Aging and predicting inferences: A diffusion model analysis

Gail McKoon *, Roger Ratcliff

The Ohio State University, United States

ABSTRACT

In the domain of discourse processing, it has been claimed that older adults (60–90-year-olds) are less likely to encode and remember some kinds of information from texts than young adults. The experiment described here shows that they do make a particular kind of inference to the same extent that college-age adults do. The inferences examined were "predictive" inferences such as the inference that something bad would happen to the actress for the sentence "The director and cameraman were ready to shoot close-ups when suddenly the actress fell from the 14th story" (McKoon & Ratcliff, 1986). Participants read sentences like the actress one and then later they were asked to decide whether words that expressed an inference (e.g., "dead") had or had not appeared explicitly in a sentence. To directly compare older adults' performance to college-age adults' performance, we used a sequential sampling diffusion model (Ratcliff, 1978) to map response times and accuracy onto a single dimension of the strength with which an inference was encoded. On this dimension, there were no significant differences between the older and younger adults.

© 2012 Elsevier Inc. All rights reserved.

Introduction

For many of the experimental procedures used in cognitive psychology, older adults are slower than young adults and they are sometimes less accurate. However, for reading, common observation suggests that there may be little or no impairment. Many older adults appear to read fiction, nonfiction, email, and cooking recipes easily, perhaps as easily as young adults. In the experiment reported here, we investigated one aspect of older adults' comprehension of the meanings of sentences and their memory for them. Specifically, we looked at comprehension and memory for information that is not stated explicitly in a text but can be inferred.

Consider the sentence *The director and cameraman were ready to shoot close-ups when suddenly the actress fell from the 14th story* (McKoon & Ratcliff, 1986). The sentence carries the strong inference, what we call a predictive infer-

ence, that the actress will die. Following McKoon and Ratcliff (1986), many studies have investigated predictive inferences and other inferences like them (e.g., Beeman et al., 1994; Calvo & Castillo, 1996; Calvo, Castillo, & Estevez, 1999; Casteel, 2007; Cook, Limber, & O'Brien, 2001; Fincher-Kiefer, 1993, 1995; Gueraud, Tapiero, & O'Brien, 2008; Keefe & McDaniel, 1993; Klin, Guzman, & Levine, 1999; Lassonde & O'Brien, 2009; Linderholm, 2002; Linderholm & van den Broek, 2002; Murray & Burke, 2003; Murray, Klin, & Myers, 1993; Peracchi & O'Brien, 2004; Sanford & Garrod, 2005; van den Broek, 1990; Weingartner, Guzman, Levine, & Klin, 2003).

For college-age adults, the conclusion has been that predictive inferences are encoded minimally (see Gerrig & O'Brien, 2005, and van den Broek, Rapp, & Kendeou, 2005, for reviews). They are not instantiated as specific lexical items but rather as general features that could be instantiated by several lexical items. For the actress sentence, the encoded information might be "something bad happened," which would be consistent with, for example, "hurt" or "blood" as well as "dead". Also, it should be mentioned, death is not an absolutely necessary outcome. A 29-year-old window-washer fell 47 floors from an apartment

* Corresponding author. Address: Department of Psychology, The Ohio State University, Columbus, OH 43210, United States. Fax: +1 614 688 3984.

E-mail address: mckoon.1@osu.edu (G. McKoon).

building and survived, as did a 29-year-old man who plunged 17 stories in the atrium of a hotel in Minneapolis, a 22-year-old amateur sky diver who went into free fall more than a mile above the earth (New York Times, 2007), and a cat who fell 19 stories from a high-rise window (Associated Press, 2012).

Our question here was whether older readers understand predictive inferences and encode them into memory to the extent that young adults do. Predictive inferences have been classified as elaborative inferences because they add information to what is explicitly stated. Some studies suggest that older adults do encode some types of elaborative inference (e.g., Morrow, Stine-Morrow, Leirer, Andrassy, & Kahn, 1997; Radvansky, Copeland, Berish, & Dijkstra, 2003; Radvansky, Copeland, & Zwaan, 2003; Radvansky & Dijkstra, 2007; Radvansky, Zwaan, Curiel, & Copeland, 2001; Stine-Morrow, Gagne, Morrow, & DeWall, 2004; Stine-Morrow, Morrow, & Leno, 2002) but others that they do not (e.g., Cohen, 1979, 1981; Dixon, Simon, Nowak, & Hultsch, 1982; Hamm & Hasher, 1992; Hess, 1995; Light & Capps, 1986; Noh & Stine-Morrow, 2009; Stine, 1990; Stine & Wingfield, 1988; Till, 1985; Till & Walsh, 1980; Zacks, Hasher, Doren, Hamm, & Attig, 1987; Zelinski & Miura, 1990).

To our knowledge, only two studies (Valencia-Laver & Light, 2000; Zipin, Tompkins, & Kasper, 2000) have looked specifically at whether older adults understand elaborative inferences of the actress-dead kind, and for both the results were equivocal. For the experiment described below, we used a new, more analytic, procedure and found no significant differences between older and young adults.

How can performance by older adults be compared to performance by young adults?

Comparisons of performance between young and older adults face three critical problems. One is that older adults often set more conservative speed/accuracy response criteria than young adults, that is, they are more concerned to avoid errors even if doing so slows performance (e.g., Ratcliff, Thapar, & McKoon, 2001, 2003, 2004, 2006a, 2006b, 2007, 2010, 2011; Spaniol, Madden, & Voss, 2006; Thapar, Ratcliff, & McKoon, 2003; the Ratcliff et al. papers are henceforth referenced as RTM). The second is that older and young adults often have different baseline levels of performance: older adults' overall response times (RTs) tend to be slower and, depending on the task, their overall accuracy may be higher than young adults' or lower. The differences between older and young adults in response criteria and baselines mean that their performance cannot be directly compared. For example, suppose the word "daffodil" was presented for lexical decision. Older adults' accuracy might be better than young adults' because the quality of the information they have in lexical memory about "daffodil" is better. Or the quality might actually be worse for them and their higher accuracy the result of more conservative response criteria. For RTs, older adults might be slower than young adults because their criteria are more conservative or because the quality of their information is worse, or both. These two problems lead to the third, a scaling problem. Continuing the lexical decision

example, older adults might have a 50 ms difference between conditions on a baseline of 800 ms and young adults a 25 ms difference on a baseline of 500 ms. Older adults might have a 2% difference in accuracy on a baseline of 95% and young adults a 5% difference on a baseline of 90%. The larger difference between conditions in RTs or accuracy for older adults than young might be due to better information or it might be that they actually have worse information but their larger difference is due to their more conservative criteria.

To directly compare the quality of information for older and young adults requires a computational model that can solve the three problems just described. The model should separate information quality from the effects of speed/accuracy criteria and it should explain how accuracy and RTs arise from the same underlying cognitive processes. To accomplish this, we use Ratcliff's (1978; Ratcliff & McKoon, 2008) diffusion model, a member of the class of sequential sampling models. We have found that in some (although not all) memory and perceptual tasks, the quality of the information on which performance is based is as good for older as younger adults (e.g., RTM papers). The reason older adults are slower than young is usually due to more conservative speed/accuracy criteria (and also to slowdowns in processes outside those of interest, such as encoding a stimulus or executing whatever response is required in an experiment, e.g., pressing a key).

In most applications of the diffusion model, the number of observations in an experiment has been large, often several hundred per condition. Large numbers have been required to produce accurate estimates of the components of processing identified by the model. Also, large numbers reduce the variance in the components for a single participant's performance such that it is smaller than the variance between participants. Only with smaller variance within than between participants can individual participants be compared. However, experiments with textual materials usually cannot have large numbers of observations per condition (we call this the "small-n" problem, Ratcliff, 2008). One reason is that such materials can be difficult to construct. Another is that too many items of a particular kind might lead participants to adopt special strategies. Below, we describe a method for addressing this problem.

Before presenting the diffusion model and our method for addressing the small-n problem, we first situate predictive inferences in a broader context and review the two studies that have investigated predictive inferences for older adults.

What kinds of elaborative inferences do readers generate?

In the past, whether college students draw inferences that elaborate on the information that is explicitly stated in a text was the focus of much debate. Until the early 1980's, many researchers believed that readers generate and encode all the inferences from a text that would be needed to construct a "mental model" of the real-life situation described by the text (e.g. Anderson & Ortony, 1975; Anderson et al., 1976; Black & Bower, 1980; Bower, Black, & Turner, 1979; Bransford, Barclay, & Franks, 1972;

Bransford & Franks, 1971; Johnson, Bransford, & Solomon, 1973; Johnson-Laird, 1980; Mandler, 1978; Mandler & Johnson, 1977; Paris & Lindauer, 1976; Rumelhart, 1975, 1977; Schank & Abelson, 1977; Stein & Glenn, 1979; van Dijk & Kintsch, 1983). However, in the mid-1980's, McKoon and Ratcliff (McKoon & Ratcliff, 1986, 1988, 1989a, 1989b, 1989c; Seifert, McKoon, Abelson, & Ratcliff, 1986; see also Alba & Hasher, 1983) developed evidence that supports a different view, that the only types of inferences made automatically and passively during reading are minimal ones, not the full set of inferences that would be needed to form a mental model (McKoon & Ratcliff, 1992). One of the types of minimal inferences that McKoon and Ratcliff defined is predictive inferences. For the actress sentence, a full inference would be that the actress died. As characterized above, a minimal inference might be only that "something bad" happened to the actress.

