MEMORY, 2006, 14 (6), 672–691

Psychology Press
Taylor & Francis Group

Recollection rejection of false narrative statements

C. J. Brainerd, V. F. Reyna, and S. Estrada
Cornell University, Ithaca, NY, USA

Our research was focused on a false-memory editing operation that is posited in fuzzy-trace theory—recollection rejection. The main objectives were (a) to extend model-based measurement of this operation to a narrative task that ought to ensure high levels of recollection rejection and (b) to study five manipulations that ought to influence recollection rejection by affecting the accessibility of verbatim traces of narrative statements: recency of narrative presentation, narrative repetition, type of false-memory item, testing delay, and repeated testing. The results showed that the narrative task did indeed yield high levels of recollection, with an estimated 49% of gist-consistent distractors being rejected in this way on initial memory tests. Consistent with current theoretical conceptions of false-memory editing, the results also showed that recollection rejection increased as a function of manipulations that should enhance the accessibility of verbatim traces of narrative statements, with repeated testing delivering especially large increases in verbatim accessibility.

We report some research that revolved around a false-memory editing operation that is posited in fuzzy-trace theory (FTT)—recollection rejection. The crux of the operation is that participants rely on verbatim traces of actual experience (e.g., ‘‘I ate a hamburger and drank a Pepsi at the baseball game’’) to perform rejections of false events that preserve the gist of their experience (‘‘ate a hot dog’’, ‘‘drank a Coke’’). Importantly, recollection rejection involves positive, principled suppression, a feature that contrasts with the passive rejections that result when false but gist-consistent events produce such weak memory evidence that they are not deemed to be viable candidates for acceptance. With recollection rejection, the evidence that accumulates from memory is strong, but it is disconfirmatory. According to the current theoretical model of recollection rejection, this process occurs because verbatim traces neutralise the seductively high familiarity of false but gist-consistent events (Brainerd, Reyna, Wright, & Mojardin, 2003). It follows that, other things being equal, recollection rejection is most likely to occur in circumstances in which verbatim traces are readily accessible in memory (e.g., see Gallo, 2004; Lampinen, Odegard, & Neuschatz, 2004). A situation of this sort was explored in our research.

Recollection rejection differs in two key respects from the older and more extensively studied notion of recollection (e.g., Jacoby, 1991). The first difference is concerned with the type of memory response that each supports (see Rotello, 2001). Recollection rejection, as just noted, produces principled suppression of false but gist-consistent events, whereas recollection produces principled acceptance of true events. The other difference is concerned with the types of memory representations that the two operations tap (see Brainerd, Reyna, & Mojardin, 1999). Recollection rejection is assumed to involve retrieval of verbatim traces of the surface form of experience, whereas recollection is traditionally assumed to involve retrieval of semantic memories.

Prior studies of recollection rejection have dealt almost entirely with memory for word lists, usually long lists of unrelated words (for a review, see Brainerd & Reyna, 2005). In contrast, the

---

Address correspondence to: C. J. Brainerd, Department of Human Development, Cornell University, Ithaca, NY, USA.

This research was supported by a National Science Foundation grant (SBR-9730143) and by a National Institute of Health grant (NIH31620).

http://www.psypress.com/memory DOI:10.1080/09658210600648449

TABLE 1
Examples of Reyna and Kiernan's (1994) narrative memory paradigm

Narratives Sentences	*1*	*2*
Targets	The bird is in the cage. The cage is on the table. The bird has green feathers.	The coffee is hotter than the tea. The tea is hotter than the cocoa. The cocoa is sweet.
Target probes		
TPO1	The bird is in the cage.	The coffee is hotter than the tea.
TPO2	The cage is on the table.	The tea is hotter than the cocoa.
False-memory probes		
TPN	The table is under the cage.	The cocoa is cooler than the tea.
TIO	The bird is on the table.	The coffee is hotter than the cocoa.
TIN	The table is under the bird	The cocoa is cooler than the coffee.
Control probes		
FPO	The table is on the cage.	The cocoa is hotter than the tea.
FPN	The cage is under the table.	The tea is cooler than the cocoa.
FIO	The table is on the bird.	The cocoa is hotter than the coffee.
FIN	The bird is under the table.	The coffee is cooler than the cocoa.

present research was focused on a narrative paradigm that was developed by Reyna and Kiernan (1994, 1995). The paradigm has the following features. First, participants are exposed to one or more three-sentence narratives that follow the pattern exemplified by the two narratives in Table 1. The first two sentences specify a familiar relation (spatial or magnitude in Table 1) that holds among a trio of everyday objects (between a bird, a cage, and a table, or between coffee, tea, and cocoa in Table 1). These two sentences are called premises. The third sentence in each narrative is a filler target that refers to one of the three objects but not to the relation that is specified in the premises. Second, following such narratives, sentence-recognition tests are administered that consist of three types of probes: (a) the relation-specifying targets (the two premises; TPO1 and TPO2 in Table 1); (b) three unpresented but gist-consistent distractors (the false-memory probes in Table 1); and (c) four distractors that do not preserve the gist of experience and serve as response-bias measures (the control distractors in Table 1). Concerning the three false-memory probes, note that the first (TPN in Table 1) is a true paraphrase of a premise that is generated by inserting a new word (e.g., "The coffee is hotter than the tea" becomes "The tea is *cooler* than the coffee"); the second (TIO in Table 1) is a valid inference that uses only old words to connect two objects that were not paired in either premise (e.g., "The coffee is hotter than the cocoa"); and the third (TIN in Table 1) is a valid interference that uses a new word in conjunction with old words to connect two objects that were not paired in either premise (e.g., "The cocoa is *cooler* than the coffee"). Finally, concerning the response-bias items, note that there is one control probe that matches each of the gist-consistent probes (targets and false-memory probes) in degree of *surface overlap* with presented sentences, thereby controlling simultaneously for hits or false alarms that are due to nonmemorial factors, such as guessing, as well as for hits or false alarms that are due to the overall familiarity of the words and word orderings that comprise these probes. Thus, FPO probes (false paraphrases of premises that contain only old words) are control items for premises; FPN probes (false paraphrases of premises that contain new words) are control items for TPN distractors; FIO probes (invalid inferences that contain only old words) are control items for TIO distractors; and FIN probes (invalid inferences that contain new words) are control items for TIN distractors.[1]

---

[1] Because it was necessary to administer eight distinct types of recognition probes to unconfound surface familiarity from meaning resemblance, it is necessary to use three-letter acronyms for these probes to avoid numerous repetitions of phrases such as "true premises with original wording". As these acronyms appear, their meanings can be regenerated merely by remembering that they follow a 2 $\times$2 $\times$2 factorial structure: The first letter indicates whether or not a probe is true, in the sense of being consistent with the gist of a narrative (T = true or F = false); the second letter indicates whether a probe mentions two objects that were connected in one of the premise sentences or whether it is an inference that involves two objects that were not connected in the premises (P = premise or I = inference); and the third letter indicates whether the probe contains only old words from the narrative or contains a new word (O = old or N = new).

