was directly measured as the proportion of responding “yes” to test sentences from the absent condition (distractors). Since there was no content overlap between the absent test sentences and the passages studied in the first phase, any acceptances were considered as guesses.

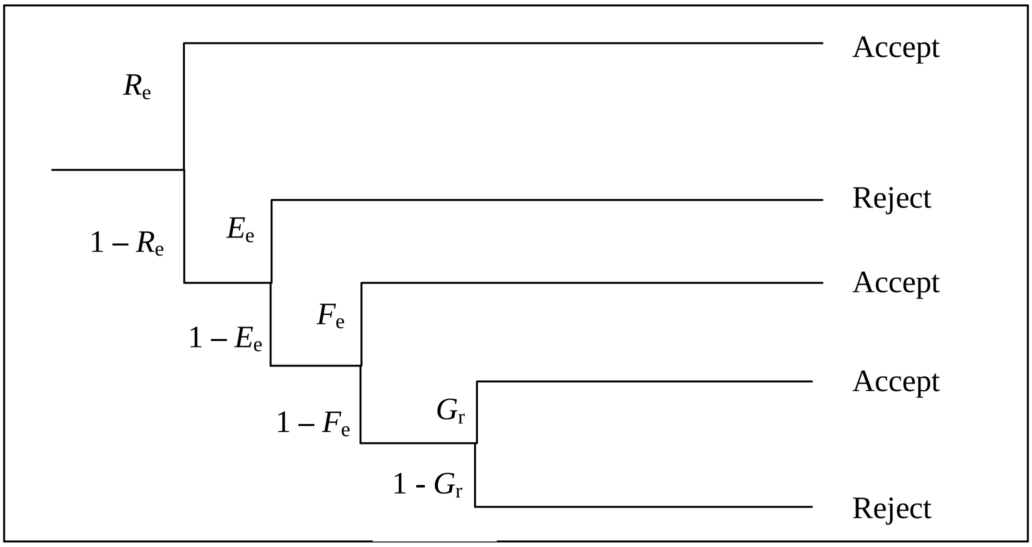
The equations for the explicit-recognize (*er*) and explicit-imply (*ei*) instructional conditions are:

$$p_{\mathrm{er}} = (R_{\mathrm{e}}) + (1 - R_{\mathrm{e}})(1 - E_{\mathrm{e}})(F_{\mathrm{e}}) + (1 - R_{\mathrm{e}})(1 - E_{\mathrm{e}})(1 - F_{\mathrm{e}})(G_{\mathrm{r}})$$

$$p_{\mathrm{ei}} = (1 - R_{\mathrm{e}})(E_{\mathrm{e}}) + (1 - R_{\mathrm{e}})(1 - E_{\mathrm{e}})(F_{\mathrm{e}}) + (1 - R_{\mathrm{e}})(1 - E_{\mathrm{e}})(1 - F_{\mathrm{e}})(G_{\mathrm{i}})$$

Each element of each equation indicates the possible pathway that will result in acceptance of a target sentence at test in each respective condition. Another way to conceptualize each equation is by using a multinomial processing tree diagram. Figure 1 outlines the tree diagram for equation B1.

Figure 1. *Multinomial processing tree diagram for the explicit condition, under the recognize instructions*.



In equation B1 (explicit-recognize condition), (R_e) indicates that the recollection of the studied passage will result in target acceptance because the participant remembers that they had indeed studied the sentence word-for-word in the first phase of the experiment. The expression $(1 - R_e)(1 - E_e)(F_e)$ indicates that in the absence of recollection and erroneous recollection, familiarity of the item will result in target acceptance. Finally, $(1 - R_e)(1 - E_e)(1 - F_e)(G_r)$ indicates that when recollection, erroneous recollection, and familiarity all fail, then target acceptance is still possible if the participant simply guesses “yes.”

For the explicit-imply condition (equation B3) there will be no target acceptance if the participant is able to recollect the studied passage. Recollection would result in remembering that the sentence was stated word-for-word, and not implied, and as per the imply instructions the participant would respond “no”. The only difference between the explicit-imply condition and the explicit-recognize condition is in the first element, $(1 - R_e)(E_e)$, which indicates that if the item elicits erroneous recollection of related studied material, the participant may respond “yes”.

The equations for the inference-recognize (*ir*) and inference-imply (*ii*) conditions are:

$$p_{ir} = (1 - R_i)(P_i) + (1 - R_i)(1 - P_i)(F_i) + (1 - R_i)(1 - P_i)(1 - F_i)(G_r)$$

$$p_{ii} = (R_i) + (1 - R_i)(1 - P_i)(F_i) + (1 - R_i)(1 - P_i)(1 - F_i)(G_i)$$

In the inference condition, the test sentence was not-stated during study. Therefore in the recognize instructional condition, participants would only accept the test sentence if there was evidence of phantom recollection at play, $(1 - R_i)(P_i)$; if the item failed to induce recollection and phantom recollection, but exceeded the acceptance threshold for a sense of familiarity, $(1 - R_i)(1 - P_i)(F_i)$; or if the participant simply guessed, $(1 - R_i)(1 - P_i)(1 - F_i)(G_r)$. In contrast, in the imply instructional condition, participants would accept the test sentence if they explicitly recollected reading the study passage, (R_i). They would remember that the passage had not directly stated the test sentence, but had rather implied it, and respond “yes” as per the imply instructions. A sense of familiarity that exceeded the acceptance threshold would also lead to target acceptance, $(1 - R_i)(1 - P_i)(F_i)$, as well as guessing, $(1 - R_i)(1 - P_i)(1 - F_i)(G_i)$.

In the contradictory condition, the test sentence was stated during study as well as at test. The critical factor however is that the first sentence of the study passage had contradicted the stated target sentence. In this situation there would be no use of a phantom recollection (P) parameter, since the target sentence was actually presented the same way at study as at test. It is predicted that the contradictory nature of the passage would cause the passage to appear distinct within the study phase, and lead to better memory for the verbatim sentences. However, there is the possibility that the contradictory nature of the study passage will influence erroneous encoding of the sentences, such as believing the target sentence was *not* presented in the "stated" form during study due to the opposing nature of the first sentence. This type of erroneous recollection is not the same as phantom recollection, therefore the same parameter *E* will be used, as was used for the explicit condition equations.

The equations for the contradictory-recognize (*ur*) and contradictory-imply (*ui*) conditions are:

$$p_{ur} = (R_u) + (1 - R_u)(1 - E_u)(F_u) + (1 - R_u)(1 - E_u)(1 - F_u)(G_r)$$

$$p_{ui} = (1 - R_u)(E_u) + (1 - R_u)(1 - E_u)(F_u) + (1 - R_u)(1 - E_u)(1 - F_u)(G_i)$$

Using the above described reasoning and applying the parameters *R* (recollection), *F* (familiarity), *P* (phantom recollection), *E* (erroneous recollection of related targets), and *G* (guessing), I created the full model that will be used to estimate the individual contributions of recollection, familiarity, and phantom recollection to discourse comprehension. The full model is included in Appendix B. The model was evaluated using the multinomial processing tree analysis, GPT and multiTree software (MPT; Hu & Phillips, 1999; Moshagen, 2010).

Method

Participants. One hundred and ninety two participants were recruited from the Introductory Psychology subject pool at the University of Manitoba. Participants were required to have English as their first language. Participants must not have taken part in either of the pilot or norming experiments. Sixty-four participants were randomly assigned to each of the three instructional conditions: recognize, verify, and imply. Within each instructional condition, participants were randomly assigned to one of the five counterbalanced lists. This resulted in either 12 or 13 participants being assigned to each list.