McKoon and Ratcliff's (1992) minimalist proposal led to controversy. Some researchers continued to argue that readers do automatically construct a full mental model of the events described by a text (the term "mental models" has been replaced in much current research by the term "situation models"; e.g., Graesser, Singer, & Trabasso, 1994; Long, Seely, & Oppy, 1996; Madden & Dijkstra, 2010; Morrow et al., 1997; Radvansky & Copeland, 2006a, 2006b; Radvansky & Curiel, 1998; Radvansky & Dijkstra, 2007; Radvansky et al., 2001; Radvansky, Copeland, Berish, & Dijkstra, 2003; Radvansky, Copeland, & Zwaan, 2003; Singer, Graesser, & Trabasso, 1994; Stine-Morrow, Parisi, Morrow, Greene, & Park, 1997; Stine-Morrow et al., 2002, 2004; Suh & Trabasso, 1993; Zwaan, 2008; Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995; Zwaan & Radvansky, 1998). In this body of research, predictive inferences have not been explicitly studied. By at least some definitions, they would not qualify as inferences to be included in situation models. For example, Singer and Ferreira (1983) and Radvansky and Dijkstra (2007) have said that predictive inferences are "forward" inferences, that the event they express occurs after the time frame described by a text. If so, they would not be part of a text's situation model.

Other researchers have advanced the minimalist position into a theoretical framework called "memory-based" processing (e.g., Albrecht & Myers, 1995, 1998; Albrecht & O'Brien, 1993; Cook, Halleran, & O'Brien, 1998; Gerrig & McKoon, 1998, 2001; Gerrig & O'Brien, 2005; Greene, Gerrig, McKoon, & Ratcliff, 1994; McKoon, Gerrig, & Greene, 1996; McKoon & Ratcliff, 1995; Myers, Cook, Kambe, Mason, & O'Brien, 2000; Myers & O'Brien, 1998; Myers, O'Brien, Albrecht, & Mason, 1994; O'Brien, Albrecht, Hakala, & Rizzella, 1995; O'Brien, Raney, Albrecht, & Rayner, 1997; O'Brien, Rizzella, Albrecht, & Halleran, 1998; Rizzella & O'Brien, 2002; van den Broek et al., 2005). In this view, as Gerrig and O'Brien (2005, p. 229) put it, incoming text "serves as a signal to all of long-term memory, including both the inactive portion of the discourse representation [for the text currently being read] as well as general world knowledge. The signal proceeds autonomously and is unrestricted. Memory elements that are contacted by the initial signal in turn signal to other elements. During this resonance process, activation builds and when the process stabilizes, the most active elements" become part of the reader's understanding of the text. The conclusion from this research has been that predictive inferences are automatically, passively, and minimally encoded during reading (e.g., Calvo & Castillo, 1996; Cook, Limber, & O'Brien, 2001; Fincher-Kiefer, 1993, 1995; Gueraud, Tapiero, & O'Brien, 2008; Keefe & McDaniel, 1993; Klin, Guzman, et al., 1999; McKoon & Ratcliff, 1986; Murray et al., 1993; Peracchi & O'Brien, 2004; Till, Mross, & Kintsch, 1988; Weingartner et al., 2003; see Gerrig & O'Brien, 2005, for a review).

How does aging affect inference generation?

The situation-model framework has guided many investigations of the effects of aging on inference (e.g., Madden & Dijkstra, 2010; Morrow et al., 1997; Radvansky, 1999; Radvansky, Copeland, Berish, et al., 2003; Radvansky, Copeland, & Zwaan, 2003; Stine-Morrow et al., 2002, 2004). Radvansky and Dijkstra (2007) reviewed these studies and concluded that comprehension at the situation-model level is well preserved. However, most of these studies were conducted before the difficulties of comparing performance between young and old, and the necessity to explain RTs and accuracy simultaneously, were well understood. When differences between old and young were observed, it was usually concluded that they were due to differences in the quality of the information available to old versus young. When differences were not observed, it was usually concluded that there were no differences in quality. Moreover, comparisons were often made in the context of different baselines. The problems presented by differing baselines were illustrated above for a lexical decision task, but they apply to any task. For example, reading times for sentences in an experiment might average 1500 ms for older adults and 1200 ms for young adults, but the difference in reading times between two conditions, e.g., 150 ms, might be the same. This difference cannot be interpreted because, it is reasonable to think, older adults set different criteria for reading, perhaps giving more, or less, weight to a full understanding of a sentence than young adults.

A predictive inference study by Zipin et al. (2000) illustrates how interpretations based on accuracy data and interpretations based on RT data can conflict with each other. Zipin et al. asked participants to respond to test words presented either after predictive sentences or after control sentences. The task was to respond "yes" or "no" according to whether a test word had or had not appeared in the sentence. For example, the test word "cut" was presented either after *While swimming in the shallow water near the rocks, Sharon stepped on a piece of glass. The glass came from a bottle tossed carelessly on the rocks* or after *While walking by a large display near the entrance, Helen saw a beautiful piece of glass. It was a hand blown glass bottle made by a local artist.* The question was whether, in the predicting condition, older participants (ages 55–78) were as likely to infer "cut" as younger participants (ages 18–25).

The finding for RTs for correct responses (responding "no" to "cut") was that the older adults were slower in

the predicting condition than the control condition, by about 200 ms, but the younger adults were not; their RTs were about the same in the two conditions. However, for accuracy, both groups were more likely to respond "yes" in error in the predicting than the control conditions, by amounts that were not significantly different. Thus, the conclusion that might be drawn would be different for the two measures. For RTs, it might be concluded that in the predicting condition, older adults better understood Sharon being cut than did the young adults. For accuracy, it might be concluded that there was no difference in their understanding. There were also baseline differences. For RTs for the predicting and control conditions, the older adults averaged 1435 ms whereas the young averaged 1218 ms. For accuracy, the older adults' probability of an error averaged .24 whereas for the young adults, it averaged only .12.

Valencia-Laver and Light (2000; Expt. 3) also obtained equivocal results about age effects on predictive inferences. Like Zipin et al., they tested predictive inferences with older (ages 63–81) and younger (ages 18–30) adults. They used predicting sentences like *The angry swarm of bees flew out of the hive and landed on Joan's hand* (from McKoon & Ratcliff, 1986). The participants' task was perceptual identification. Several seconds after they read a sentence, a test word was displayed and then masked. The task was to name the word. The display time was set for each participant individually such that the probability of correct identification was about 50%. Comparing predicted to control test words (e.g., "sting" to "baked"), the young participants were 22% more likely to identify predicted than control words, whereas the older participants were only 4% more likely. Although this difference was not significant in Valencia-Laver and Light's study, it is large enough to suggest a difference between older and younger participants in their likelihood of drawing predictive inferences. However, the difference cannot be interpreted because RTs were not measured.

The diffusion model

In any experiment, the response to a stimulus is both a choice that a participant makes and the time taken to make that choice and respond. In lexical decision, for example, a participant chooses whether to respond "word" or "nonword" and takes some amount of time to do so. In reading sentences, a participant chooses some level of comprehension and takes some time to reach that level.

For two-choice tasks for which mean RTs are short (e.g., less than about 1.5 s), the diffusion model provides the mapping between accuracy and RT that is needed. In the model, evidence about a stimulus is accumulated over time from a starting point (z) to one or the other of two criterial amounts, or boundaries, one for each choice. The higher the quality of the evidence, the higher the rate at which it is accumulated. The rate of accumulation is called drift rate, v. Stimuli that differ in difficulty (e.g., in lexical decision, low-frequency versus high-frequency words) differ in drift rates. A response is executed when the amount of accumulated evidence reaches a criterion, either 0 for a negative response or a for a positive response. The

processes outside the decision process (e.g., encoding, memory access, and response execution) are combined into a single parameter of the model that has mean duration T_{er} ms. Noise (within-trial variability) in the accumulation of evidence from the starting point to the criteria results in processes with the same mean drift rate terminating at different times (producing RT distributions) and sometimes terminating at the wrong criterion (producing errors).

The values of drift rates, the nondecision component, and the starting point are assumed to vary from trial to trial. The assumption of across-trial variability is required if participants cannot accurately set these parameters at the same values from trial to trial (e.g., Laming, 1968; Ratcliff, 1978). Across-trial variability in drift rate is assumed to be normally distributed with SD η, across-trial variability in the nondecision component is assumed to be uniformly distributed with range s_t, and across-trial variability in the starting point is assumed to be uniformly distributed with range s_z. (These assumptions are not crucial because the model is robust to the forms of the distributions.) Across-trial variability in drift rate and starting point are necessary for the model to account for the relative speeds of correct and error RTs (Ratcliff, Van Zandt, & McKoon, 1999).

Performance usually includes "contaminant" responses – responses that are spurious in that they do not come from the decision process of interest (e.g., distraction, lack of attention). To accommodate these responses, on some proportion of trials (p_o), a random delay is added to the decision RT. The across-trial variability in p_o is uniform between the maximum and minimum RTs for each experimental condition. (Like the other variability parameters, the model is robust to the form of the distribution.) We do not consider this parameter further in this article because it was very small in the fits of the model to the data, .004 for college students, .002 for 60–74-year-olds, and .004 for 75–90-year-olds.

The diffusion model is designed to explain all aspects of data – accuracy, mean correct and mean error RTs, the shapes and locations of RT distributions, and the relative speeds of correct and error responses – and it is designed to do so at the level of individual subjects. With a single 45-min experimental session, the model can successfully fit data for individual participants, with standard deviations in the parameter estimates for criteria, nondecision time, and drift rate typically 3–5 times smaller than the standard deviations across participants.