Baseline data on this paradigm are available for normal children (Brainerd & Mojardin, 1998; Reyna & Kiernan, 1994, 1995), young adults (Gerkens & Smith, 2004; Reyna & Kiernan, 1995), and learning-disabled persons (Brainerd & Reyna, 2002; Karibian, 2003). Such data show that participants from these populations exhibit false-memory effects for unpresented but gist-consistent sentences (i.e., false-alarm rates are higher for the TPN, TIO, and TIN distractors than for their control distractors), even on recognition tests that are administered immediately following the presentation of a single narrative. These false-memory effects are also observed on recognition tests that are administered after the presentation of multiple narratives, or after delays of a few days to a few months.

This paradigm has four major advantages as a methodology for investigating recollection rejection. First and most obviously, unlike word lists, the paradigm involves narrative statements that describe everyday experiences, so that recollection rejection can be studied in highly meaningful contexts that emulate circumstances in which suppression of false memories occurs in everyday life. Second, these short, meaningful narratives ought to yield strong, readily accessible verbatim traces of the individual statements, thereby creating the memory conditions in which, according to theory, recollection rejection is especially likely to operate. Third, the test probes precisely control three forms of surface information that might encourage acceptance of unpresented sentences independently of their meaning overlap with premises: familiarity of words, word orderings within sentences, and sentence locus of words. As noted, each false-memory item (TPN, TIO, TIN) is paired with a control item (FPN, FIO, FIN) that matches it with respect to whether it contains a new word versus only old words, whether word ordering is the same as in the premises, and whether words appeared in the same premise. A semantic false-memory effect is therefore present whenever false-alarm rates are higher for the former than for the latter items, because the only difference between the two types of probes is whether they overlap in meaning with premises. Owing to precise control of the familiarity of surface features, when recollection rejection is measured for these particular false-memory items, it is a relatively pure estimate of the tendency to suppress false recognition of familiar meanings, rather than an estimate that is contaminated by uncontrolled influences of surface familiarity. Fourth, this control feature should also encourage the use of recollection rejection because test probes provide three simple surface clues that are perfectly diagnostic of whether the probes were presented: Are new words present? Are words presented in a different order than in the premises? Did any words not appear together in the same premise? A "yes" answer to any of these questions means that a probe was not presented, and each question is readily answered by comparing probes to verbatim traces of the premises.

Reyna, Mills, Estrada, and Brainerd (in press) recently pointed out that the use of recollection rejection to suppress false memories leads to a series of predictions about fine-grain aspects of false-memory data, and they reviewed data bearing on them. Examples of important predictions include: (a) that the stochastic relation between true- and false-memory responses should vary systematically as time passes, beginning as negative dependency, followed by independence, and ending as positive dependency; and (b) that because recollection rejection does not depend on metacognitive knowledge or complex decision strategies, it should be within the capabilities of young children. Reyna et al. concluded that available data are consistent with such predictions. However, because responses to false-memory items involve a mixture of cognitive processes, recollection rejection cannot be directly measured without the aid of mathematical models that disentangle this operation from other cognitive processes. Two models are currently available for this purpose, the phantom ROC model of Lampinen and associates (e.g., Lampinen, Odegard, Backshear, & Toglia, 2005; Lampinen et al., 2004; see also Rotello, 2001, for a parallel technique) and the conjoint-recognition model (e.g., Brainerd et al., 1999; Brainerd, Reyna, Wright, & Mojardin, 2001; Brown & Gorfein, 2004; Odegard & Lampinen, in press). We used the latter, a choice that was dictated by the nature of our recognition tasks. The conjoint-recognition model is applicable to tasks in which participants make binary decisions about probes, such as accept or reject, whereas the phantom ROC model is applicable to tasks in which participants make graded confidence judgements about probes, usually ranging from sure–old to sure–new. In the conjoint-recognition model, participants' responses to gist-consistent distractors are assumed to be controlled by four processes: recollection rejection, phantom recollection, similarity judgement, and response bias. Phantom recollection refers to circumstances in which such a distractor is accepted owing to illusory vivid recollections of its prior ‘‘presentation’’ (i.e., its prior ‘‘presentation’’ flashes in the mind's eye or echoes in the mind's ear), whereas similarity judgement refers to circumstances in which acceptance occurs because a distractor's meaning overlaps strongly with premises, notwithstanding that its ‘‘presentation’’ cannot be recollected. In a standard recognition condition in which participants are instructed to accept only targets, the model's expression for the probability of false alarms to such distractors is

$$p_{V,D} = (1-R)P + (1-R)(1-P)S_D + (1-R)(1-P)(1-S_D)\beta_V, \qquad (1)$$

where R is the probability of recollection rejection, P is the probability of phantom recollection, S_D is the probability of similarity judgement, and β_V is the response-bias probability. This is known as the ‘‘verbatim’’ (V) condition. Obviously, the parameter of interest, R, cannot be estimated with this equation because there are four theoretical processes on the right side but only one empirical degree of freedom on the left, $p_{V,D}$. This problem is resolved by requiring additional groups of participants to respond to the same series of recognition probes under two further types of instructions: gist (G: accept only gist-consistent distractors) and verbatim+gist (VG: accept targets and gist-consistent distractors). The conjoint-recognition model supplies other expressions for the probability of accepting gist-consistent distractors under these alternate instructions. The other expressions, together with the model's expressions for acceptance of distractors that are not gist consistent, provide sufficient degrees of freedom to (a) obtain independent estimates of R, P, and S_D, (b) to evaluate goodness of fit, and (c) to test statistical hypotheses about which experimental manipulations affect levels of recollection rejection. Mathematical details of these features of the model were reported by Brainerd et al. (2001).