Materials. In Experiment 2, mean item ratings were evaluated. An item which showed higher ratings in the not-implied/stated condition than in the implied/stated condition was judged to be a bad item since it elicited the opposite response than expected. Also, if an item showed minimal difference in ratings across the four conditions it was judged to be a bad item since the Statement x Implication manipulation did not alter readers' representations of the passages. These criteria resulted in four of the 40 items being considered bad items. The four bad passages were excluded from Experiment 3.

The materials for this experiment were created from 20 randomly selected passages from the 36 passages evaluated in the norming experiment. Lists were created in much the same way as in the pilot experiment except that five counter-balanced lists were created in which four experimental passages appeared in each of the five conditions: explicit, inference, contradictory, control and absent. Each of these groups was assigned a verbal set from one to five. For list 1, each verbal set was randomly assigned to one of the five conditions. The passages were then randomly assigned to a list position, again with the restriction that two passages from each condition appear in each half of the list. The remaining four lists were created by cycling the verbal sets across conditions using a Latin-square pattern. This resulted in every passage appearing in each of the five conditions across the four lists. List order remained constant across lists and the condition of each passage was varied across each list. The absent passages were assigned a list position in their respective lists, but were not presented during study. However, the corresponding test sentences from each of the absent passages were presented during the test phase. Thus, 20 experimental passages were assigned to each list but only 16 of those 20 passages were actually presented during study. Three filler passages preceded the 16 presented experimental passages, and three filler passages followed the experimental passages, which served as primacy and recency buffers, respectively. The filler passages were unchanged across the five lists. Participants encountered 22 passages in total in the study phase of the experiment.

The test list consisted of the stated versions of each of the second sentences from all 20 experimental passages used. These test probes appeared in a fixed random order and were preceded by six filler probes which corresponded to the filler passages from study. Four of the filler probes mimicked each of the initial four experimental conditions. The other two filler probes were unrelated to their corresponding passage, so they mimicked the absent condition.

Procedure. Participants were tested in groups of one to four and were seated at individual stations with a computer screen and keyboard on a desk in front of them. The computer screen was positioned 22 cm away from the near edge of the table. The participants first read and signed an informed consent form and also read a one-page sheet of instructions for the study phase. The instructions outlined the procedure of the study phase including the consistency rating task. The experimenter answered any questions if necessary before starting the experimental program on the computer.

The procedure was very similar to the pilot Experiment 1, with two main differences. One difference was that the rating system during the study phase was changed from the pilot study. This was done to ensure that participants were reading the passages carefully and that they encoded the specific details more accurately. Most importantly, I wanted to ensure that participants were taking note of the peculiarity of the contradictory passages. Participants responded with the 1 to 4 keys on the keyboard indicating their rating of level of consistency conveyed by the two sentences of each passage, rather than degree of activity as they had done in the pilot experiment. A rating of “1” indicated “very inconsistent”, while a rating of “4” indicated “very consistent.”

The second difference was that, instead of receiving all of the instructions prior to beginning the experiment, participants received instructions for the study phase only at the onset of testing. Only after completing the study phase were participants given the instructions for the test phase. This was done because of the complexity of the three different instructional conditions. Giving the full set of instructions before the study phase could have resulted in participants forgetting precisely how they were to respond in the test phase. Also, the test phase instructions included an example passage and test sentences as well as a practice passage and test sentences that participants were required to complete according to the instructional condition they were assigned to. Across the three instructional conditions, the example and practice passages remained the same, but the answers to the test questions varied according to the given instructions. More details regarding the practice passage are provided in the following procedure section.

In the study phase, before each passage was presented, the word “READY” was displayed in the centre of the screen at which time participants were required to press the space bar to display the passage. Participants were first presented with a fixation X for 500 ms, left-justified halfway down the computer screen to orient their gaze to the correct location on the screen, after which the first sentence of the two-sentence passage appeared in the same location. They read at their own pace and pressed the space bar to proceed to the next sentence. Upon pressing the key, the first sentence was removed from the screen and the second sentence immediately appeared. Once both sentences had been read and understood, participants were required to make a 4-point rating for each passage judging the level of consistency conveyed by the whole passage. They were reminded of this task by having the phrase “Reply 1, 2, 3, 4” appear at the top of the screen when the second sentence of each passage was presented. A rating of “1” indicated “very inconsistent” while a rating of “4” indicated “very consistent.” Participants responded with the 1 to 4 keys indicating their rating of consistency. After the response was registered, the “READY” screen was displayed again after an interval of 3000 ms to prompt the start of the next trial. There was no answer time limit. Response times were measured and recorded within the experimental program.

After the study phase was completed, a message appeared in the centre of the screen instructing participants to open the door to the testing room and wait for the experimenter to come by to provide the instructions for the test phase. Each participant was then given a two-sided page of colour-coded instructions as per the instructional group the participant had been randomly assigned to. The participants were unaware that other participants may have received different instructions. The experimenter provided the test phase instructions, and asked each participant to read through carefully and to complete the practice passage and test sentences by writing in the appropriate “yes” or "no" response for each test sentence according to the instructions. For the test phase, participants were assigned to one of three instructional conditions: (1) recognize: to only accept sentences that were stated verbatim in the study phase; (2) verify: to accept sentences that were stated verbatim or that were implied from the study phase; or (3) imply: to only accept sentences that were implied, but not stated, in the study phase. The presentation of the recognition test was the same for all participants regardless of instructions.

Participants were given a few minutes to read through the instructions and complete the practice portion. After this the experimenter returned to the room, answered any questions, and checked the answers to the practice sentences. If the participant had responded incorrectly to any of the test sentences, they were given another sheet containing a different practice passage and corresponding test sentences to complete. After allowing a few more minutes for completion of the second practice passage, the experimenter checked the answers again. If there were any incorrect responses, the experimenter discussed it with the participant and explained the instructions again ensuring the participant understood the task. Once this was done, the experimenter started the test phase of the computer program before leaving the room.

At the start of the test phase, a message was displayed in the middle of screen prompting participants to switch their response keys to the marked "NO" and "YES" keys on the keyboard. The "x" and "." keys were labeled "NO" and "YES", respectively. The message also reminded participants of which sentences to respond "yes" to and which to respond "no" to according to their instructional group. Participants pressed the READY key (space bar) to start the recognition test. A fixation X appeared in the same location as in the study phase for 500 ms before each test sentence was displayed. Once the test sentence appeared, participants responded “yes” or “no” according to the given instructions. After the response was registered, the sentence was removed from the screen. Following a 1000 ms interval, the next trial automatically began, again with a fixation X preceding the test sentence. There was no time limit applied to responding during the recognition test, but participants’ response times were automatically recorded by the experimental software.

Results and model analyses

Two participants were eliminated from the recognize condition due to having a higher or equal proportion of acceptances in the absent condition than in the other four text conditions combined. One participant was eliminated from the verify condition due to accepting all targets in the control and absent conditions and only one out of 12 targets in the explicit, inference and contradictory conditions combined. Four participants were eliminated from the imply condition: one participant responded “no” to all 20 target sentences and told the experimenter “I think I did the whole experiment wrong” after completing the study; one participant had very fast response times in all conditions with an average of less than 400 milliseconds which was deemed not enough time to have fully read and understand each target sentence; and two participants were eliminated for having a higher or equal proportion of acceptances in the absent condition than in the other four conditions combined. After removing these outlying data, the final number of participants was 185: 62 in the recognize condition, 63 in the verify condition, and 60 in the imply condition. Removal of these problem data did not significantly affect condition means. The results for this experiment are displayed in Table 4.