The model is tightly constrained. The most powerful constraint comes from the requirement that the model fit the right-skewed shape of RT distributions (Ratcliff, 1978, 2002; Ratcliff & McKoon, 2008; Ratcliff et al., 1999). In addition, across experimental conditions that vary in difficulty (and are randomly intermixed at test), changes in accuracy, RT distributions, and the relative speeds of correct and error responses must all be captured by changes in only one parameter of the model, drift rate. Across experimental conditions that vary in speed/accuracy criteria (e.g., speed versus accuracy instructions), all the changes in accuracy, RT distributions, and the relative speeds of correct and error responses are usually captured

by changes only in the settings of the response criteria. The criteria cannot be adjusted as a function of difficulty because it would be necessary for the system to know which level of difficulty was being tested before drift rate could be determined. It is also usually assumed that the processes that make up the nondecision component of the model do not vary with difficulty.

The diffusion model has been applied to a range of experimental tasks with younger and older adults as participants (RTM). For example, item recognition data show large increases in RTs with age coupled with small changes in accuracy or no changes in accuracy at all. The RT data suggest large decrements in the match between a test item and memory with age whereas the accuracy data suggest only small decrements. As just outlined, the diffusion model reconciles these seemingly inconsistent results by mapping the two dependent variables onto the same underlying decision process. We have found (RTM) in many but not all tasks, that the large increases in RTs with age are due mainly to increases in criteria settings and the duration of the nondecision processes. The small or nonexistent deficits in accuracy are due to small or nonexistent decreases in drift rates. From these findings (RTM, 2004, 2010, 2011) we have concluded that, for item recognition, drift rates change little with age. In the experiment described below, we used item recognition as a paradigm to compare older and younger adults' comprehension of predictive inferences.

Examining age effects on predictive inferences by applying the diffusion model

To test for predictive inferences of the actress kind, McKoon and Ratcliff (1986; Experiment 3) used a recognition memory paradigm and compared predicting sentences like *The director and cameraman were ready to shoot close-ups when suddenly the actress fell from the 14th story* to control sentences that used many of the same words but did not lead to the predictive inference. For the actress sentence, the control was *The director fell upon the cameraman, demanding that he get close-ups of the actress on the 14th story.*

In McKoon and Ratcliff's experiment, college students were presented with lists of several unrelated sentences to read and then an immediately following test list of single words. For each test word, participants were to respond "yes" or "no" according to whether it had appeared in any of the sentences they had just read. The target test words for the control and predicting sentences were the same (e.g., "dead" for the actress sentences). Because a target test word did not appear in any sentence, the correct response was "no." Responding "no" should be more difficult in the predicting than the control conditions, to the extent that an inference was at least partially encoded for the predicting sentence.

McKoon and Ratcliff found this result but only when the critical target test word was immediately preceded in the test list by a word from its sentence. For the "dead" example, "dead" was preceded by "actress" (in both the predicting and control conditions) or "dead" was preceded by a word from another sentence (the same word for the predicting and control conditions). Responses to "dead" for the predicting sentence were slower and less accurate than for the control sentence only when "dead" was immediately preceded by "actress," not when it was preceded by the word from another sentence. From this, McKoon and Ratcliff concluded that predictive inferences are encoded strongly enough to affect responses when tested in the context of their sentences, but not strongly enough to affect responses outside the contexts of their sentences. It is this result that McKoon and Ratcliff interpreted as minimal encoding.

It should be noted that predictive inferences have sometimes been thought to be transient (e.g., Allbritton, 2004; Campion, 2004; Fincher-Kiefer, 1996; Keefe & McDaniel, 1993; Valencia-Laver & Light, 2000), perhaps understood on-line as a sentence is read but not encoded into the representation of the sentence in memory. However, McKoon and Ratcliff's result shows that at least some information about a predictive inference is available well after it has left working memory (see also, e.g., Klin, Guzman, et al., 1999; Klin, Murray, Levine, & Guzman, 1999).

McKoon and Ratcliff (1986) used the recognition memory paradigm just described in order to ensure that two possible mechanisms for responses to target test words could be ruled out. First, if a test word were presented immediately after a sentence was read (e.g., "dead" immediately after the actress sentence), it could be that the predictive inference was constructed by some mechanism that worked backwards to attempt to integrate the test word with the just-read sentence (Forster, 1981). This cannot occur if the test word and the sentence are not in short-term memory at the same time. Second, tests of recognition for single words generate fast responses, under about 1000 ms. These responses are faster than could be generated from strategic processes that attempt to construct the inference at the time of the memory test (e.g., Ratcliff & McKoon, 1981) and they are fast even relative to the availability of relational information (Dosher, 1984; Ratcliff & McKoon, 1982, 1989).

In the experiment described below, we used the same paradigm as McKoon and Ratcliff (1986). To illustrate application of the diffusion model to single-word recognition for young and older adults, we used the model to generate predicted data (RTs and accuracy). These data, summarized in Table 1, show the effects that changes in speed/accuracy criteria can have on RTs and accuracy.

RTs and accuracy were generated for two age groups, young and old, and two conditions: test words for predicting sentences and test words for control sentences (for both cases, we assumed that the test words were immediately preceded in the test list by the same word from their sentence, e.g., "actress"). The drift rates were set to be the same for the young and old adults, 0.3 for predicting sentences and 0.2 for control sentences. The two groups of participants differed in the distance between their criteria: 0.15 for the young participants and 0.25 for the old participants (a distance between the criteria of 0.08 is also included in the table for reference). Despite the fact that drift rates were the same for the two groups, the predicting/control differences were different, 38 ms and .095 probability correct for the young group, and 77 ms and

Table 1
Predictions from the diffusion model for primed and unprimed words as a function of boundary separation.

Boundary separation (a)	Primed mean RT (ms)	Unprimed mean RT (ms)	Priming effect (ms)	Primed accuracy	Unprimed accuracy
0.08	596	605	9	.795	.710
0.15	751	789	38	.858	.763
0.25	948	1025	77	.878	.775

The parameter values for the predictions were: starting point $z = a/2$, mean nondecision component of response time $T_{er} = 500$ ms, SD in drift across trials $\eta = 0.25$, range of the distribution of starting point $s_z = 0.05$, proportion of contaminants $p_o = 0.002$, range of the distribution of nondecision times $s_t = 200$ ms, unprimed drift rate $v_1 = 0.2$, and primed drift rate $v_2 = 0.3$.

.103 for the old group (Table 1). From these values, especially the RTs, it might be concluded that older participants make a stronger inference than younger participants (perhaps the older participants make more effort to comprehend the sentences than do the younger participants). However, this would be a mistake: the difference in drift rates between test items for predicting and control sentences was the same for the older adults as the younger adults.

Experiment

The materials were pairs of sentences, each made up of a predicting sentence and its control. The actress sentences above are one example; the sentences below, with the target word "lift," are another. In test lists, the test word immediately preceding the target for the predicting and control conditions, the "prime" for "dead," was the same – the character who, for the predicting sentence, would engage in the predicted event. "Lift," for example, was immediately preceded by "mover" in both the predicting and control conditions.

For each pair, we attempted to ensure, as much as possible, that the control sentence would not lead to the predictive inference. As the data presented later indicate, we did not completely succeed in this. For the lift sentence, for example, lift might be associated to the control sentence even though it is not a predicted event. Nevertheless, the data from the experiment show that the inferences generated from the predicting sentences were stronger than the associations between target words and control sentences.

Predicting: The mover bent his knees, put his arms around the box, and took a deep breath.
Control: The mover laid down the box, rested his knees and arms, and caught his breath.

To conduct this experiment, we needed to address the problem described at the beginning of this article, the small-n problem. As we pointed out, an application of the diffusion model usually requires that there be many test items per condition in order for the model to accurately estimate the values of the components of processing that best predict the data. In the experiment here, the number of sentences per condition was only 16 (16 predicting sentences and 16 control sentences).

Ratcliff and colleagues (McKoon & Ratcliff, 2012; Ratcliff, 2008; White, Ratcliff, Vasey, & McKoon, 2009; White, Ratcliff, Vasey, & McKoon, 2010a, 2010b) showed that the small-n problem can be addressed by including large num-

bers of filler items in an experiment. The diffusion model is fit simultaneously to all conditions, experimental and filler, and so the conditions with large n's largely determine the values of the response criteria, nondecision, and across-trial variability parameters. Since these parameters cannot vary across conditions, drift rates for the test words in the predicting and control conditions are determined by their RT and accuracy data alone. In the experiment here, there were 416 filler test words that had appeared in studied sentences and 384 words that had not.

The estimates of drift rates obtained in this way have less variability than would be obtained if the other parameters were not fixed across conditions and less variability than in the raw accuracy and RT values. White et al. (2009, 2010a, 2010b) demonstrated this with studies of anxiety. The question was whether individuals with high trait anxiety could be differentiated from those with low trait anxiety in terms of their responses to what are called threat words (e.g., "anger", "fear"). Testing such words in lexical decision, previous studies had failed to find consistent effects on either RTs or accuracy. However, when the diffusion model was applied, there was a significant effect in drift rates: anxious individuals had larger drift rates for threat words compared to neutral words, whereas non-anxious individuals did not. The power for detecting the drift-rate difference was about double the power for RTs or accuracy.