In our research, then, the conjoint-recognition model was used to measure levels of recollection rejection in the narrative paradigm. There were two primary objectives. The first was simply to gather baseline data about the levels of recollection rejection that can normally be expected for the three types of gist-consistent distractors. As mentioned, it was anticipated that these levels would be high in comparison to those than have typically been observed with word lists, owing to features of the narrative paradigm that should redound to the benefit of verbatim memory. The other objective was to conduct tests of the current theoretical conception of recollection rejection as a verbatim-based operation. According to this conception, recollection rejection relies on retrieval of verbatim traces of actual experience to suppress false recognition of unpresented items that preserve the gist of experience. Consequently, if the degree of meaning overlap between targets and gist-consistent distractors is held constant, observed levels of recollection rejection ought to be affected by two types of manipulations—namely, those that affect whether verbatim traces are available in storage and those that affect the accessibility of such traces. We included five specific manipulations of this sort in the design of our research.

The first was recency of narrative presentation. During the study phase, narratives were presented in pairs, followed by a recognition test for one but not the other. The normal expectation is that verbatim traces would be more readily accessible for a just-presented narrative than for a previously presented one (e.g., Brown & Gorfein, 2004). The second manipulation was narrative repetition. A number of findings in the false-memory literature suggest that repeated presentation of targets strengthens verbatim memory for their surface form relative to gist memory for their meaning (e.g., Budson, Sullivan, Mayer, Daffner, Black, & Schacter, 2002; Kensinger, & Schacter, 1999). Narrative repetition was manipulated between-condition and within-condition in order assess both item-level recollection rejection (e.g., rejecting individual distractors on the basis of verbatim memory for their corresponding targets) and another verbatim-based form of suppression that is metacognitive in nature, the distinctiveness heuristic (e.g., Schacter, Israel, & Racine, 1999). Recent research suggests that both processes operate to varying degrees in false-memory tasks that involve word lists (see Hege & Dodson, 2004). The third manipulation was the type of gist-consistent distractor. The expectation, based on earlier research (Gerkens & Smith, 2004; Reyna & Kiernan, 1994, 1995), was that some of these distractors would support recollection rejection better than others and that, in particular, recollection rejection would increase as the level of surface mismatch between gist-consistent distractors and their corresponding targets increased. By this criterion, TIN distractors should be superior to TPN or TIO distractors.

The fourth manipulation was delay. Recognition tests for some narratives were administered immediately following presentation, whereas others were tested following a 1-week delay. Given the steep forgetting function for verbatim memory for narrative sentences (Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Murphy & Shapiro, 1994), it was naturally assumed that such traces would be less accessible 1 week after presentation and that recollection rejection would be correspondingly less likely to occur. The fifth and final manipulation was repeated testing. During the 1-week delayed session, recognition tests were administered for narratives that had been tested immediately after presentation, as well as for narratives that had not been previously tested. Prior studies of the effects of immediate recognition tests on delayed recognition suggest that they help to inoculate verbatim memory against the effects of forgetting, enhancing the accessibility of verbatim traces on delayed tests (Brainerd & Reyna, 1996; Reyna & Lloyd, 1997; Warren & Lane, 1995). Hence, if recollection rejection depends on accessing verbatim traces it should be more likely to occur, after a 1-week delay, for narratives that were tested immediately after presentation than for narratives that were not.

METHOD

Participants

A total of 144 undergraduate students from the University of Texas at Arlington participated in the study in order to fulfil a course requirement. First, participants were randomly assigned to the three instructional conditions of conjoint recognition, V, G, and VG. Second, within these instructional conditions, each participant was randomly assigned to one of four repetition conditions: (a) high (H: all narratives presented twice); (b) low (L: all narratives presented once); (c) mixed one (M1: the first member of narrative pairs was presented twice and the second was presented once); (d) mixed two (M2: the first member of narrative pairs was presented once and the second was presented twice).

Materials

The study and test materials were obtained from Reyna and Kiernan (1994) and consisted of 16 three-sentence narratives like those in Table 1, together with their corresponding recognition probes. Twelve narratives and probes were randomly selected from the 16 sets for presentation and testing of individual participants. The recognition test for each narrative consisted of 10 probes: 3 targets (TPO1, TPO2, and the filler), 3 gist-consistent distractors (TPN, TIO, and TIN), and 4 control distractors (FPO, FPN, FIO, and FIN).

Procedure

There were two experimental sessions, separated by a 1-week interval. Participants received general memory instructions at the start of the first session. These instructions informed them that they would be listening to a series of three-sentences stories, and memory tests would be administered for some but not all of the stories. They were further informed that each memory test would consist of a series of sentences, that some sentences would be identical to the ones they had heard, that the rest of the sentences would be new, that some of the new sentences would be false in that the two objects would have a different relation than in the story, and the other new sentences would be true in that the two objects would have the same relation as in the story.

Following these general instructions, which were the same for all participants, each participant received further specialised instructions that were specific to his or her condition. Participants in the V condition were told: “Whenever you hear a sentence that you heard before when I read you the story, say YES. Whenever you hear a sentence that you did not hear in the story and is false, say NO. Whenever you hear a sentence that you did not hear in the story, but think is true say NO.” Participants in the G condition were told: “Whenever you hear a sentence that you heard before when I read you the story say NO. Whenever you hear a sentence that you did not hear in the story and is false say NO. Whenever you hear a sentence that you did not hear in the story, but think is true say YES.” Participants in the VG condition were told: “Whenever you hear a sentence that you heard before when I read you the story say YES. Whenever you hear a sentence that you did not hear in the story and is false say NO. Whenever you hear a sentence that you did not hear in the story, but think is true say YES." The specialised instructions were illustrated with the practice narrative "The ball is in the toy box. The toy box is in the room. The room belongs to Michael." All 10 recognition probes of this narrative were read in random order, and the correct yes–no response for each was indicated by the experimenter. Each participant was informed that the experimenter would repeat the instructions and examples as many times as needed for the participant to gain full understanding.