Table 4. *Condition means for proportion "yes" responses and corresponding mean* ***Response Times (in milliseconds)*** *for Experiment 3*

Condition	Type of Instruction		
	Recognize	Verify	Imply
Explicit (Implied/Stated)	0.83 (0.22)	0.89 (0.17)	0.29 (0.26)
	3229 (996)	***3074 (1180)***	***4236 (1658)***
Inference (Implied/Not-Stated)	0.30 (0.27)	0.57 (0.28)	0.68 (0.25)
	3799 (1954)	***3906 (2197)***	***5472 (2707)***
Contradictory (Not-Implied/Stated)	0.80 (0.20)	0.87 (0.19)	0.28 (0.27)
	3603 (1009)	***3362 (1208)***	***5788 (3317)***
Control (Not-Implied/Not-Stated)	0.25 (0.23)	0.49 (0.25)	0.54 (0.28)
	3799 (1447)	***3950 (1808)***	***5325 (2946)***
Absent	0.01 (0.04)	0.04 (0.10)	0.26 (0.29)
	3499 (1498)	***3554 (1415)***	***5589 (3210)***

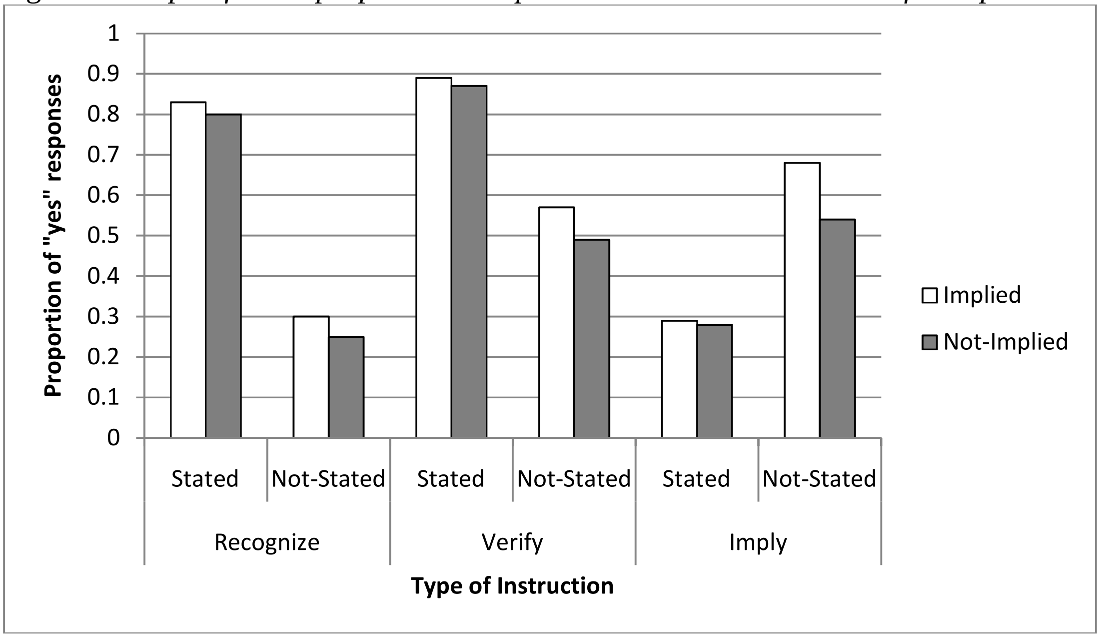
Note: Standard deviations in parentheses.

The ANOVAs reported herein are part of routine data analyses for this type of experiment. However, in this case it is the multinomial modelling that is the critical analysis. The modelling is reported after the ANOVA results.

The results were analyzed by using subjects as a random variable (F1 values), as well as by using items as a random variable (F2 values). For both types of analyses, two ANOVAs were conducted; one using proportion acceptances, and the other using response times (RTs) for "yes" responses. For the subjects-random analyses, the ANOVAs had a 2 x 2 x 3 x 5 mixed design. The three types of instructions and five counter-balanced lists were the between-subjects variables. The two statement conditions and the two implication conditions made up the within-subjects variables. Similarly, the items-random analyses also utilized a 2 x 2 x 3 x 5 mixed design, however in this case verbal set was the only between-items variable, while statement (x 2), implication (x 2), as well as instructional group (x 3) comprised the within-items variables. The absent condition was not evaluated within the factorial ANOVAs. Rather, a simple t-test was conducted between the absent and control conditions to confirm that the pattern of responding for these two conditions were significantly different from each other and from zero. Planned comparisons were also conducted to evaluate the individual effects, if any, of implication and statement. All results reported were evaluated using $\alpha = 0.05$, unless otherwise indicated.

Proportion acceptances. For proportion acceptances, there was no significant effect of list in the subjects-random analysis, $F1(4, 170) = 2.09$, $p = 0.085$, $MS_e = 0.064$. Also, there was no significant effect of verbal set in the items-random analysis, $F2(4, 15) = 0.61$, $p = 0.66$, $MS_e = 0.050$. Therefore, it was concluded that neither between- variable affected the remaining results, and were excluded from further analyses. Figure 2 presents proportion acceptance rates across all conditions. The figure provides a clearer visual which helps show the differences in patterns of responding for Statement and Implication, as well as across instructions. The absent condition was excluded from the figure since it was not a crucial aspect of the ANOVA design and analyses.

Figure 2. *Graph of mean proportion acceptance rates across conditions for Experiment 3*



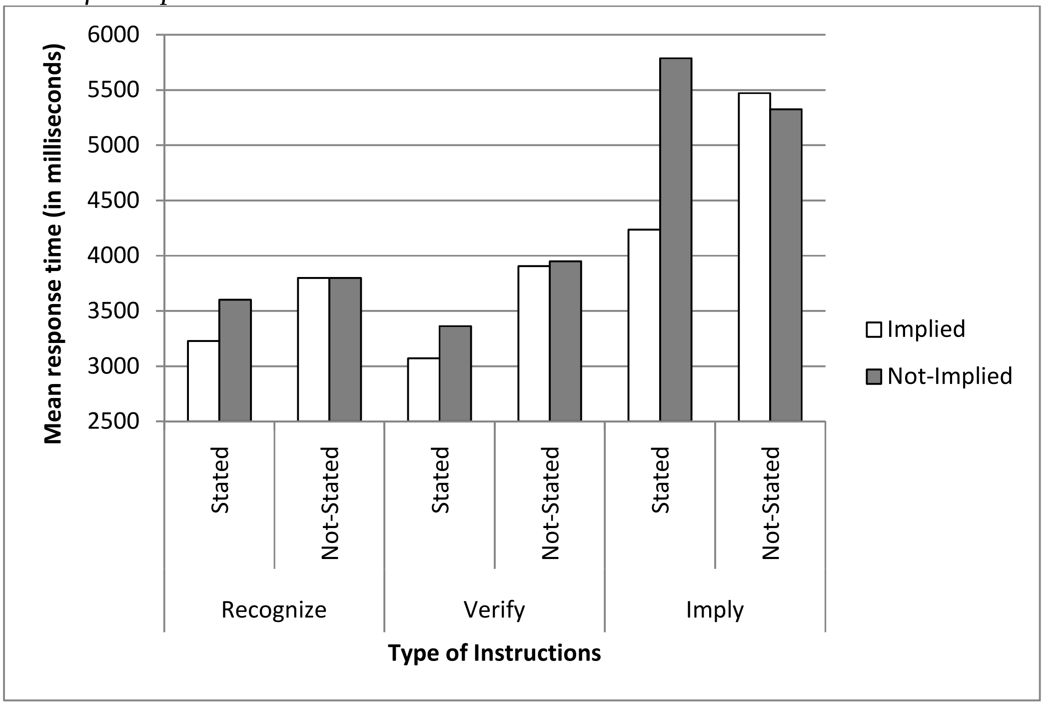
There was a significant main effect of statement, $F1(1, 170) = 98.71$, $MS_e = 0.07$, $F2(1, 15) = 65.77$, $MS_e = 0.03$. That is, stated items were accepted significantly more often than unstated items. There was also a significant main effect of implication, $F1(1, 170) = 13.81$, $MS_e = 0.04$, $F2(1, 15) = 7.48$, $MS_e = 0.03$: Implied items were accepted significantly more often than not-implied items across all conditions. The Statement x Implication interaction was also found to be significant, $F1(1, 170) = 5.71$, $MS_e = 0.04$; $F2(1, 15) = 5.48$, $MS_e = 0.02$, such that there was a significantly larger difference between both types of not-stated items (inference and control conditions) than between both types of stated items (explicit and contradictory conditions).