Method

Materials

We used 32 pairs of sentences, each pair made up of a predicting sentence and a control sentence, and each pair associated with its "target" test word, i.e., the word that expressed the predicted event (e.g., "dead"). There were three additional test words from each sentence, the same three for both sentences of a pair. One of them was the individual who would engage in the predicted event, i.e., the prime for the target word (e.g., "actress," "mover"). The other two were words (e.g., "cameraman," "director," "arms," "box") chosen randomly from the other content words of the sentence. For the predicting sentences, the mean number of words per sentence was 18, and for the control sentences the mean was 20. For the predicting sentences, there were 18 that were two lines long (as displayed on the PC monitor in the experiment) and 14 that were three lines. For the control sentences, there were 13 two-line sentences and 17 three-line sentences.

There was also a pool of 67 filler sentences of approximately the same length (mean number of words 15) as

the predicting and control sentences. There were five test words from each of these sentences. There were 1 one-line sentence, 52 two-line sentences, and 17 three-line sentences.

Negative test words, words that did not appear in any sentence, included the target word for the experimental sentences (e.g., "dead"). The other negative words were drawn from a pool of 2710 words with frequencies ranging from 0 to 999 (Kucera & Francis, 1967).

Procedure

Stimuli were displayed on a PC screen and responses were collected from the PC's keyboard. For the words of the sentences and test words, individual letters were .4° wide and .8° high at a normal viewing distance of 57 cm.

The experiment began with 30 lexical decision test items, used for practice with the PC keyboard. Then there were 32 blocks of sentences and test words (preceded by 1 practice block which contained only filler sentences). For each of the 32 blocks, there were three sentences for study, two filler sentences and one experimental sentence (i.e., the predicting or control sentence from one of the pairs). The experimental sentence was always the second of the three sentences. Participants began each block by pressing the space bar on the keyboard. Then the study sentences were displayed one at a time, for 6 s for one- and two-line sentences and 8 s for three-line sentences. There was a 500 ms pause between sentences.

The three sentences were followed by 26 test words, 13 from the three sentences just studied (one of them was the prime for the target word) and 13 that did not appear in any of the sentences in the experiment (one of which was the target word). For each word, participants were asked to respond "yes" or "no" as quickly and accurately as possible according to whether the word had appeared in any of the just-studied sentences, using the ?/key on the PC keyboard for "yes" responses and the Z key for "no" responses. Each test word was displayed until the participant made a response and then the screen was cleared. If the response was correct, the next test word appeared in 300 ms. If the response was not correct, the word ERROR appeared for 900 ms and then the screen was cleared for 300 ms. The test words were presented in random order except that the target test word for a predicting or control sentence was immediately preceded by its prime and the three test words immediately preceding the prime could not be words from the experimental sentence.

The pairs of experimental sentences were counterbalanced across sentences and participants such that half of the sentences for each participant and half of the participants for each sentence were in the predicting condition and the other half in the control condition.

Participants

There were 30 college-age participants, 37 participants ages 60–74, and 30 participants ages 75–90. The college-age participants were recruited at Ohio State University and Columbus, OH, area community recreation centers. The older adults were community-dwelling volunteers from the Columbus area, recruited at senior citizens centers and senior communities. Each participated in two sessions, one to collect the background information listed in Table 2 and the other to participate in the experiment. The college-age participants were paid $12 per session and the older adults were paid $15 per session (they were paid more because they had to travel to a senior center to participate). As shown in Table 2, the participants in the three age groups were matched on a number of background measures, in particular, IQ, the Mini-mental State Examination, and the Center for Epidemiological Studies depression scale.

Results

RTs longer than 3500 ms and shorter than 300 ms were eliminated from the analyses (less than 1% of the data). Mean RTs and accuracy values are shown in Table 3.

For the filler positive and negative words, the data showed the pattern typical in item recognition for the older and younger participants. Differences in accuracy between the three groups were small ($F(2,94) < 1.0$, standard error, $SE = .007$), at most 6%, whereas differences in RTs were large ($F(2,94) = 45.5$, $SE = 14.0$, $p < .05$ for all the anovas reported here). For correct "yes" responses, the difference in RTs between the college-age and 75–90 year olds was 277 ms, and for correct "no" responses, the difference was 376 ms.

For the college-age participants, for the target test words, the probability of a "yes" response when the predicting sentence was read (i.e., the probability of an error) was .63; when the control sentence was read, it was .41. The probability of a "yes" response to the target in the control condition was higher than the probability of a "yes" response for other test words that had not appeared in any sentence. We attribute the high probability of errors for the target words in the control condition to overlap in meaning between the control sentences and the target words (e.g., overlap between "laid down box," "arms," knees" and "lift").

For the older participants, the probability of a "yes" response when the predicting sentence was read was .77 for the 60–74-year-olds and .79 for the 75–90-year-olds, and when the control sentence was read, it was .54 and .57. Although the older participants were biased to respond "yes" relative to the college-age participants, the difference in accuracy between the predicting and control conditions was almost identical ($F(2,94) < 1.0$, $SE = .016$) for the three groups: .22 for the college-age participants, .23 for the 60–74-year-olds, and .22 for the 75–90-year-olds.

In contrast, the difference in RTs between the predicting and control conditions increased substantially with age: 91 ms for the 60–74-year-olds and 106 ms for the 75–90-year-olds, compared to 45 ms for the college-age participants. However, the variability in RTs was large enough that this difference was not significant ($F(2,94) = 1.46$, $SE = 14.7$).

Diffusion model analyses

As discussed above, the model needed to account for the shapes and locations of the distributions of RTs for correct responses and errors, accuracy, and the relative speeds of correct and error responses, and it needed to do this for

Table 2
Participant characteristics.

Measure	College age		60–74 year olds		75–90 year olds	
	M	SD	M	SD	M	SD
Mean age	20.8	1.9	68.7	3.6	79.8	3.3
Years education	13.4	1.8	15.3	3.4	13.9	2.5
MMSE	29.2	0.8	27.7	4.9	28.0	1.5
WAIS-III vocabulary (scaled)	11.4	3.1	11.6	2.3	11.7	2.4
WAIS-III vocabulary (raw)	41.7	11.8	47.2	9.2	45.1	9.3
WAIS-III matrix reasoning (scaled)	12.0	3.2	11.0	3.3	13.1	3.2
WAIS-III matrix reasoning (raw)	19.2	4.7	13.0	6.0	13.1	5.0
WAIS-III IQ	109.7	16.0	107.5	13.7	113.4	14.9
CES-D	10.1	7.5	7.1	5.4	7.7	4.2

Note: MMSE = Mini-Mental State Examination; WAIS-III = Wechsler Adult Intelligence Scale-3rd edition; CES-D = Center for Epidemiological Studies-Depression Scale.

Table 3
Response proportions and median RTs for experimental and filler conditions.

Statistic	Participant group	Predicting ("old" responses)	Control ("old" responses)	Filler old ("old" responses)	Filler new ("new" responses)
Response proportions	College	0.63	0.41	.82	.83
	60–74	0.77	0.54	.88	.80
	75–90	0.79	0.57	.87	.78
Median RTs	College	703	748	684	737
	60–74	928	1019	878	1024
	75–90	1012	1118	961	1113

all three groups of participants and for all the conditions in the experiment.

To fit the model to the data, the RT distributions were represented by 5 quantiles, the .1, .3, .5 (the median), .7, and .9 quantiles. The model was fit with a chi-square minimization method, which was fully described by Ratcliff and Tuerlinckx (2002). Correct and error RT distributions were fit for all the conditions simultaneously and these were weighted by the number of observations (because the chi-square method uses frequencies). The model was fit to the data for each participant individually. Tables 4 and 5 show the means and SDs of the best-fitting parameter values averaged over participants.

We fit the model to eight categories of test words: target words in the predicting condition (16), target words in the control condition (16), words from filler sentences (320), words that did not appear in any sentence (384, not including the target word), words that immediately preceded the target words in the predicting condition (16), words that immediately preceded the target words in the control condition (16), filler words from the predicting sentences (32), and filler words from the control sentences (32). The prime words that immediately preceded the target words (e.g. "actress," "mover") were the same in the predicting and control conditions, but we treated them separately in fitting the model in order to increase the power of the data to constrain the model. Likewise, the filler words for the predicting and control sentences were the same, but we treated them separately for the same reason.

Table 4 shows the chi-square values from fitting the model to the data. The degrees of freedom were computed as follows: for the 5 quantile RTs for correct responses and the 5 quantile RTs for error responses, there were 6 bins (2 outside the .1 and .9 quantiles and 4 between the pairs of quantiles). This gives 12 degrees of freedom for each set of test words minus 1 (because the 12 numbers must sum to 1). With 11 degrees of freedom for each of the eight

Table 4
Diffusion model parameters.

Participant group and statistic	a	T_{er} (ms)	η	s_z	p_o	s_t (ms)	z	χ^2
College age, mean	0.149	0.519	0.246	0.074	0.004	0.169	0.066	95.9
60–74, mean	0.178	0.675	0.219	0.034	0.002	0.189	0.070	89.2
75–90, mean	0.187	0.718	0.193	0.031	0.002	0.223	0.074	97.8
College age, SD	0.029	0.049	0.051	0.044	0.008	0.069	0.018	35.5
60–74, SD	0.041	0.080	0.055	0.036	0.002	0.078	0.017	31.5
75–90, SD	0.036	0.072	0.058	0.040	0.005	0.086	0.018	46.4

The parameters were: boundary separation a, starting point $z = a/2$, mean nondecision component of response time, T_{er}, SD in drift across trials η, range of the distribution of starting point s_z, range of the distribution of nondecision times, s_t, and proportion of contaminants p_o.