After full understanding was achieved, the 12 narratives were presented (read aloud) in pairs, with each pair being followed by a recognition test for one of the narratives (randomly selected) that was composed of the 10 probes described above, presented in random order. The probes, like the narratives, were read aloud, with participants responding "yes" or "no" to each. Thus, the procedure for the first session consisted of six cycles of narrative presentation followed by a recognition test. The recognition test was for the first narrative in three of these cycles and for the second narrative in the other three cycles. For participants in the H condition, both narratives in each cycle were presented twice. For participants in the L condition, both narratives in each cycle were presented once. For participants in the M1 condition, the first narrative in each cycle was presented twice and the second was presented once. For participants in the M2 condition, the first narrative in each cycle was presented once and the second was presented twice.

One week after the first session, the participants returned for the delayed session. They had been told at the end of the first session that they would receive another memory test during the delayed session, but the nature of the test was not specified. The instructions that each participant had received about how to respond to recognition tests were repeated. No narratives were presented during the delayed session. Instead, only recognition tests were administered. Specifically, a full 10-item recognition test was now administered for each of the six narratives that were not tested during the first session. For the previously tested narratives, the 10-item recognition test for each narrative was repeated. The order of probe presentation on the test for each narrative was random, and the order in which the 12 narratives were tested was also random.

RESULTS

The findings are presented in two waves. First, we consider results for the immediate recognition tests. Second we consider results for the delayed recognition tests. Within each of these sections, we briefly sketch overall patterns of qualitative results. This is followed by a detailed report of the process-level results that were obtained from the conjoint-recognition model.

Immediate tests

Qualitative findings. Acceptance probabilities for the immediate tests are reported in Table 2 for the first narrative in each pair, with probabilities being reported separately by instructional condition (V, G, VG), level of repetition (once versus twice), and type of memory probe (premise targets, gist-consistent distractors, control distractors). Concerning the level of repetition factor, probabilities are not reported separately by repetition condition (H, L, M1, M2) because preliminary analyses revealed that the effects of narrative repetition did not vary as a function of whether repetition was manipulated within- versus between-condition. All subsequent analyses of this variable were therefore collapsed across within- versus between-condition repetition. Acceptance probabilities for the second narrative in each pair are reported in the same manner in Table 2.

The data in Table 2 are not of primary interest because they do not involve the measurement of recollection rejection or any other memory process. However, five patterns will be mentioned, each of which is readily apparent in these tables, which are in the nature of manipulation checks. First, note that the acceptance probabilities for targets are much higher in the V and VG conditions than in the G condition, which shows that participants readily followed instructions. Second, note that the acceptance probabilities for gist-consistent distractors were much higher in the G and VG conditions than in the V condition, which also shows that participants readily followed instructions. Third, participants displayed excellent memory for the premises, in that mean

TABLE 2
Acceptance probabilities on the immediate recognition test

Instructional Condition/Sentences	*Verbatim*	*Gist*	*Verbatim+Gist*
First narrative			
Presented once:			
TPO1	.82	.19	.83
TPO2	.77	.25	.77
TPN	.22	.66	.72
TIO	.35	.67	.70
TIN	.14	.58	.53
FPO	.18	.13	.20
FPN	.08	.23	.31
FIO	.16	.23	.20
FIN	.05	.18	.25
Presented twice:			
TPO1	.87	.21	.91
TPO2	.85	.26	.93
TPN	.14	.76	.72
TIO	.35	.64	.90
TIN	.16	.65	.83
FPO	.13	.09	.12
FPN	.07	.15	.22
FIO	.12	.13	.12
FIN	.03	.19	.13
Second narrative			
Presented once:			
TPO1	.77	.26	.84
TPO2	.80	.33	.86
TPN	.23	.53	.84
TIO	.27	.59	.82
TIN	.09	.57	.77
FPO	.07	.18	.23
FPN	.09	.26	.25
FIO	.07	.22	.14
FIN	.07	.20	.23
Presented twice:			
TPO1	.82	.26	.90
TPO2	.84	.26	.85
TPN	.25	.56	.75
TIO	.37	.68	.83
TIN	.12	.65	.69
FPO	.16	.25	.17
FPN	.00	.28	.28
FIO	.12	.25	.13
FIN	.02	.21	.06

TABLE 3
Estimates of false-memory processes on the immediate recognition test

	Repetition/probe					
Parameter	*R*	*P*	*S_D*	β_V	β_G	β_{VG}
First narrative						
Presented once:						
TPN	.44	.14	.28	.10	.23	.18
TIO	.31	.08	.42	.18	.23	.20
TIN	.32	.00	.18	.06	.19	.18
Presented twice:						
TPN	.56	.00	.27	.07	.16	.21
TIO	.55	.59	.40	.12	.13	.12
TIN	.62	.45	.00	.03	.19	.13
Second narrative						
Presented once:						
TPN	.51	.50	.00	.09	.25	.26
TIO	.54	.53	.03	.07	.22	.15
TIN	.55	.19	.00	.07	.18	.26
Presented twice:						
TPN	.44	.46	.00	.02	.28	.28
TIO	.49	.36	.08	.12	.24	.13
TIN	.59	.24	.01	.06	.21	.13

hit rates for TPO1 and TPO2 probes in the V and VG conditions were above .80, whereas the corresponding false-alarm rates for their control distractors (FPO) were below .20. Fourth, semantic false-memory effects were clearly present in the V condition. Over narratives and repetition levels, for instance, the average false-alarm rate for the three types of gist-consistent distractors was three times the average false-alarm rate for the corresponding control distractors (.22 versus .07). Fifth, as in prior conjoint-recognition experiments (e.g., Brainerd et al., 1999, 2001; Odegard & Lampinen, in press), false-alarm rates for distractors that did not preserve the gist of experience (FPO, FPN, FIO, FIN) were influenced by instructional condition. In particular, as in prior experiments, the false-alarm rate in the V condition ($M = .09$) was lower than in either the G condition ($M = .20$) or the VG condition ($M = .19$), the usual interpretation being that participants are more susceptible to response bias in the latter two conditions.