There was a significant effect of instruction across both analyses, $F1(2, 170) = 66.14$, $MS_e = 0.06$; $F2(2, 30) = 28.98$, $MS_e = 0.05$. This indicates that participants' pattern of responding was distinct across the three instruction conditions. The Statement x Instruction interaction was also significant, $F1(2, 170) = 189.44$, $MS_e = 0.07$; $F2(2, 30) =$

92.09, $MS_e = 0.04$, which reinstates the result above that test sentences in the stated conditions were accepted more often than those from the not-stated conditions under the Recognize and Verify instructions, but the opposite was true under the Imply instructions. The Implication x Instruction interaction was non-significant for both the subjects-random analysis and the items-random analysis, *F*s < 1, showing that the pattern of responding between items from implied and not-implied conditions did not significantly differ across instructional condition.

Simple *t*-tests were conducted for both the proportion acceptances as well as the mean response times for acceptances comparing the control and absent conditions. The absent condition test sentences were derived from passages that did not appear during study at all. This condition was compared with the control condition, which was not-stated as well as not-implied. Both of these conditions required participants to make “no” responses across the three instructions. Although I expected a low proportion of acceptances in the control condition, the rate should have still been higher than in the absent condition due to lexical overlap between control passages and their corresponding test sentences. The *t*-test revealed that there were significantly more acceptances to target probes in the control condition than in the absent condition, $t(184) = 14.66$.

Figure 3. *Graph of mean response times (RTs) in milliseconds for "yes" responses across all conditions for Experiment 3*



Response times. Mean RTs from the 15 conditions are illustrated in Figure 3.The RT analyses followed the same factorial designs as the proportion acceptance subjects-random and items-random designs. However, the pattern of results for the RT analyses was not as clear-cut as the proportion acceptance results, which can be seen in Figure 2. There was a significant effect of list for the subjects-random analysis, $F1(4, 99) = 2.61$, $MS_e = 6{,}204{,}646.85$, so it was not possible to collapse all results across list. There was no significant effect of verbal set in the items-random analysis, $F2(4,8) = 1.15$, $p = 0.40$, $MS_e = 6{,}927{,}614.50$. The effect of instruction was significant for both the subjects-random and the items-random analyses, $F1(2, 99) = 17.86$, $MS_e = 6{,}204{,}646.85$, and $F2(2, 16) = 47.81$, $MS_e = 742{,}945.76$, indicating that RT was significantly influenced by the type of instructions given, such that participants took significantly longer to respond in the imply condition, than in the recognize and verify conditions.

The effect of statement was close to significant in the subjects-random analysis, *F1*(1, 99) = 3.47, $p = 0.07$, MS_e = 2,166,664.49, and was not significant in the items-random analysis, *F2*(1, 8) = 1.83, $p = 0.21$, MS_e = 1,421,048.82. The effect of implication was significant in the subjects-random analysis, *F1*(1, 99) = 6.55, MS_e = 2,744,253.55, but was not significant in the items-random analysis, *F2*(1, 8) = 2.15, $p = 0.18$, MS_e = 1,605,306.28. The Statement x Implication interaction was significant in the subjects-random analysis, but not the items-random analysis, *F1*(1, 99) = 6.97, MS_e = 2,043,645.23, and *F2*(1, 8) = 1.73, $p = 0.23$, MS_e = 1,730,536.86.

As with proportion acceptances, a *t*-test was conducted to evaluate the response times in the control condition against the response times of the absent condition. There was no significant difference found between response times in these two conditions, $t(42) = -1.27$, $p = 0.21$.

These RT results were not indicative of many specific predictions as in the proportion acceptance analyses. The most important result from these RT analyses is that it took significantly longer to respond “yes” or “no” to test probes under the imply instructions than under the recognize and verify instructions. This suggests that the imply instructional condition is indeed more difficult for participants than either the recognize or verify instructions (Singer & Remillard, 2008).

Model fitting. The phantom recollection model was analyzed by using the proportion acceptances measured in each combination of text and instructional conditions. These values were used as the *p* value in each of the respective model equations. The models were initially going to be evaluated using Hu and Phillips (1999) GPT multinomial processing tree software. However, due to glitches in the modelling program caused when models exceeded nine trees (my phantom recollection model was comprised of 15 trees), an alternative program was sought out. I used the alternative multiTree multinomial processing tree software (Moshagen, 2010). This program was more user-friendly than GPT. I ran my phantom recollection model through both programs, GPT and multiTree, and both provided the exact same parameter estimate results. Therefore multiTree was used for all subsequent analyses. Goodness of fit testing was also conducted.

Based on previous research by Brainerd et al. (2001) and Singer and Remillard (2008), I expected that the phantom recollection model would fit the resulting data. The complexity of the research design made it difficult to flesh out individual parameter estimate predictions. The most important predictions I had were that phantom recollection would make a significant contribution to the acceptance of targets in the inference condition compared with the other three conditions. Also, recollection was predicted to be greater for the contradictory condition than for the explicit condition due to the increased distinctiveness of the contradictory passages. This increased distinctiveness was expected to cause deeper encoding of the studied passages, leading to better verbatim recollection at test.

I also compared the phantom recollection model with the simpler, two-process conjoint recognition model (Brainerd et al., 1999; 2001) and the process-dissociation inference model (Singer & Remillard, 2004), as was done by Singer and Remillard (2008). The fits of the three models was assessed using G^2, a chi-square goodness of fit statistic (Singer & Remillard, 2008).

The main distinctions among these models are that the conjoint recognition model does not propose a phantom recollection component and that the process dissociation inference model posits an illusory recollection parameter rather than a simple recollection parameter. The latter parameter would cause acceptance of a target that had been only inferred from the studied text. However, the process dissociation inference model is missing a recognize-to-reject process, which means it does not allow the reader to be able to correctly recollect the studied material and make the judgment that the presented probe was only implied by the text, leading to a “no” response. The literature has shown that the phantom recollection model has a generally better fit to discourse inference material than both of the two-process models because it encompasses both true recollection and phantom recollection (Brainerd et al., 2001; Singer & Remillard, 2008). However, if we find that either the conjoint recognition model or the process dissociation inference model results in a better fit to the data, it would be the preferred model due to use of fewer parameters (i.e., principle of parsimony). I did not expect that this would be the case. I predicted that the phantom recollection model would result in a superior fit to the experimental data.