Table 5
Diffusion model drift rates.

Participant group and statistic	Predicting	Control	Word from filler sentence	New word	Prime for predicting	Prime for control	Filler from predicting	Filler from control
College age, mean	0.105	−0.064	0.239	−0.275	0.301	0.262	0.247	0.238
60–74, mean	0.167	−0.001	0.243	−0.231	0.312	0.293	0.312	0.263
75–90, mean	0.180	0.017	0.214	−0.191	0.280	0.243	0.275	0.232
College age, SD	0.135	0.138	0.090	0.104	0.181	0.127	0.134	0.105
60–74, SD	0.108	0.106	0.063	0.103	0.126	0.092	0.098	0.104
75–90, SD	0.136	0.125	0.072	0.092	0.137	0.116	0.113	0.119

sets of test words, there were a total of 88 degrees of freedom in the data. There were 15 model parameters (shown in Tables 4 and 5) and so the number of degrees of freedom was 88 − 15 = 73.

The model fit the data well. The chi-square test is a very conservative test so, even when chi-square values are only a little lower than the critical value, the fit of the model to data is good (see Ratcliff et al., 2010, for a discussion of model fitting and chi-square values). In fact, of the chi-square values for individual participants, only 12 for college-age, 11 for 60–74 year-olds, and 9 for 75–90-year-olds were significant (out of 30, 39, and 30 participants, respectively).

The best-fitting values for all the parameters of the diffusion model except drift rates are given in Table 4. As expected, the distance between the criteria was larger for the older participants and the time taken up by nondecision processes was longer ($F_{(2,94)} = 9.0$, $SE = .003$ and $F_{(2,94)} = 69.0$, $SE = .007$). These two factors explain the longer RTs for the older participants.

The best-fitting values of drift rates for the eight categories of test words are shown in Table 5. The difference in drift rates for the target test words (e.g., "dead") between the predicting and control conditions was remarkably consistent across the three age groups: .169, .168, and .163, for the college age, 60–74-year-olds, and 75–90-year-olds, respectively ($F_{(2,94)} < 1.0$, $SE = .014$). This finding shows that the oldest participants understood the predictive inferences to the same degree as the young participants. The only difference among the age groups was that the older participants were biased toward "yes" responses.

For test words other than the target, memory declined only moderately with age. To test the differences statistically, we used the test words from the filler sentences and the words that had not appeared in any sentence, the categories with the largest numbers of items. Adding the absolute values for the two categories, the drift rates were .514 for the college-age participants, .474 for the 60–74-year-olds, and .405 for the 75–90-year-olds. This represents a relatively small decline in performance: from college-age to 60–74-year-olds, drift rates decreased by only 8%. From college-age to 75–90-year-olds, the decrease was larger, 21%, but 75–90-year-olds' drift rates still showed good discrimination. The effect of age was significant ($F_{(2,94)} = 3.9$, $SE = .015$).

The words used as primes for the target word were the same in the predicting and control conditions, but as noted above, the model was fit to their data separately. The difference in drift rates in the two cases was not significant ($F_{(2,94)} < 1.0$, $SE = .014$). The words used as fillers from the predicting and control sentences were also the same and the difference in drift rates was also not significant ($F_{(2,94)} = 1.31$, $SE = .010$).

The across-trial variability parameters of the model were significantly different across age groups (Table 4). Not surprisingly, the range of the nondecision component was larger for the older participants, $F_{(2,94)} = 3.8$, $SE = .008$, suggesting that either their encoding processes or their response execution processes or both were more variable than for the college-age participants. The SDs in drift rate across trials and the range of the starting point were significantly smaller for the older participants than the college-age ($F_{(2,94)} = 7.0$, $SE = .005$; $F_{(2,94)} = 10.9$, $SE = .004$). The direction of the difference in drift rate variability is important. If the SDs had been larger for the older participants, it might have been that their differences in discriminability between the predicting and control conditions were actually smaller than those of the college-age participants, with drift rates disguised by their larger variability.

Discussion

Predicting inferences are, by definition, implicit in a text, not explicit, and so might be thought to require more processing resources than explicitly-stated information. A number of hypotheses have attempted to explain why older adults' cognitive processing often shows decrements relative to young adults', including, for example, short-term memory deficits (e.g., Craik & Byrd, 1982; Hasher & Zacks, 1988), limitations on attentional resources (e.g., Craik, 1983), and decreases in processing speed (e.g., Salthouse, 1996). For our experiment, these hypotheses do not make definitive predictions. Short-term memory might be limited enough that the information necessary to form an inference is not all available at the same time and so no inference can be generated – or short-term capacity might not be limited to this extent. In the same way, a limit on attentional resources might or might not be sufficient to prevent inferences. How speed of processing applies to inference generation is somewhat of an open question because many studies, including studies of inference processing, have not taken speed/accuracy settings and baseline levels of performance into account (e.g., Valencia-Laver & Light, 2000; Zipin et al., 2000). For our experiment, none of these possible limits on older adults' resources applied: the strength with which predictive inferences were encoded and remembered was not significantly lower for

older adults than young adults. We believe that future research, with large numbers of participants, should be conducted to explore how differences among individuals in short-term memory and attentional resources interact with inference processing.

We concentrate the following discussion on the implications of our result for theories of discourse processing and how age affects discourse processing. It has been proposed that textual information is represented in four levels, for both encoding and memory. The most basic level is made up of the exact words of a text. The second is made up of the syntactic structures expressed by the concepts denoted by the words (e.g., Fodor, Bever, & Garrett, 1974; Frazier & Fodor, 1978), the third is made up of the propositions expressed by the concepts, and the fourth is a model of the whole situation described by the text. We first consider the effects of age on memory for exact words.

Aging and memory for words

Drift rates for the words from the filler sentences and the words that did not appear in any sentence showed good discrimination for all the age groups. They declined by only 8% from college-age to 60–74-year-olds and only 21% from college-age to 75–90-year-olds. This finding, especially for the 60–74-year-olds, calls into question the results of previous studies that have been interpreted as showing that older adults have considerably poorer memory for the words of texts than college-age adults (e.g., Radvansky, Copeland, Berish, et al., 2003, and other studies referenced therein). However, these studies did not have available a method for separating the strength with which items are represented in memory from other components of decision processes.

Drift rates about as good for 60–74-year-olds as college-age participants and only moderate declines for 75–90-year-olds are consistently obtained in recognition memory for words. In a study by Ratcliff et al. (2011), participants were given lists of single words, unrelated to each other, 12 words per list. The drop in drift rate from college-age to 60–74-year-olds was only 3%, and only 19% for 75–90-year-olds. In four similar studies (Ratcliff et al., 2004; Ratcliff et al., 2006a, 2007, 2010), drift rates were not significantly different for 60–74-year-olds than college-age participants and declined only moderately for 75–90-year-olds. In another study (McKoon & Ratcliff, 2012), participants were given pairs of words to study, with the words of a pair unrelated to each other or to any other words in the list, 12 pairs per list. Drift rates for college-age and 60–74-year-olds differed by −3%. The difference between college-age and 75–90-year-olds was larger, 31%, but again still shows good discrimination.

For these reasons, we conclude that the question of whether older adults have poorer memory for the words of texts than young adults should be re-investigated. What is needed are systematic studies that compare older' and younger' adults memory (for a range of ages for the older adults) for all the sorts of words in texts: words more or less relevant to the text's topic, higher and lower frequency words, syntactically more prominent and less prominent words, words expressing concepts referenced several times and only once, and so on.

Aging and propositions

Propositions represent the individual units of meaning in a text (e.g., Kintsch & Keenan, 1973). A proposition is made up of a relation and its arguments. For *The director and cameraman were ready to shoot close-ups*, there are three propositions: (1) "and, director, cameraman;" (2) "shoot, 1, close-ups;" and (3) "ready to, 1, 2." For the first, "and" is the relation that joins the cameraman and the director. For the second, "shoot" is the relation that joins the conjunction of director and cameraman to the close-ups that they are going to shoot. For the third, "ready to" relates the conjunction of director and cameraman to the shooting activity.

The propositional representation of a text (sometimes called the "textbase," e.g., Radvansky & Dijkstra, 2007) includes all the concepts that are explicitly stated (e.g., cameraman, director, shoot, ready to), the propositions that relate them to each other (e.g., "and, director, cameraman"), and the connections among the propositions (e.g., "ready to, 1, 2" connects proposition 1 to proposition 2). The representation also includes information that is easily available from long-term memory, including memory for the earlier parts of a text that are no longer being processed and information from semantic memory that is related to the explicitly-stated concepts and propositions.

For predictive inferences, the question is whether they are automatically encoded as part of a text's propositional representation. Kintsch's (1988) construction–integration model provides a mechanism by which this could be accomplished. In the model, propositions, the concepts in them, and easily-available information from memory are integrated during reading through a repeated recycling of activation such that information associated only weakly to a relatively small portion of a text is further weakened, and information associated more strongly to multiple concepts in the text is strengthened. For example, for *The townspeople were amazed to find that all the buildings had collapsed except the mint* (Till et al., 1988), the concept "earthquake" would not be strongly associated with any individual concept in the sentence, and so not immediately available to support inferences (see, e.g., Casteel, 1998, 2007; Cook, Limber, & O'Brien, 2001; Estevez & Calvo, 2000; Till et al., 1988). However, as integration proceeded, the recycling of activation from multiple sources would make "earthquake" available as an inference. This same integration process could produce the inference that something bad happened to the actress.