Process-level findings. The data of principal interest for the immediate tests were those that were obtained by estimating the conjoint-recognition model's parameters. Parameter estimates for false-memory items appear in Table 3, and corresponding estimates for targets appear in Table 4. We consider the process findings for the two types of probes separately.

False-memory items. In Table 3 we report estimates of recollection rejection (R), the other two memory processes (phantom recollection, P, and similarity judgement, S_D), and the response bias parameters for the three instructional conditions of conjoint recognition (β_V, β_G, and β_{VG}). Before examining these findings, some statistical preliminaries must be discussed. The statistical machinery for applying the conjoint-

TABLE 4
Estimates of true-memory processes on the immediate recognition test

	Repetition/probe					
Parameter	*I*	*E*	S_T	β_V	β_G	β_{VG}
First narrative						
Presented once:						
TPO1	.62	.02	.42	.20	.13	.20
TPO2	.52	.07	.44	.20	.13	.20
Presented twice:						
TPO1	.70	.00	.64	.03	.09	.11
TPO2	.54	.00	.47	.03	.09	.11
Second narrative						
Presented once:						
TPO1	.55	.08	.49	.08	.18	.23
TPO2	.51	.11	.57	.08	.18	.23
Presented twice:						
TPO1	.71	.33	.48	.15	.35	.18
TPO2	.64	.09	.53	.15	.35	.18

recognition is provided elsewhere (e.g., Brainerd et al., 2001). Generally speaking, the model is applied to data by implementing multinomial modelling software, such as Hu's (1995) general processing tree program, to conduct four types of analyses: (a) goodness of fit; (b) parameter estimation; (c) within-condition parameter comparisons (Do two different parameters tend to have the same value within experimental conditions?); and (d) between-condition parameter comparisons (Do manipulations produce different values of a certain parameter?). Goodness-of-fit tests are necessary preliminaries to the other types of analyses. Because the method of maximum likelihood is used to estimate the model's parameters, these particular analyses involve standard likelihood ratio tests that are asymptotically distributed as chi-squared, and, therefore, goodness-of-fit tests boil down to a series of chi-squared tests of the null hypothesis that the model provides a statistically tolerable approximation to the data, with one such test for each experimental condition (see Brainerd et al., 1999, for worked examples).

There were a total of total of 12 conditions for which goodness-of-fit tests were computed—3 types of false-memory probes (TPN, TIO, TIN); ×2 narrative presentation positions (first narrative versus second narrative); ×2 levels of narrative repetition (one presentation versus two presentations). The data space over which the conjoint-recognition model is currently defined (cf. Brainerd et al., 2001) contains 12 degrees of freedom (i.e., 12 independent empirical probabilities), and there are 9 theoretical processes to estimate: the three false-memory parameters (R, P, and S_D), the response bias parameters for the three instructional conditions (β_V, β_G, β_{VG}), and three true-memory parameters (see below). Therefore, the likelihood ratio test for each of these conditions was a $\chi^2(3)$ statistic, with a critical value of 7.82 being required to reject the null hypothesis that the model fits the data of any given condition. These tests showed that the model delivered acceptable fits, producing a mean value of 5.07 over the various conditions. We therefore proceed to analyses that involve the model's parameters.

Four findings for the recollection rejection parameter, R, are of interest. First, consider the absolute values of R in the first data column of Table 3. We mentioned earlier that this narrative task has features that, relative to word-list tasks, should ensure excellent verbatim memory—so that if recollection rejection involves retrieval of verbatim traces of targets, absolute values of R ought to be high. That prediction was confirmed. Across conditions and probe types, the mean probability of recollection rejection was .49, which is more than twice the corresponding probability, for instance, in the conjoint-recognition experiments reported by Brainerd et al. (2001), Brainerd and Wright (2005), and Odegard and Lampinen (in press), and in the phantom ROC experiments reported by Lampinen et al. (2005) and Rotello (2001). In all of the latter experiments, participants responded to recognition tests after studying long word lists. Beyond the baseline datum of high levels of recollection rejection for gist-consistent narrative statements, if the retrieval of verbatim traces of targets is involved in recollection rejection, one would expect that, other things being equal, R should be significantly affected by manipulations that, on the face of it, ought to enhance verbatim memory. The second, third, and fourth results of interest are concerned with this latter prediction.

The second result is concerned with recency of narrative presentation. The normal expectation would be that verbatim traces of the sentences that comprise the second narrative in each pair would be more readily accessible and, therefore, produce higher levels of recollection rejection than verbatim traces of the sentences that comprise the first narrative in each pair, particularly for narratives that are only presented once. Consistent with this hypothesis, the average value of R for the TPN, TIO, and TIN probes was .53 for the second narrative when it was presented once, but this same average value was .34 for the first narrative when it was presented once. Between-condition likelihood ratio tests showed that R was reliably larger for the second narrative for each of the three types of false-memory items, showing that it was easier to reject them on the basis of target information when they referred to the most recent narrative in the narrative pairs: The null hypothesis that the estimated values were the same for the two narrative positions was rejected for TPN, TIO, and TIN with a $\chi^2(1)$ test that had a critical value (.05 level of confidence) of 3.84.[2] In contrast, recency of narrative presentation did not affect estimates of R for narratives that were presented twice, which is not remarkably because repetition itself is a manipulation that should enhance verbatim memory.

The third result is concerned with narrative repetition, which, like recency of presentation, should affect R because it ought to make verbatim traces of narrative sentences more accessible on recognition tests, particularly for the first narrative in each pair. In line with this prediction, it can be seen in Table 3 that the average value of R for the TPN, TIO, and TIN probes for first narratives was .58 for narratives that were presented twice, but the same average value was .36 for first narratives that were presented once. Between-condition likelihood ratio tests showed that R was reliably larger for the twice-presented narratives than for once-presented narratives for each of the three types of false-memory items, showing that it was easier to reject them for narratives that should have produced more accessible verbatim traces of targets: The null hypothesis that the estimated values were the same for the two levels of repetition was rejected for TPN, TIO, and TIN with a $\chi^2(1)$ test that had a critical value (.05 level of confidence) of 3.84. In contrast, narrative repetition did not increase estimates of R for narratives that were presented as the second narrative of each pair. Again, this is not surprising because, as we just saw, recency of narrative presentation substantially elevates recollection on its own.