The fits of the competing models were assessed using a chi-square goodness-of-fit statistic, G^2. A significant G^2 value would indicate the lack of fit of a model. Specific null hypotheses were evaluated by subtracting G^2 for the full model from the G^2 associated with the submodel constrained by the null hypothesis. The difference between these two values is also a G^2 statistic with one degree of freedom because the full model and submodel differ by one free parameter.

The formulas for the probability of accepting a target sentence for both the conjoint recognition model and the process dissociation inference model are provided in the Appendix. Each of these models is composed of 11 free parameters, whereas the experimental procedure results in 15 observations. This leads to four degrees of freedom being associated with the G^2 statistic for both two-process models. On the other hand, the phantom recollection model is composed of 14 free parameters, which are measured from 15 observations, leading to one degree of freedom associated with the G^2 statistic for the model. The two parameters E_e and E_u represented the erroneous recollection of other targets that were related to the test probes. The text passages used in this study were unrelated to one another, so I assigned both of the *E* parameters a constant value of 0.05 throughout the modelling. This value was chosen from the value directly measured by Brainerd, Reyna, and Estrada (2006) in their testing of explicit probes from narrative texts. Setting both *E* parameters to a constant value resulted in the phantom recollection model gaining two extra degrees of freedom, thus having three degrees of freedom in total.

Rather unfortunately, none of the three models provided a satisfactory fit to the data. The phantom recollection model resulted in the least significant lack-of-fit of the three model, $G^2(3) = 30.56$, $p < 0.001$. The conjoint recognition model resulted in a significant lack-of-fit of $G^2(4) = 34.43$, $p < 0.001$, and the process dissociation inference model also resulted in a significant lack-of-fit of $G^2(4) = 71.43$, $p < 0.001$. The parameter estimates for the three models are shown in Table 5. P_{i} and R_{c} were both indistinguishable from zero, $G^2(1) = 0.00$ (as the model fit was exactly the same with and without setting either $P_{\mathrm{i}} = 0$ or $R_{\mathrm{c}} = 0$).

Table 5. *Parameter estimates for the phantom recollection model, the conjoint recognition model, and the process dissociation inference model.*

Model	Condition	*R*	*F*	*P*	*E*	G_R	G_V	G_I
Phantom Recollection	Explicit (IS)	0.61	0.64		0.05[a]			
	Inference (I-S)	0.26	0.40	0.00[b]				
	Contradictory (-IS)	0.60	0.57		0.05[a]			
	Control (-I-S)	0.00[b]	0.35					
	Absent					0.01	0.06	0.27
Conjoint Recognition	Explicit (IS)	0.61	0.63					
	Inference (I-S)	0.26	0.40					
	Contradictory (-IS)	0.59	0.57					
	Control (-I-S)	0.00[b]	0.35					
	Absent					0.01	0.06	0.27
Process Dissociation Inference	Explicit (IS)	0.61	0.62					
	Inference (I-S)	0.00	0.43					
	Contradictory (-IS)	0.60	0.56					
	Control (-I-S)	0.00[b]	0.34					
	Absent					0.00	0.08	0.30

Note: [a] refers to parameter estimates that were set at a constant value; [b] refers to being statistically indistinguishable from 0; IS = Implied/Stated, I-S = Implied/Not-Stated. -IS = Not-Implied/Stated, -I-S = Not-Implied/Not-Stated

Because none of the models seemed to provide a satisfactory fit to the data, I next tested the fit of my data with the three models using only the explicit, inference, and absent conditions. This design matched the design used by Singer and Remillard (2008) in their evaluation of the same three models. Fitting the data with the phantom recollection model to just these three conditions did in fact result in a very satisfactory fit with $G^2(1) = 1.97$, $p = 0.16$. The conjoint recognition model provided a slightly less satisfactory fit of $G^2(2) = 4.04$, $p = 0.13$, whereas the process dissociation inference model showed a significant lack-of-fit even for these three conditions, $G^2(2) = 44.46$, $p = 0.00$. Although the phantom recollection model fit the data of the explicit, inference, and absent conditions exceptionally well, the value of the P_i parameter was still indistinguishable from zero.

Discussion

The fit of the phantom recollection model was slightly better than the conjoint recognition model for both the full data set as well as the explicit, inference, and absent subset of the data, but the P_i parameter estimate was negligible. This could have been due to a few factors. First, perhaps the participants had not encoded the study materials deeply enough to allow for integration of the inference material into the situation model of the given text passages. This would be similar to the result obtained by Singer and Remillard (2008) in which when participants were given a shallow orienting task during study, such as being asked to read text passages and look for spelling errors. This study had found that the shallow orienting task prevented the extraction of text inferences and gist representations to the situation model, which consequently caused no effect of phantom recollection (Singer & Remillard, 2008, Experiment 3). Even though we had implemented the novel consistency rating of study passages in Experiment 3, perhaps it was not a strong enough orienting task to allow participants to process and encode the implied ideas from the passages. Perhaps placing greater emphasis on comprehension of the text passages during initial instructions could have countered this.

Second, it could be that the materials used in this study did not manipulate the inference generation strongly enough. Brainerd and Reyna (2002) stated that falsely remembered information is consistent with the gist of readers' experience, but perhaps the passages in this experiment did not elicit the desired inferences. Many of the inferences evoked from the created passages were elaborative inferences rather than bridging inferences. Because bridging inferences have been found to be more consistently and deeply encoded than elaborative inferences, this could also have led to the inferences we were expecting to be generated, to not be generated as strongly in all participants. Bridging inferences are more often than not necessary to preserve the coherence of a text passage, whereas elaborative inferences are merely predictions or elaborations about the text (Keefe & McDaniel, 1993; Potts et al., 1988; Schmalhofer et al., 2002; Singer, 1980; Singer & Ferreira, 1983). Because elaborative inferences are not required to be generated to preserve passage coherence, perhaps in this experiment they were not generated strongly enough, or at all.

The remaining parameter estimates seemed to be reasonable across the three models. The three guessing parameters were directly measured from the proportion of acceptances from the absent condition test probes. The Recollection parameters in the explicit and contradictory conditions were nearly identical, indicating that there may not have been a distinctiveness advantage for the contradictory passages. This could also be attributed to the possibility that participants had read the study passages using a shallow encoding orientation. As mentioned earlier, the consistency rating task used in this experiment may not have led to the desired level of encoding of text ideas. Perhaps if the study task was made to reflect deeper encoding strategies a distinctiveness effect could have been found for the contradictory condition. Another possibility is that the contradictory passages were not illogical or bizarre enough to cause readers to encode them as being distinctive. The Familiarity parameters were almost the same for the explicit and contradictory passages as well. Taken with the Recollection parameter pattern, this is indicative of very similar processes being at play for both explicit text ideas and explicitly-stated-but-contradicting text ideas.

Familiarity had a larger influence than Recollection on target acceptances for both the inference condition and the control condition. This makes sense, since for both of these conditions the target probes had not been presented explicitly during study, so a higher sense of familiarity was more likely to evoke “yes” responding. Recollection was zero for the control condition across the three models. This is also feasible, since any recollection of the control condition passages would have resulted in participants remembering that the study passage was quite different from the presented test probe, and consequently lead to “no” responding. Recollection in the inference condition was zero only for the process dissociation inference model. Since the R parameter in this model was representative of illusory recollection rather than true recollection, this result also makes sense.

General Discussion

I would like to discuss briefly three theoretical issues that have been debated in the literature regarding the multiprocess procedures I used in my thesis project. I will discuss the simplified CR procedure, the issue of parameter heterogeneity, and the apparent advantages and disadvantages of single-process versus dual- or multi-process theories of memory.