According to this analysis, predictive inferences would be part of the propositional representation of a text. If so, then our finding that the inferences are understood and remembered to the same degree for older adults as young adults contradicts conclusions from previous studies that older adults' memory for propositions is poorer (e.g., Cohen, 1979; Engelkamp & Zimmer, 1997; Kemper, 1987; Kemtes & Kemper, 1997; Light & Capps, 1986; Meyer & Rice, 1981; Stine & Wingfield, 1988, 1990; Stine-Morrow, Ryan, & Leonard, 2000; Zurif, Swinney, Prather, Wingfield,

& Brownell, 1995). However, again, these studies compared performance only on empirical measures, RTs and accuracy, not the individual components of processing that give rise to those measures.

Predictive inferences and situation models

Another possibility is that predictive inferences are represented in situation models, and older adults' memory for information from situation models is as good as young adults'. However, this idea is questionable because there are problems with the definition of situation model and with the evidence that has been said to support the hypothesis that situation models are constructed automatically and passively during reading. These problems, reviewed in the next paragraphs, strongly suggest that it is not useful to attribute predictive inferences to situation models.

Deciding what information should be included in a situation model and what should not be is difficult. Glenberg, Meyer, and Lindem (1987, p. 69) attempted a full definition: "A situation model is the result of interactions between information given in a text and knowledge about linguistics, pragmatics, and the real world; a situation model can be modified as new information comes in to produce a completely new interpretation of the text; the information in a situation model can be manipulated to produce emergent relations; a situation model is perceptual-like; a situation model guides interpretation of referential terms; and a situation model guides the generation of inferences."

Definitions like this may be intuitively compelling but they are not easy to implement. As Morrow, Bower, and Greenspan (1989, p. 300) described it, comprehension of *We flew from Paris to New York last week* is unlikely to involve information about the Atlantic ocean despite its being an essential part of the event described by the sentence. Many years ago, Alba and Hasher (1983, p. 25) concluded that the basic problem confronting mental models is their definition. This is still the case in current research. Relevant to our experiment, there is no way to independently verify whether predictive inferences should or should not be considered part of a situation model.

Among the studies that have been prominently claimed to support the automatic construction of situation models during reading is one by Schmalhofer and Glavanov (1986). In their paradigm, which is still sometimes used, participants were given texts to read followed by test sentences for which they were to decide if the sentence was exactly the same as one they have read. Test sentences were exactly the same as sentences that were read, paraphrases of sentences that were read, sentences consistent with the meaning of a text, or sentences inconsistent with the meaning. Participants were more likely to false alarm to paraphrases than consistent test sentences, which was thought to indicate a propositional level of representation, and more likely to false alarm to consistent than inconsistent test sentences, which was thought to indicate a situation-model level.

However, Schmalhofer and Glavanov (1986) gave two interpretations of their data that do not require situation-

model representations. They said: "One may argue that the difference between propositional and situational information, which were distinguished from one another by their contents, is just that and does not have any further implication with respect to their representation in memory. Thus only a single (propositional) memory representation would be postulated" (p. 292). By this interpretation of the data, meaning-consistent test sentences are simply easier than paraphrase test sentences. The other interpretation Schmalhofer and Glavanov proposed, following Reder (e.g., 1982), was that there are two retrieval processes that operate on a single, propositional, level of representation, one searching for an exact match between the test sentence and memory (a less accurate process) and the other judging the plausibility of the test sentence with respect to memory (a more accurate process).

Another study that has been influential was conducted by Glenberg et al. (1987). Participants read short texts like *A girl was enjoying the warm spring weather. She walked up to the entrance of a park, and bent down to an ornamental display to pick a flower for her sister. Then she walked into the park and down to a small stream where some ducks were feeding. She smiled to see seven tiny ducklings trailing behind their mother.* At the end of a text, a word was presented for recognition. For the girl-in-the-park text, the test word was "flower." If participants constructed a situation model as they read, then at the end of the text, the girl is holding the flower and so RTs to say "yes" to "flower" should be shorter than in a control condition where the girl only smelled the flower and did not take it with her, and this is what Glenberg et al. found. However, McKoon and Ratcliff (1992) showed that RTs for "display" were also shorter in the picking-the-flower version than the smelling version. Since the girl could not have the display with her at the end of the story, McKoon and Ratcliff concluded that something other than information in a situation model was speeding responses to "flower" for the picking compared to smelling versions (they suggested that propositions about picking flowers for a sister might be more salient in the text than propositions about smelling flowers).

One of the earliest studies to suggest that readers encode situation models was Bransford et al.'s (1972) "turtles on a log" experiment. Participants heard sentences like *Three turtles rested on a floating log and a fish swam beneath them*. After hearing the sentences, the participants were given test sentences and asked to decide, for each test sentence, whether it was the same as a sentence they had heard. When a test sentence was consistent with the heard sentence but not identical to it (*Three turtles rested on a floating log, and a fish swam beneath it*), participants often responded positively, in error. This result was usually attributed to participants' construction of a situation model for the heard sentence that did not include exact information about its wording. However, Jahn (2004) showed that participants were not automatically constructing situation models. In the turtles sentence, the turtles are not in danger from the fish. Others of Bransford et al.'s items did describe danger, as in *Two robins crouched (on/beside) their nest as the hawk flew above (it/them)*. Jahn tested danger and non-danger items with the same paradigm as

Bransford et al. and found that only the danger items showed Bransford et al.'s result. For non-danger items, participants were well able to discriminate consistent sentences from identical sentences, showing that they might not have been constructing situation models.

Situation models and aging

If hypotheses about situation models are on somewhat shaky ground because of the problems just discussed, then so also are hypotheses about the effects of age on the construction of them. Nevertheless, there have been many studies that have investigated aging and situation models. The results are mixed: some appear to support the preservation with age of situation-model encoding and memory and some do not.

There is a long list of supporting studies (e.g., Gilinsky & Judd, 1994; Miller & Stine-Morrow, 1998; Morrow, Leirer, & Altieri, 1992; Morrow et al., 1997; Radvansky, Copeland, Berish, et al., 2003; Radvansky, Copeland, & Zwaan, 2003; Radvansky & Curiel, 1998; Radvansky, Gerard, Zacks, & Hasher, 1990; Radvansky, Zacks, & Hasher, 1996; Radvansky et al., 2001; Shake, Noh, & Stine-Morrow, 2009; Soederberg & Stine, 1995; Stine-Morrow, Loveless, & Soederberg, 1996; Stine-Morrow et al., 2002, 2004). Findings include: older adults slow in their reading times for causally important elements of a text compared to unimportant elements (Stine-Morrow et al., 2004); they read slower about characters described in a text as interacting with each other than characters described as not interacting (Radvansky, Copeland, & Zwaan, 2003); and they read slower when a temporal shift from one scene to the next in a text is large compared to when it is small (Radvansky et al., 2001). Also, when probed after reading, they are slower to recognize a test word from an earlier than a more recent scene (Radvansky, Copeland, Berish, et al., 2003) and slower for tests of goals a character has already completed than goals not yet completed (Radvansky & Curiel, 1998). In addition, older adults show patterns of data similar to young adults in the Glenberg et al. (1987) paradigm (Radvansky, Copeland, Berish, et al., 2003) and the Schmalhofer and Glavanov (1986) paradigm (Radvansky, Copeland, Berish, et al., 2003; Radvansky, Copeland, & Zwaan, 2003; Radvansky et al., 2001). They also appear to understand which characters in a text are the most important (Morrow et al., 1992).

The many studies that do not support preservation include those by Noh and Stine-Morrow (2009), Light and Capps (1986), Dixon et al. (1982), Stine, Cheung, and Henderson (1995), and Hamm and Hasher (1992). Compared to young adults, older readers have difficulty accessing an earlier character in a text after a new character has been introduced (Noh & Stine-Morrow, 2009), identifying the correct referent of a pronoun when one sentence intervenes between pronoun and referent (Light & Capps, 1986), and matching one concept in a sentence to a related concept in the previous sentence (Zelinski & Miura, 1990). Other studies suggest difficulties for older adults in organizing the elements of a text with respect to their importance (Dixon et al., 1982), differentially allotting time to comprehension of concepts new to a text (Stine, Cheung,

& Henderson, 1995), and appropriately discarding irrelevant information as they read (Hamm & Hasher, 1992). These results are all consistent with a failure on the part of older adults to construct, while they read, a situation model complete enough or strong enough to allow appropriate connections between text elements across sentences. Beyond relations among the entities of a text, older adults sometimes appear to have difficulties with inferences. Till and Walsh (1980) gave an inference word as a cue for recall (analogous to giving "dead" as a cue for recall of the actress sentence) and found that, relative to free recall, older adults benefited from the cue less than young adults. In studies by Light, Zelinski, and Moore (1982), Cohen (1979), Zacks et al. (1987), and Till (1985), older adults did not appear to generate inferences to the extent that young adults did.

Although this is only a brief review, the studies are representative: there is no decisive evidence to tell us whether older adults do or do not construct a full model of the events described by a text.

Conclusion

The question of whether predictive inferences should be considered part of the propositional representation of a text or the situation-model level cannot at this time be resolved. To assign predictive inferences to a situation-model level would require that there be a clear definition of what kinds of information are included in a situation model and what kinds are not, something that we do not have.

Whether propositional or situation-model information, the older adults in our study appear to have encoded and remembered predictive inferences as well as the young adults. In other words, they do appear to understand that something bad happened to the actress. We can draw this conclusion because the diffusion model gives a way to separate the strength with which information is represented in memory from speed/accuracy criteria and nondecision components of processing. The model can be applied to examine encoding and memory for particular kinds of information such as predictive inferences without any commitment to where they fall in or between propositional and situation-model levels.