The fourth finding about recollection rejection is that there was evidence that it varied as a function of the type of false-memory item. Earlier, we mentioned that it should be easier to successfully reject gist-consistent distractors, using verbatim traces of targets, as the surface differences between such distractors and targets multiply. From this perspective, the TIN distractor, which contains new wording and words that did not appear together in narrative sentences, should be easier to reject than either TPN or TIO distractors. There was support for this hypothesis for narratives that were presented twice (for which verbatim traces should be especially easy to access). For such narratives, the mean value of R was .61 for TIN probes and .51 for TPN and TIO probes. Within-condition likelihood ratio tests showed that R was reliably larger for TIN than for either TPN or TIO. Like the between-condition parameter tests, above, each of these within-condition tests involved a $\chi^2(1)$ statistic and a critical value (.05 level of confidence) of 3.84.[3]

Summing up the model-based results for recollection rejection, the findings were congruent with the view, derived from current theory, that

---

[2] Because the conjoint-recognition model's parameters are estimated by the method of maximum likelihood, between-condition significance tests of those estimates also involve likelihood ratios. The specific procedure for comparing the estimated values of a given parameter, R in this instance, between pairs of conditions involves three steps (see Brainerd et al., 1999, for details). First, the model's likelihood function is used to compute the joint likelihood of the data of the two conditions when all of the model's theoretical parameters are free to vary. Call this value L_{18} because a total of 18 parameters, 9 for each condition, are estimated when all parameters are free to vary. Second, the model's likelihood function is used to recompute the joint likelihood of the data, subject to the single constraint that R has the same value in each condition. Call this value L_{17} because one less parameter is free to vary. Third, the value of the test statistic $-2\ln[L_{17} \div L_{18}]$ is computed, which has an asymptotic $\chi^2(1)$ distribution and a critical value of 3.84 for rejection of the null hypothesis of no between-condition difference in parameter values at the .05 level of confidence.

[3] Mathematically, these tests are the same as between-condition tests, apart from the fact that they are computed for pairs of *parameters* using the data of individual conditions, rather than for a single parameter using pairs of *conditions* (see Brainerd et al., 1999, for details). First, the model's likelihood function is used to compute the likelihood of the data of a target condition when all nine of the model's theoretical parameters are free to vary. Call this value L_9. Second, the model's likelihood function is used to recompute the likelihood of the same data, subject to the constraint that two of the nine parameters (in this instance, β_V and β_G or β_V and β_G) have the same value. Call this value L_8 because one less parameter is free to vary. Third, the value of the test statistic $-2\ln[L_8 \div L_9]$ is computed, which has an asymptotic $\chi^2(1)$ distribution and a critical value of 3.84 for rejection of the null hypothesis of no difference in parameter values at the .05 level of confidence.

this operation depends on verbatim memory for targets. Other than the results for recollection rejection, the response bias parameters and the other memory parameters also reacted to some of the experimental manipulations. To begin, visual inspection of the estimates of the bias parameters in the last three data columns of Table 3 shows that for most trios of estimates, participants were more conservative (response bias was lower) when they were told to accept only target sentences (V condition) than when they were told to accept gist-consistent distractors (G or VG conditions). When within-condition likelihood ratio tests were computed for the 12 sets of parameter estimates in Table 3, β_V was found to be reliability smaller than β_G for all sets other than those in the second and fifth rows, and β_V was found to be reliability smaller than β_{VG} for all sets other than those in the second, fifth, and eleventh rows.

Turning to the other memory processes, phantom recollection (parameter *P)* and similarity judgement (parameter S_D), no specific predictions were made about these processes because the focus of the present research was squarely on recollection rejection. Further, our experimental paradigm and the manipulations that we included were ones that were designed to enhance verbatim memory for targets and thereby facilitate recollection rejection, whereas phantom recollection and similarity judgement are processes that, theoretically, are primarily influenced by gist memory (e.g. Reyna et al., in press). However, each process reacted to one of the experimental manipulations in ways that are consistent with this theoretical interpretation. Concerning phantom recollection, note, first, that this interesting illusory phenomenology was in evidence with the narrative memory task: The mean value of P in Table 3 was .30, which means (cf. Equation 1) that when participants were unable to perform recollection rejection on a gist-consistent distractor, they accepted it 30% of the time owing to illusory vivid remembrance of its "presentation". Perhaps these high levels of recollection rejection are grounded in the fact that the baseline surface resemblance between premises and gist-consistent distractors was quite high (see Table 1). Further, for narratives that were presented once, phantom recollection was significantly higher (i.e., between-condition null likelihood ratio tests produced null hypothesis rejections) for the second narrative in each pair ($M = .41$) than for the first narrative ($M = .08$). Although a theoretical explanation of this pattern could be constructed post hoc, it would be speculative without additional manipulations that were designed to test it. Turning to similarity judgement, this process was also influenced by recency of narrative presentation but in the opposite direction: The mean value of S_D in Table 3 was much higher for the first narrative in each pair ($M = .26$) than for the second ($M = .02$). Again, although a theoretical explanation of this pattern and of the double dissociation between phantom recollection and similarity judgement could be constructed post hoc, it would be speculative without additional manipulations that were designed to test it.

True-memory items. In addition to the false-memory processes of recollection rejection, phantom recollection, and similarity judgement, the conjoint-recognition model delivers estimates of three parallel true-memory processes that control participants' responses to target probes (TPO1 and TPO2): identity judgement (I), erroneous recollection rejection (E), and similarity judgement (S_T). Identity judgement occurs when participants perceive that there is an exact match between the surface form of a target probe (e.g., "The coffee is hotter than the tea") and a retrieved verbatim trace. Erroneous recollection rejection is an ersatz form of "false" memory editing that can occur when a target probe (e.g., "The coffee is hotter than the tea") fails to produce retrieval of its verbatim trace but produces retrieval of the verbatim trace of another target (e.g., "The tea is hotter than the cocoa"), and the perceived mismatch causes the probe to be mistakenly rejected. Last, similarity judgement occurs when a target probe fails to produce retrieval of verbatim traces but produces retrieval of gist traces of its meaning content, and the meaning overlap is perceived to be sufficiently strong to warrant acceptance. In a standard recognition condition (i.e., V instructions), the model expression for the target hit rate is

$$p_{V,T} = I + (1-I)(1-E)S_T + (1-I)(1-E)(1-S_T)\beta_V, \qquad (2)$$

where I is the probability of identity judgement, E is the probability of erroneous recollection rejection when identity judgement fails, S_T is the probability of similarity judgement when identity judgement and recollection rejection both fail, and β_V is the probability that response bias produces a hit when all three memory processes fail.