Simplified CR Procedure

The CR model as reported herein has only been implemented a handful of times in the literature (Brainerd et al., 1999, 2001; Erdfelder et al., 2009; Singer & Remillard, 2004, 2008; Stahl & Klauer, 2008). The limited use of the experimental paradigm and related model has been attributed to the complex design involving three distinct sets of instructions as a between-subjects variable. This design element requires that researchers recruit three times as many participants than if the paradigm could be conducted using a single set of participants. Another potential shortcoming of the original CR paradigm was that the imply instructions seemed to be too complex for some participants to reliably understand (Stahl & Klauer, 2008).

Stahl and Klauer (2008) suggested a new paradigm and model which they labelled the simplified CR model. In their design, a single group of participants would read a study list just as in the original CR paradigm. At test they would receive verbatim targets, related distractors, and unrelated distractors. The crucial difference from the original paradigm was that instead of a recognition test where different groups of participants would have received one of three different sets of response instructions, participants were asked to classify each test item according to the type of probe they believed it was. Participants were instructed to respond “target” if they believed the item had been presented verbatim during study, to respond “related” if they believed that the item was a related distractor, and to respond “new” to items believed to be unrelated to the study list. Stahl and Klauer found that this simplified model and paradigm elicited very similar results and parameter estimates as the original CR model, while utilizing a smaller number of participants than the original model as well as having a singular set of instructions.

However, the simplified CR procedure has not yet been extended to incorporate additional processes such as phantom recollection or erroneous recollection of related targets. Therefore, it would not have been appropriate to use in the current study, as one of the objectives was to compare and contrast the results and model fits of the three models; namely the phantom recollection model, the conjoint recognition model, and the process dissociation inference model.

Another shortcoming of the simplified CR procedure is that the design is too complicated to be administered to children, which was explicitly noted by the developers (Stahl & Klauer, 2008). Brainerd, Holliday, and Reyna (2004) successfully administered the original conjoint recognition experimental paradigm and model using DRM style word lists with children aged 5- to 14-years old. The three types of recognition test instructions were administered through auditory means, and were tailored so that even the youngest children could easily understand the task. For example, the children heard the following instructions in the Imply condition:

> Every time I read you a word, I want you to say “yes” if it’s a word like “carrot” that’s the name of a new vegetable that you didn’t just hear on the list of vegetables. But, I want you to say “no” to any word that’s not a new vegetable name. Be careful not to say “yes” to any old vegetables, like “celery,” that you just heard on the list. Please don’t guess. Just say “yes” whenever I read a word that’s a new vegetable name that you didn’t hear, and say “no” to every other word that I read (Brainerd et al., 2004, pp. 521).

They were able to track the development of the processes of recollection and familiarity in these young children. In this sense, the original CR model would be deemed superior to the simplified model due to its increased breadth of applicability. Perhaps in the future the simplified model could be modified to address these shortcomings.

Parameter Heterogeneity

The issue of parameter heterogeneity stems from the fact that the multinomial models addressed in this study, as well as the simplified CR model, rely on the assumption that parameters are homogenous across items and participants. If this assumption is violated it can lead to mistaken rejection of models as well as to biased parameter estimates. It seems intuitive when one thinks about this that parameter estimates are likely to differ from person to person. There is likely an average value associated with particular cognitive processes, but the parameter estimate can vary around that mean value for individual persons.

Stahl and Klauer (2008) tested this in their research and found that even though parameter heterogeneity was found for a large proportion of their conditions, this heterogeneity across participants did not affect their reported results and conclusions. This is an issue to be aware of in research involving multinomial modeling but needs to be studied in more detail and would be worthwhile to investigate in future studies.

Single-Process versus Multi-Process Memory Theories

There are two large and opposing camps in the literature of researchers who either believe that memory functions as the result of a single mental process, or else those who believe human memory is the function of two or more distinct processes that can work together, independently of one another, or even in opposition to affect memory for items

(Diana, Reder, Arndt, & Park, 2006; Dunn & Kirsner, 1988; Ratcliff, Van Zandt, & McKoon, 1995). There are ongoing speculations in the field as to which view is the correct one.

Those who favour single-process theories assert that a single-process based on a familiarity- or strength-based process is able to account for memory for items. Due to the principle of parsimony, one process is more favourable than two or more processes. On the other hand, supporters of dual- and multi-process theories claim that a single-process account is not able to explain all types of memory phenomena and that a recollection process is necessary in many different situations (Diana et al., 2004; Singer & Remillard, 2008). Memory models comprised of more than two processes, such as the phantom recollection model, are relatively new in the field, thus the literature focuses on comparing single- and dual-process models. However, many of the assumptions made of dual-process models are also applicable to more general multi-process models.

Diana et al. (2006) reviewed several memory phenomena that were unable to be accounted for by single-process models, but that were easily explained using a single dual-process account. Certain recognition paradigms in which a single-process was not sufficient to account for the data included plurality reversal (Malmberg, Holden, & Shiffrin, 2004) and the word frequency mirror effect (Diana et al., 2006). These types of findings have required single-process theorists to add additional assumptions to their models, and to even add a recall-like process to be able to account for the data. Also, some newer single-process models have taken the position that recollection is used in some types of recognition tasks, but not for simple single-item recognition. Alternatively, dual-process accounts do not need to instate additional assumptions for some situations but not others. In this sense dual-process theories, and by extension certain multi-process theories, are more parsimonious than single-process accounts (Diana et al., 2006)

Conclusions

From this study, I was able to lend some support for the application of multinomial models to the area of discourse comprehension. I provided further evidence for the differences in verbatim and gist representations of explicitly stated text versus inferences drawn from text passages. The main goal was to lend support to findings by Singer and Remillard (2008) that the phantom recollection model was a superior fit to discourse materials than either of the conjoint recognition or process dissociation inference models. My results did not fit as nicely as previous research has shown, but the pattern of results was in the right direction. Multi-process theories such as the phantom recollection model are new to the field of memory research, so more research is still needed to test whether or not three processes are better than two processes. However, I was able to demonstrate the separate contributions of recollection and familiarity on memory for discourse. With further modifications and possible improvements to the materials and methodologies used, I believe this line of research would be able to lend substantial support to the advocacy of the phantom recollection process and model in text memory.

References

Albrecht, J. E. & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1459-1468.

Argument. (n.d.). In *Oxford Dictionaries Online*. Retrieved from http://oxforddictionaries.com/view/entry/m_en_gb0039220#m_en_gb0039220.

Brainerd, C. J., Holliday, R. E., & Reyna, V. F. (2004). Behavioral measurement of remembering phenomenologies: So simple a child can do it. *Child Development, 75*, 505-522.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*, 164-169.

Brainerd, C. J., Reyna, V. F., & Estrada, S. (2006). Recollection rejection of false narrative statements. *Memory, 14*, 672-691.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review, 106*, 160-179.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 307-327.

Chan, J. C. K. & McDermott, K. B. (2006). Remembering pragmatic inferences. *Applied Cognitive Psychology, 20*, 633-639.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review, 13*, 1-21.

Dunn, J. C. & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95*, 91-101.

Erdfelder, E., Auer, T., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie / Journal of Psychology, 21*, 108-124.

Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 981-995.

Graesser, A. C., Millis K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163-189.

Haviland, S. E. & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 512-521.

Hu, X. & Phillips, G. A. (1999). GPT.EXE: A powerful tool for the visualization and analysis of general processing tree models. *Behavior Research Methods, Instruments, & Computers, 31*, 220-234.