The model solves a scaling problem: The older readers' responses to test words were much slower than the young adults', roughly 200–400 ms slower, and the difference between their RTs to "dead" in the predicting and control conditions was larger. For the young adults, the difference was 53 ms, whereas for the two older groups, the difference averaged 99 ms. Applying the model, we found that RTs were longer and differences between conditions larger because the older adults set their criteria farther apart: they were less willing than the young adults to go so fast that they made errors that they could have avoided by going slower.

The model also handles another problem: at the same time that the difference in RTs between the predicting and control conditions was considerably larger for the older adults than the younger, the difference in accuracy was almost nonexistent (a .23 difference between predicting

and control for the older groups, a .22 difference for the young group). The model resolved this seeming contradiction in the same way it resolved the scaling issue: the older adults' difference in memory strength (drift rate) for the target test word between the predicting and control conditions was almost identical to the young adults', even though they set their criteria further apart.

To our knowledge, the experiment reported here is the first application of a sequential sampling model to investigations of language comprehension and memory for older adults. It is our hope that, in the near future, such models will allow investigations of the degree to which older adults understand and remember many other kinds of inferences, as well as other sorts of textual information.

Acknowledgment

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93*, 203–231.

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

Albrecht, J. E., & Myers, J. L. (1998). Accessing distant text information during reading: Effects of contextual cues. *Discourse Processes, 26*, 87–107.

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1061–1070.

Allbritton, D. (2004). Strategic production of predictive inferences during comprehension. *Discourse Processes, 38*, 309–322.

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles – A problem of polysemy. *Cognitive Psychology, 7*, 167–180.

Anderson, R. C., Pichert, J., Goetz, E., Schallert, D., Stevens, K., & Trollip, S. (1976). Instantiation of general terms. *Journal of Verbal Learning and Verbal Behavior, 15*, 667–679.

Beeman, M., Friedman, R. B., Grafman, J., Perez, E., Diamond, S., & Lindsay, M. B. (1994). Summation priming and coarse semantic coding in the right hemisphere. *Journal of Cognitive Neuroscience, 6*, 26–45.

Black, J. B., & Bower, G. H. (1980). Story understanding as problem solving. *Poetics, 9*, 223–250.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*, 177–220.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*, 193–209.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2*, 331–350.

Calvo, M. G., & Castillo, M. D. (1996). Predictive inferences occur on-line, but with delay: Convergence of naming and reading times. *Discourse Processes, 22*, 57–78.

Calvo, M. G., Castillo, M. D., & Estevez, A. (1999). On-line predictive inferences in reading: Processing time during vs. after the priming context. *Memory & Cognition, 27*, 834–843.

Campion, N. (2004). Predictive inferences are represented as hypothetical facts. *Journal of Memory and Language, 50*, 149–164.

Casteel, M. (1998, November). The role of working memory in the generation of predictive inferences. Poster session presented at the 39th annual meeting of the Psychonomics Society, Dallas, TX.

Casteel, M. A. (2007). Contextual support and predictive inferences: What do readers generate and keep available for use? *Discourse Processes, 44*, 51–72.

Cohen, G. (1979). Language comprehension in old age. *Cognitive Psychology, 11*, 412–429.

Cohen, G. (1981). Inferential reasoning in old age. *Cognition, 9*, 59–72.

Cook, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available during reading? A memory-based view of text processing. *Discourse Processes, 26*, 109–129.

Cook, A. E., Limber, J. E., & O'Brien, E.-J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language, 44*, 220–234.

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

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and cognitive processes* (pp. 191–211). New York: Plenum.

Dixon, R. A., Simon, E. W., Nowak, C. A., & Hultsch, D. F. (1982). Text recall in adulthood as a function of level of information, input modality, and delay interval. *Journal of Gerontology, 37*, 358–364.

Dosher, B. A. (1984). Discriminating pre-experimental (semantic) from learned (episodic) associations: A speed–accuracy study. *Cognitive Psychology, 16*, 519–555.

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica, 96*, 43–60.

Estevez, A., & Calvo, M. G. (2000). Working memory capacity and time course of predictive inferences. *Memory, 8*, 51–61.

Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes, 16*, 99–124.

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.

Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes, 22*, 225–246.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. St. Louis: McGraw-Hill Book Co..

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology, 33A*, 465–495.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two stage parsing model. *Cognition, 6*, 291–325.

Gerrig, R., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. Invited article. *Discourse Processes, 26*, 67–86.

Gerrig, R., & McKoon, G. (2001). Memory processes and experiential continuity. *Psychological Science, 12*, 82–86.

Gerrig, R. J., & O'Brien, E. J. (2005). The scope of memory-based processing. *Special Issue: Discourse Processes, 39*, 225–242.

Gilinsky, A. S., & Judd, B. B. (1994). Working memory and bias in reasoning across the life span. *Psychology and Aging, 9*, 356–371.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26*, 69–83.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395.

Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and the management of common ground. *Journal of Memory and Language, 33*, 511–526.

Gueraud, S., Tapiero, I., & O'Brien, E. J. (2008). Situation models and the activation of predictive inferences. *Psychonomic Bulletin & Review, 15*, 351–356.

Hamm, V. P., & Hasher, L. (1992). Age and the availability of inferences. *Psychology and Aging, 7*, 56–64.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.). *The psychology of learning and motivation* (Vol. 22, pp. 193–225). New York, NY: Academic Press.

Hess, T. M. (1995). Aging and the impact of causal connections on text comprehension and memory. *Aging and Cognition, 2*, 310–325.

Jahn, G. (2004). Three turtles in danger. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 969–987.

Johnson, M. K., Bransford, J. D., & Solomon, S. K. (1973). Memory of tacit implications of sentences. *Journal of Experimental Psychology, 98*, 203–205.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science, 4*, 71–115.

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

Kemper, S. (1987). Syntactic complexity and the recall of prose by middle-aged and elderly adults. *Experimental Aging Research, 13*, 47–52.

Kemtes, K. A., & Kemper, S. (1997). Younger and older adults on-line processing of syntactic ambiguities. *Psychology and Aging, 12*, 362–371.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95*, 163–182.

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

Klin, C. M., Guzman, A. E., & Levine, W. H. (1999). Prevalence and persistence of predictive inferences. *Journal of Memory and Language, 40*, 593–604.

Klin, C., Murray, J. D., Levine, W. H., & Guzman, A. E. (1999). Forward inferences: From activation to long-term memory. *Discourse Processes, 27*, 241–260.

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Laming, D. R. J. (1968). *Information theory of choice reaction time*. New York: Wiley.

Lassonde, K. A., & O'Brien, E. J. (2009). Contextual specificity in the activation of predictive inferences. *Discourse Processes, 46*, 426–438.

Light, L. L., & Capps, J. L. (1986). Comprehension of pronouns in younger and older adults. *Developmental Psychology, 22*, 580–585.

Light, L. L., Zelinski, E. M., & Moore, M. (1982). Adult age differences in reasoning from new information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 435–447.

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes, 34*, 259–280.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology, 94*, 778–784.

Long, D. L., Seely, M. R., & Oppy, B. J. (1996). The availability of causal information during reading. *Discourse Processes, 22*, 145–170.

Madden, C., & Dijkstra, K. (2010). Contextual constraints in situation model construction: An investigation of age and reading span. *Aging, Neuropsychology, and Cognition, 17*, 19–34.

Mandler, J. M. (1978). A code in the node: The use of a story schema in retrieval. *Discourse Processes, 1*, 14–35.

Mandler, J. M., & Johnson, N. J. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9*, 111–151.

McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: Some consequences of memory based text processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 919–932.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 82–91.

McKoon, G., & Ratcliff, R. (1988). Contextually relevant aspects of meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 331–343.

McKoon, G., & Ratcliff, R. (1989a). Assessing the occurrence of elaborative inference with recognition: Compatibility checking vs. compound cue theory. *Journal of Memory and Language, 28*, 547–563.

McKoon, G., & Ratcliff, R. (1989b). Inferences about contextually-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1134–1146.

McKoon, G., & Ratcliff, R. (1989c). Semantic associations and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 326–338.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*, 440–466.

McKoon, G., & Ratcliff, R. (2012). Aging and IQ effects on associative recognition and priming in item recognition. *Journal of Memory and Language, 66*, 416–437.

McKoon, G., & Ratcliff, R. (1995). The minimalist hypothesis: Directions for research. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Meyer, B. J. F., & Rice, G. E. (1981). Information recalled from prose by young, middle, and old adult readers. *Experimental Aging Research, 7*, 253–268.

Miller, L. M. S., & Stine-Morrow, E. A. L. (1998). Aging and the effects of knowledge on on-line reading strategies. *Journal of Gerontology: Psychological Sciences, 53B*, P223–P233.

Morrow, D., Bower, G., & Greenspan, S. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language, 28*, 292–312.

Morrow, D. G., Leirer, V. O., & Altieri, P. A. (1992). Aging, expertise, and narrative processing. *Psychology and Aging, 7*, 376–388.

Morrow, D. G., Stine-Morrow, E. A. L., Leirer, V. O., Andrassy, J. M., & Kahn, J. (1997). The role of reader age and focus of attention in creating situation models from narratives. *Journal of Gerontology: Psychological Science, 52B*, P73–P80.

Murray, J. D., & Burke, K. A. (2003). Activation and encoding of predictive inferences: The role of reading skill. *Discourse Processes, 35*, 81–102.

Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. *Journal of Memory and Language, 32*, 464–473.

Myers, J. L., Cook, A. E., Kambe, G., Mason, R. A., & O'Brien, E. S. (2000). Semantic and episodic effects on bridging inferences. *Discourse Processes, 29*, 179–199.

Myers, J. L., & O'Brien, B. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*, 131–157.

Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 876–886.

Noh, S. R., & Stine-Morrow, E. A. L. (2009). Age differences in tracking characters during narrative comprehension. *Memory & Cognition, 37*, 769–778.

O'Brien, E. J., Albrecht, J. E., Hakala, C. M., & Rizzella, M. L. (1995). Activation and suppression of antecedents during reinstatement. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 626–634.

O'Brien, E. J., Raney, G. E., Albrecht, J. E., & Rayner, K. (1997). Processes involved in the resolution of explicit anaphors. *Discourse Processes, 23*, 1–24.

O'Brien, E., Rizzella, M., Albrecht, J., & Halleran, J. (1998). Updating a situation model. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 1200–1210.

Paris, S., & Lindauer, B. K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology, 8*, 217–227.

Peracchi, K. A., & O'Brien, E. J. (2004). Character profiles and the activation of predictive inferences. *Memory & Cognition, 32*, 1044–1052.

Radvansky, G. A. (1999). Memory retrieval and suppression: The inhibition of situation models. *Journal of Experimental Psychology: General, 128*, 198–206.

Radvansky, G. A., & Copeland, D. E. (2006a). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition, 34*, 1150–1156 (could not find Rad & Cope 2004).

Radvansky, G. A., & Copeland, D. E. (2006b). Situation models and retrieval interference: Pictures and words. *Memory, 14*, 614–623.

Radvansky, G. A., Copeland, D. E., Berish, D. E., & Dijkstra, K. (2003). Aging and situation model updating. *Aging, Neuropsychology and Cognition, 10*, 158–166.

Radvansky, G. A., Copeland, D. E., & Zwaan, R. A. (2003). Aging and functional spatial relations in comprehension and memory. *Psychology and Aging, 18*, 161–165.

Radvansky, G. A., & Curiel, J. M. (1998). Narrative comprehension and aging: The fate of completed goal information. *Psychology and Aging, 13*, 69–79.

Radvansky, G. A., & Dijkstra, G. (2007). Aging and situation model processing. *Psychonomic Bulletin & Review, 14*, 1027–1042.

Radvansky, G. A., Gerard, L. D., Zacks, R. T., & Hasher, L. (1990). Younger and older adults' use of mental models as representations for text materials. *Psychology and Aging, 5*, 209–214.

Radvansky, G. A., Zacks, R. T., & Hasher, L. (1996). Fact retrieval in younger and older adults: The role of mental models. *Psychology and Aging, 11*, 258–271.

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging, 16*, 145–160.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.

Ratcliff, R. (2002). A diffusion model account of reaction time and accuracy in a two choice brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review, 9*, 278–291.

Ratcliff, R. (2008). The EZ diffusion method: Too EZ? *Psychonomic Bulletin & Review, 15*, 1218–1228.

Ratcliff, R., & McKoon, G. (1981). Automatic and strategic priming in recognition. *Journal of Verbal Learning and Verbal Behavior, 20*, 204–215.

Ratcliff, R., & McKoon, G. (1982). Speed and accuracy in the processing of false statements about semantic information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 16–36.

Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology, 21*, 139–155.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*, 873–922.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging, 16*, 323–341.

Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics, 65*, 523–535.

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*, 408–424.

Ratcliff, R., Thapar, A., & McKoon, G. (2006a). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin & Review, 13*, 626–635.

Ratcliff, R., Thapar, A., & McKoon, G. (2006b). Aging, practice, and perceptual tasks: A diffusion model analysis. *Psychology and Aging, 21*, 353–371.

Ratcliff, R., Thapar, A., & McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75–90 years old. *Psychology and Aging, 22*, 56–66.

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology, 60*, 127–157.

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General, 140*, 46–487.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating the parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*, 438–481.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261–300.

Reder, L. M. (1982). Plausibility judgments vs. fact retrieval: Alternative strategies for sentence verification. *Psychological Review, 89*, 250–280.

Rizzella, M. L., & O'Brien, E. J. (2002). Retrieval of concepts in script-based texts and narratives: The influence of general world knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 780–790.

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press.

Rumelhart, D. E. (1977). Understanding and summarizing brief stories. In D. Laberge & J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, NJ: Erlbaum.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*, 403–428.

Sanford, A. J., & Garrod, S. C. (2005). Memory-based approaches and beyond. *Discourse Processes, 39*, 205–224.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language, 25*, 279–294.

Seifert, C. M., McKoon, G., Abelson, R. P., & Ratcliff, R. (1986). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 220–231.

Shake, M. C., Noh, S., & Stine-Morrow, E. A. L. (2009). Age differences in learning from text: Evidence for functionally distinct text processing systems. *Applied Cognitive Psychology, 23*, 561–578.

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

Singer, M., Graesser, A. C., & Trabasso, T. (1994). Minimal or global inference in comprehension. *Journal of Memory and Language, 33*, 421–441.

Soederberg, L., & Stine, E. A. L. (1995). Activation of emotion information in text among younger and older readers. *Journal of Adult Development, 2*, 23–36.

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 101–117.

Stein, N. L., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing*. Hillsdale, NJ: Erlbaum.

Stine, E. A. L. (1990). On-line processing of written text by younger and older adults. *Psychology and Aging, 5*, 68–78.

Stine, E. A. L., Cheung, H., & Henderson, D. T. (1995). Adult age differences in the on-line processing of new concepts in discourse. *Aging and Cognition, 2*, 1–18.

Stine, E. A. L., & Wingfield, A. (1988). Memorability functions as an indicator of qualitative age differences in text recall. *Psychology and Aging, 3*, 179–183.

Stine, E. A. L., & Wingfield, A. (1990). How much do working memory deficits contribute to age differences in discourse memory? *European Journal of Cognitive Psychology, 2*, 289–304.

Stine-Morrow, E. A. L., Gagne, D. D., Morrow, D. G., & DeWall, B. H. (2004). Age differences in rereading. *Memory & Cognition, 32*, 696–710.

Stine-Morrow, E. A. L., Loveless, M. K., & Soederberg, L. M. (1996). Resource allocation in on-line reading by younger and older adults. *Psychology and Aging, 11*, 475–486.

Stine-Morrow, E. A. L., Morrow, D. G., & Leno, R. (2002). Aging and the representation of spatial situations in narrative understanding. *Journals of Gerontology, 57B*, P291–P297.

Stine-Morrow, E. A. L., Parisi, J. M., Morrow, D. G., Greene, J., & Park, D. C. (1997). An engagement model of cognitive optimization through adulthood. *Journal of Gerontology: Series B, Psychological Sciences and Social Sciences*, 62–69.

Stine-Morrow, E. A. L., Ryan, S., & Leonard, J. S. (2000). Age differences in on-line syntactic processing. *Experimental Aging Research, 26*, 315–322.

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

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging, 18*, 415–429.

Till, R. E. (1985). Verbatim and inferential memory in young and elderly adults. *Journal of Gerontology, 40*, 316–323.

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition, 16*, 283–298.

Till, R. E., & Walsh, D. A. (1980). Encoding and retrieval factors in adult memory for implicational sentences. *Journal of Verbal Learning and Verbal Behavior, 19*, 1–16.

Valencia-Laver, D. L., & Light, L. L. (2000). The occurrence of causal bridging and predictive inferences in young and older adults. *Discourse Processes, 30*, 27–56.

van den Broek, P. (1990). Causal inferences and the comprehension of narrative texts. In A. C. Graesser & G. H. Bower (Eds.). *The psychology of learning and motivation: Inferences and text comprehension* (Vol. 25, pp. 175–196). New York: Academic Press.

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes, 39*, 299–316.

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

Weingartner, K. M., Guzman, A. E., Levine, W. H., & Klin, C. M. (2003). When throwing a vase has multiple consequences: Minimal encoding of predictive inferences. *Discourse Processes, 36*, 131–146.

White, C., Ratcliff, R., Vasey, M., & McKoon, G. (2009). Dysphoria and memory for emotional material: A diffusion model analysis. *Cognition and Emotion, 23*, 181–205.

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010a). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology, 54*, 39–52.

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010b). Anxiety enhances threat processing without competition among multiple inputs: A diffusion model analysis. *Emotion, 10*, 662–677.

Zacks, R. T., Hasher, L., Doren, B., Hamm, V., & Attig, M. S. (1987). Encoding and memory of explicit and implicit information. *Journal of Gerontology, 42*, 418–422.

Zelinski, E. M., & Miura, S. A. (1990). Anaphor comprehension in younger and older adults. *International Journal of Aging and Human Development, 31*, 111–134.

Zipin, L. M., Tompkins, C. A., & Kasper, S. C. (2000). Effects of foregrounding on predictive inference generation by normally aging adults. *Aphasiology, 14*, 115–131.

Zurif, E., Swinney, D., Prather, P., Wingfield, A., & Brownell, H. (1995). The allocation of memory resources during sentence comprehension: Evidence from the elderly. *Journal of Psycholinguistic Research, 24*, 165–182.

Zwaan, R. A. (2008). Time in language, situation models, and mental simulations. *Language Learning, 58*, 13–26.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*, 292–297.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 386–397.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185.