The conjoint-recognition model also delivers estimates of the three true-memory processes whenever it delivers estimates of the three false-memory processes. For the immediate test data of the present experiment, those estimates appear in Table 4. Three findings merit brief comment. First and most obviously, the value of the identity parameter is high for all data sets in Table 4 ($M = .60$), which means that target acceptances were dominated by identity judgements. Because identity judgement, like recollection rejection, ostensibly relies on verbatim traces, this is consistent with the aforementioned view that the narrative task produces high levels of recollection rejection because it ensures good verbatim memory for presented sentences. Second, the average value of the identity parameter is higher than the average value of the recollection rejection parameter (.49). As we have noted elsewhere (Brainerd et al., 2003), this is a test of the assumption that both processes involve retrieval of verbatim traces: By the principle of retrieval variability, targets should be better retrieval cues for their own verbatim traces (parameter *I)* than are gist-consistent distractors (parameter R). Third, the average level of similarity judgement for targets ($M = .51$) was higher than the average level of similarity judgement for gist-consistent distractors ($M = .19$). Of course, although these distractors are consistent with targets' meanings, the meaning match is more exact for targets.

One-week tests

Qualitative findings. Recall that during the delayed session, recognition tests were administered for all 12 narratives, the 6 that had been tested immediately and the 6 that had not been. Delayed acceptance probabilities for previously untested narratives are reported in Table 5, with acceptance probabilities being reported separately for first versus second narratives, instructional condition (V, G, VG), level of repetition (once versus twice), and type of memory probe (premise targets, gist-consistent distractors, control distractors). The corresponding data for previously tested narratives appear in Table 6.

TABLE 5
Acceptance probabilities on the delayed recognition test for previously untested probes

Instructional Condition/Sentences	*Verbatim*	*Gist*	*Verbatim + Gist*
First narrative			
Presented once:			
TPO1	.59	.22	.69
TPO2	.55	.35	.57
TPN	.23	.41	.61
TIO	.29	.41	.51
TIN	.08	.52	.49
FPO	.29	.32	.43
FPN	.18	.46	.47
FIO	.31	.38	.47
FIN	.27	.39	.45
Presented twice:			
TPO1	.59	.35	.63
TPO2	.63	.31	.59
TPN	.18	.63	.57
TIO	.29	.45	.51
TIN	.12	.59	.45
FPO	.23	.23	.41
FPN	.08	.35	.43
FIO	.23	.16	.43
FIN	.16	.27	.47
Second narrative			
Presented once:			
TPO1	.48	.31	.62
TPO2	.46	.35	.63
TPN	.29	.44	.58
TIO	.34	.53	.53
TIN	.16	.56	.52
FPO	.17	.22	.38
FPN	.14	.33	.40
FIO	.22	.22	.40
FIN	.22	.36	.54
Presented twice:			
TPO1	.50	.29	.68
TPO2	.43	.35	.58
TPN	.24	.45	.64
TIO	.29	.38	.58
TIN	.25	.38	.61
FPO	.19	.27	.32
FPN	.15	.31	.42
FIO	.16	.36	.38
FIN	.18	.40	.39

Inspection of Table 5 reveals some obvious differences in data patterns for the three types of probes, relative to the immediate test data in Table 2. Specifically, hit rates for targets dropped considerably over 1 week in the V and VG conditions, false-alarm rates for control distractors increased substantially in all instructional conditions, and hit rates for gist-consistent distractors dropped considerably in the G and VG conditions. Taken together, these patterns show that there was substantial forgetting of information about the actual sentences that had been presented a week earlier, as indexed by the difference between target hit rates and false-alarm rates for control distractors, and there was

TABLE 6

Acceptance probabilities on the delayed recognition test for previously tested probes

Instructional Condition/Sentences	*Verbatim*	*Gist*	*Verbatim + Gist*
First narrative			
Presented once:			
TPO1	.74	.23	.77
TPO2	.68	.25	.73
TPN	.23	.64	.73
TIO	.43	.55	.72
TIN	.20	.61	.68
FPO	.27	.17	.24
FPN	.19	.35	.35
FIO	.17	.23	.31
FIN	.12	.25	.30
Presented twice:			
TPO1	.84	.31	.89
TPO2	.71	.36	.86
TPN	.25	.64	.78
TIO	.28	.67	.85
TIN	.14	.76	.81
FPO	.24	.19	.20
FPN	.13	.19	.27
FIO	.17	.21	.15
FIN	.13	.19	.17
Second narrative			
Presented once:			
TPO1	.73	.31	.73
TPO2	.68	.33	.81
TPN	.11	.52	.81
TIO	.34	.65	.75
TIN	.09	.60	.79
FPO	.21	.19	.23
FPN	.09	.27	.46
FIO	.14	.27	.21
FIN	.20	.40	.27
Presented twice:			
TPO1	.66	.37	.85
TPO2	.70	.39	.72
TPN	.30	.53	.68
TIO	.32	.72	.83
TIN	.21	.58	.72
FPO	.23	.35	.21
FPN	.15	.32	.28
FIO	.17	.25	.17
FIN	.15	.35	.15

substantial forgetting of meaning content, as indexed by the difference between hit rates for gist-consistent distractors and false-alarm rates for control distractors.

In contrast, inspection of Table 6 reveals dramatically different patterns of delayed performance. Specifically, hit rates for targets declined only slightly in the V and VG conditions, false-alarm rates for control distractors increased but by much smaller amounts than for previously untested narratives, and hit rates for gist-consistent distractors did not decline in the G and VG conditions. Thus, the prior recognition tests had powerful inoculating influences on memory: Memory for information about the actual sentences that had been presented, as indexed by the difference between target hit rates and false-alarm rates for control distractors, was inoculated to a considerable degree, while inoculation of meaning content, as indexed by the difference between hit rates for gist-consistent distractors and false-alarm rates for control distractors, was virtually complete.