Hunt, R. R. & Lamb, C. A. (2001). What causes the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1359-1366.

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

Keefe, D. E. & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language, 32,* 446-463.

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

Kintsch, W., & Bates, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 150-159.

Kintsch, W. & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology, 5*, 257-274.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior, 14*, 196-214.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133-159.

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30,* 319-330.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252-271.

McDaniel, M. A., Dornburg, C. C., & Guynn, M. J. (2005). Disentangling encoding versus retrieval explanations of the bizarreness effect: Implications for distinctiveness. *Memory & Cognition, 33*, 270-279.

McDaniel, M. A., Einstein, G. O., DeLosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 422-435.

McDermott, K. B. & Chan, J. C. K. (2006). Effects of repetition on memory for pragmatic inferences. *Memory & Cognition*, *34*, 1273-1284.

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language, 26*, 165-187.

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*, 42-54.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language, 27*, 399-415.

Predicate. (n.d.). In *Oxford Dictionaries Online*. Retrieved from http://oxforddictionaries.com/view/entry/m_en_gb0656650#m_en_gb0656650.

Radvansky, G. A. & Copeland, D. E. (2000). Functionality and spatial relations in memory and language. *Memory & Cognition, 28,* 987-992.

Ratcliff, R. & McKoon G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning & Verbal Behavior, 17*, 403-417.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1995). Process dissociation, single-process theories, and recognition memory. *Journal of Experimental Psychology: General, 124*, 352-374.

Rinck, M., Hahnel, A., & Becker, G. (2001). Using temporal information to construct, update, and retrieve situation models of narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 67-80.

Saldana, D. & Frith, U. (2007). Do readers with autism make bridging inferences from world knowledge? *Journal of Experimental Child Psychology, 96*, 310-319.

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes, 33*, 105-132.

Shears, C. & Chiarello, C. (2004). Knowledge-based inferences are not general. *Discourse Processes*, *38*, 31-55.

Singer, M. (1980). The role of case-filling inferences in the coherence of brief passages. *Discourse Processes, 3*, 185-201.

Singer, M. (1990). *Psychology of language: An introduction to sentence and discourse processes*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Singer, M. & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior, 22*, 437-448.

Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language, 31*, 507-524.

Singer, M. & Remillard, G. (2004). Retrieving text inferences: Controlled and automatic influences. *Memory & Cognition,*1223-1237.

Singer, M. & Remillard, G. (2008). Veridical and false memory for text: A multiprocess analysis. *Journal of Memory and Language, 59*, 18-35.

St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience, 9,* 776-787.

Stahl, C. & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 570-586.

Suh, S. & Trabasso, T. (1993). Inferences during reading: Converging evidence for discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language, 32*, 279-300.

Tompkins, C. A., Fassbinder, W., Lehman Blake, M., Baumgaertner, A., & Jayaram, N. (2004). Inference generation during text comprehension by adults with right hemisphere brain damage: Activation failure versus multiple activation. *Journal of Speech, Language, and Hearing Research, 47,* 1380-1395.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Wiley, J. & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes, 36*, 109-129.

Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 55-89.

Appendix A

Materials created from Potts et al., 1988.

KEY: First set of parentheses indicate the (not-implied/implied) manipulation. Second set of parentheses indicate the (not-stated/stated) manipulation.

1. Sharon cautiously pulled on her (beach clogs/bathing cap). The bather overlooked the piece of glass in the shallow water and (stepped/cut her foot) on it.

2. The living room of Sally the artist had an odd (fleecy/purple) wall. No longer able to control his fury, the boyfriend (hurled/broke) the delicate porcelain vase against the wall.

3. Joan kept (tame/honey) bees to pollinate her large apple orchard. A large swarm emerged from the hive and (landed on/stung) her hand.

4. The problem tooth was (brown from the coffee/the bottom-left molar). Ann opened her mouth as the dentist (worked on/drilled) it.

5. The careful Mercedes owner had installed (an immobilizer/fancy tires). The car thief pried the car window and (broke in/drove away).

6. The doctor warned him that the cast would (require standing except in bed/help his leg heal faster). After standing through the long debate, the tired speaker (went over to/sat down in) the chair.

7. The owner made sure to give his dog some (water/exercise) before the outing. After playing in the sun, the dog (jumped in/drank from) a fountain.

8. She wanted to tell her mother she had broken her right (hand/ankle). The debutante went to her desk and (picked up some paper/wrote a letter).

9. The trapeze artist was the most (experienced/well-known) member of his team. In the middle of the act, the performer suddenly (lost his grip/fell).

10. The assistant rushed to (prepare the bill/change the oil) for the car owner. Before leaving, he went over to the bathroom sink (to clean up/wash his hands).

11. The student found the course to be very (easy/interesting). With his exam coming up, he (took out/studied) his notes.

12. The wooden sailing ship's hull had been newly (strengthened/painted) before the Arctic voyage. In dense fog an iceberg suddenly (struck/sunk) the schooner.

13. The painter (sealed/moved) the can of paint when he was done. Then the worker accidentally kicked it and (knocked over/spilled) it.

14. The baseball player had (sprained his wrist/washed his uniform) the day before the big game. At practice, he picked up the ball and (gave/threw) it to his teammate.

15. Near the woods, Mildred looked through her bag for (her phone/a cigarette). The Calgary salesgirl's carelessness (had a tragic result/caused a tragic fire).

16. Don put on his (indestructible/favourite) shirt before ascending the roof. While the labourer fixed a shingle, the garment (caught/tore) on a loose nail.

17. The decorator particularly liked a (bright/luxurious) room. When sun streamed through the window he (noticed/closed) the blinds.

18. The first-time marathoner was extremely (conditioned/nervous). As he reached the end of the race, he (struggled to/collapsed at) the finish line.

19. The pianist did her best ever in playing for the (examiners/audience). The listeners (acknowledged/applauded) her when she rose from her seat.

20. George's decision to build on (higher/lower) ground now had huge implications. He watched in awe when the rising river (surrounded/flooded) his cabin.

21. The (calm/capable) contestant had prepared for this event for an extremely long time. After being announced the winner, she (reacted/cried) with emotion.

22. The (gentle/cautious) policeman approached the dangerous suspect. When the culprit tried to flee, the cop drew his gun and (subdued/shot) him.

23. Mary was (terrified/prepared) when the severe labour pains began. She (went into a coma/gave birth) that evening in the hospital.

24. Superman's prison was controlled with undetectable (kryptonite rays/video monitoring). At the stroke of midnight the Man of Steel (pulled on/bent) the cell bars.

25. Surprisingly, the (husband/intruder) was discovered in the kitchen making a salad. The disgusted wife picked up a knife and (approached/stabbed) him.

26 The surgeon had decided on a (local/reliable) anesthetic. The patient (lay quietly/slept) as the incision was made.

27. The ball sailed into the arms of the (goalkeeper/receiver) at the goal line. The crowd cheered as the official signaled (the result/a touchdown).

28. The heaping table astounded the (waiters/diners). They took the cutlery and (went into action/began to eat).

29. When he transferred from Paris, he still had to take the required (French/chemistry) course. Jacques did no studying and (wrote/failed) the midterm.

30. Jan was (careful to drive slowly through/listening to the radio at) the speed trap. The police car turned on its flashing red lights and (sped up behind/ticketed) her car.

31. The onlookers grinned as the (hostile actress and actor/loving bride and groom) recited the wedding vows in the rehearsal. The couple (turned to/kissed) one another.

32. The crew didn't know that the bombs had been (disabled/enhanced). They dropped them from the plane and watched them (strike/explode on) the ground.

33. The robin was extremely (immature/hungry). The little bird spread his wings and (looked/flew) down to the ground.

34. The teacher didn't know that the equipment was (broken/new). She heated the oven and (put in/baked) the birthday cake.

35. The (Queen/pleased diner) enjoyed the restaurant feast. When the bill was tallied, she (complimented/paid) the owner.

36. Before the big race, he made sure to (oil his wheelchair/tie his laces). Stretching his muscles at the starting line, the competitor began to (accelerate/run) at the whistle.

37. Steve looked out and saw that the grass looked very (green/dry). The owner went outside and decided to (cut/water) it.

38. The mover had just (finished/started) the huge task. He approached the heavy box in the back of the truck and(looked at/lifted) it.

39. The crew was astonished as the (stuntwoman/actress) disappeared from the roof of the 14th story building. Moments later, the medic on set (ran to check on her/pronounced her dead).

40. Phil's a new bicycle had (tubeless tires/an advanced braking system). On his very first ride, a large spike in the road (marked/flattened) his front tire.

Appendix B

Phantom Recollection Model:

Explicit condition (implied/stated)

$p_{er} = (R_e) + (1 – R_e)(1 – E_e)(F_e) + (1 – R_e)(1 – E_e)(1 – F_e)(G_r)$ (B1)

$p_{ev} = (R_e) + (1 – R_e)(E_e) + (1 - R_e)(1 – E_e)(F_e) + (1 – R_e)(1 – E_e)(1 – F_e)(G_v)$ (B2)

$p_{ei} = (1 – R_e)(E_e) + (1 – R_e)(1 – E_e)(F_e) + (1 – R_e)(1 – E_e)(1 – F_e)(G_i)$ (B3)

Inference condition (implied/not-stated)

$p_{ir} = (1 – R_i)(P_i) + (1 – R_i)(1 – P_i)(F_i) + (1 – R_i)(1 – P_i)(1 – F_i)(G_r)$ (B4)

$p_{iv} = (R_i) + (1 – R_i)(P_i) + (1 – R_i)(1 – P_i)(F_i) + (1 – R_i)(1 – P_i)(1 – F_i)(G_v)$ (B5)

$p_{ii} = (R_i) + (1 – R_i)(1 – P_i)(F_i) + (1 – R_i)(1 – P_i)(1 – F_i)(G_i)$ (B6)

Contradictory condition (not-implied/stated)

$p_{ur} = (R_u) + (1 – R_u)(1 – E_u)(F_u) + (1 – R_u)(1 – E_u)(1 – F_u)(G_r)$ (B7)

$p_{uv} = (R_u) + (1 – R_u)(E_u) + (1 – R_u)(1 – E_u)(F_u) + (1 – R_u)(1 – E_u)(1 – F_u)(G_v)$ (B8)

$p_{ui} = (1 – R_u)(E_u) + (1 – R_u)(1 – E_u)(F_u) + (1 – R_u)(1 – E_u)(1 – F_u)(G_i)$ (B9)

Control condition (not-implied/not-stated)

$p_{cr} = (1 – R_c) + (1 – R_c)(F_c) + (1 – R_c)(1 – F_c)(G_r)$ (B10)

$p_{cv} = (1 – R_c) + (1 – R_c)(F_c) + (1 – R_c)(1 – F_c)(G_v)$ (B11)

$p_{ci} = (1 – R_c) + (1 – R_c)(F_c) + (1 – R_c)(1 – F_c)(G_i)$ (B12)

Absent condition (guessing)

$p_{gr} = G_r$ (B13)

$p_{gv} = G_v$ (B14)

$p_{gi} = G_i$ (B15)

Conjoint Recognition Model:

Explicit Condition (Implied/Stated)

$p_{er} = R_e + (1-R_e)(F_e) + (1-R_e)(1-F_e)(G_r)$

$p_{ev} = R_e + (1-R_e)(F_e) + (1-R_e)(1-F_e)(G_v)$

$p_{ei} = (1-R_e)(F_e) + (1-R_e)(1-F_e)(G_i)$

Inference condition (Implied/Not-Stated)

$p_{ir} = (1-R_i)(F_i) + (1-R_i)(1-F_i)(G_r)$

$p_{iv} = R_i + (1-R_i)(F_i) + (1-R_i)(1-F_i)(G_v)$

$p_{ii} = R_i + (1-R_i)(F_i) + (1-R_i)(1-F_i)(G_i)$

Contradictory condition (Not-implied/Stated)

$p_{ur} = R_u + (1-R_u)(F_u) + (1-R_u)(1-F_u)(G_r)$

$p_{uv} = R_u + (1-R_u)(F_u) + (1-R_u)(1-F_u)(G_v)$

$p_{ui} = (1-R_u)(F_u) + (1-R_u)(1-F_u)(G_i)$

Control condition (Not-implied/Not-stated)
$p_{cr} = (1\text{-}R_c)(F_c) + (1\text{-}R_c)(1\text{-}F_c)(G_r)$
$p_{cv} = (1\text{-}R_c)(F_c) + (1\text{-}R_c)(1\text{-}F_c)(G_v)$
$p_{ci} = (1\text{-}R_c)(F_c) + (1\text{-}R_c)(1\text{-}F_c)(G_i)$

Absent condition (Guessing)
$p_{gr} = G_r$
$p_{gv} = G_v$
$p_{gi} = G_i$

<u>Process Dissociation Inference Model:</u>

Explicit Condition (Implied/Stated)
$p_{er} = R_e + (1\text{-}R_e)(F_e) + (1\text{-}R_e)(1\text{-}F_e)(G_r)$
$p_{ev} = R_e + (1\text{-}R_e)(F_e) + (1\text{-}R_e)((1\text{-}F_e)(G_v)$
$p_{ei} = (1\text{-}R_e)(F_e) + (1\text{-}R_e)(1\text{-}F_e)(G_i)$

Inference Condition (Implied/Not-Stated)
$p_{ir} = R_i + (1\text{-}R_i)(F_i) + (1\text{-}R_i)(1\text{-}F_i)(G_r)$
$p_{iv} = R_i + (1\text{-}R_i)(F_i) + (1\text{-}R_i)(1\text{-}F_i)(G_v)$
$p_{ii} = (1\text{-}R_i)(F_i) + (1\text{-}R_i)(1\text{-}F_i)(G_i)$

Contradictory Condition (Not-Implied/Stated)
$p_{ur} = R_u + (1\text{-}R_u)(F_u) + (1\text{-}R_u)(1\text{-}F_u)(G_r)$
$p_{uv} = R_u + (1\text{-}R_u)(F_u) + (1\text{-}R_u)(1\text{-}F_u)(G_v)$
$p_{ui} = (1\text{-}R_u)(F_u) + (1\text{-}R_u)(1\text{-}F_u)(G_i)$

Control Condition (Not-Implied/Not-Stated)
$p_{cr} = R_c + (1\text{-}R_c)(F_c) + (1\text{-}R_c)(1\text{-}F_c)(G_r)$
$p_{cv} = R_c + (1\text{-}R_c)(F_c) + (1\text{-}R_c)(1\text{-}F_c)(G_v)$
$p_{ci} = (1\text{-}R_c)(F_c) + (1\text{-}R_c)(1\text{-}F_c)(G_i)$

Absent Condition (guessing)
$p_{gr} = G_r$
$p_{gv} = G_v$
$p_{gi} = G_i$