Process-level findings. As was the case for the immediate tests, the data of principal interest were those that were obtained by estimating the conjoint-recognition model's parameters. Parameter estimates for false-memory items appear in Table 7, with estimates being separately reported for previously untested and previously tested narratives. Corresponding parameter estimates for targets appear in Table 8.

False-memory items. Concerning goodness of fit, once again we computed a total of 12 goodness-of-fit tests—3 types of false-memory probes (TPN, TIO, TIN) ×2 narrative presentation positions (first narrative versus second narrative) ×2 levels of narrative repetition (one presentation versus two presentations)—for the 1-week recognition tests of previously untested narratives. We conducted a further, parallel series of 12 goodness-of-fit tests for the 1-week recognition tests of previously untested narratives. As before, the fit test was a likelihood ratio statistic that was asymptotically distributed as $\chi^2(3)$, so that a critical value of 7.82 was required to reject the null hypothesis that the model fits the data of any given condition at the .05 level. The pattern for fit was similar to the immediate test in that the both series of fit tests showed that the model delivered statistically acceptable approximations to the data, with a grand mean 5.88 for the $\chi^2(3)$ statistic.

(1) Untested narratives. Turning to the parameter estimates in the upper half of Table 7, we reconsider the same four findings for the recollection rejection parameter, R, that were of interest on the immediate recognition test. First, consider the absolute values of R in the first data column of Table 7. Even a week after presentation, participants were still able to perform recollection rejec-

TABLE 7
Estimates of false-memory processes on the 1-week recognition test

	Repetition/probe					
Parameter	*R*	*P*	S_D	β_V	β_G	β_{VG}
First narrative: previously untested						
Presented once:						
TPN	.10	.13	.00	.18	.44	.45
TIO	.04	.00	.00	.32	.39	.49
TIN	.17	.00	.00	.23	.34	.45
Presented twice:						
TPN	.00	.00	.42	.10	.35	.33
TIO	.17	.00	.14	.23	.18	.39
TIN	.23	.00	.14	.17	.35	.37
Second narrative: previously untested						
Presented once:						
TPN	.19	.19	.07	.14	.32	.45
TIO	.12	.00	.21	.22	.24	.44
TIN	.23	.00	.00	.22	.37	.41
Presented twice:						
TPN	.22	.11	.07	.16	.28	.44
TIO	.34	.16	.05	.16	.31	.41
TIN	.14	.18	.00	.17	.38	.41
First narrative: previously tested						
Presented once:						
TPN	.44	.16	.14	.19	.35	.34
TIO	.22	.18	.36	.17	.23	.31
TIN	.40	.10	.16	.12	.26	.29
Presented twice:						
TPN	.56	.46	.12	.10	.26	.27
TIO	.57	.44	.24	.16	.21	.15
TIN	.66	.18	.19	.13	.19	.17
Second narrative: previously tested						
Presented once:						
TPN	.47	.17	.00	.09	.25	.45
TIO	.40	.21	.34	.16	.27	.21
TIN	.57	.16	.00	.18	.36	.28
Presented twice:						
TPN	.33	.26	.12	.15	.32	.28
TIO	.50	.38	.00	.17	.25	.17
TIN	.53	.46	.00	.14	.34	.15

TABLE 8
Estimates of true-memory processes on the 1-week recognition test

	Repetition/probe					
Parameter	*I*	*E*	S_T	β_V	β_G	β_{VG}
First narrative: previously untested						
Presented once:						
TPO1	.40	.00	.10	.28	.30	.43
TPO2	.18	.00	.20	.28	.30	.43
Presented twice:						
TPO1	.23	.00	.27	.23	.25	.37
TPO2	.28	.00	.23	.23	.25	.37
Second narrative: previously untested						
Presented once:						
TPO1	.17	.00	.23	.17	.22	.50
TPO2	.14	.00	.23	.17	.22	.50
Presented twice:						
TPO1	.37	.13	.13	.17	.22	.50
TPO2	.33	.31	.05	.17	.22	.50
First narrative: previously tested						
Presented once:						
TPO1	.52	.08	.31	.27	.17	.24
TPO2	.45	.13	.26	.27	.17	.24
Presented twice:						
TPO1	.58	.15	.62	.24	.19	.20
TPO2	.49	.30	.50	.24	.19	.20
Second narrative: previously tested						
Presented once:						
TPO1	.40	.00	.43	.20	.17	.25
TPO2	.45	.29	.34	.20	.17	.25
Presented twice:						
TPO1	.51	.40	.36	.23	.35	.21
TPO2	.39	.46	.40	.23	.35	.21

tion, the mean value of *R* in the first data column ($M = .16$) suggesting that 16% of the gist-consistent distractors were rejected via surface mismatch with verbatim traces of targets. Although participants were still capable of recollection rejection, the average rate of recollection on the immediate test (.49) was three times higher. With respect to the other three findings, the effects on recollection rejection of some of the verbatim manipulations of the first session were still detectable. For instance, the second finding from the first session, that more recently presented narratives produced higher levels of recollection, was still present after a week: The average value of *R* for the TPN, TIO, and TIN probes was .21 for the second members of narrative pairs and .11 the first members of narrative pairs. Between-condition likelihood ratio tests showed that this recency difference in *R* was reliable for TPN and TIO probes but not for TIN probes. With respect to the third finding from the first session, inspection of the *R* estimates in Table 7 reveals that there was no consistent tendency for narrative repetition to increase recollection rejection a week after presentation. Last, with respect to the fourth finding from the first session, there was again evidence that recollection rejection varied as function of the type of false-memory item. As on the immediate test, the distractors whose surface forms differed most from targets, TIN, displayed higher values of *R* ($M = .19$) than the other two types of gist-consistent distractors ($M = .15$). Within-condition likelihood ration tests showed that the TIN versus TPN difference in *R* was reliable but the TIN versus TIO difference was not.

Considering findings for other parameters, the response-bias parameters again varied as a function of instructional condition. As on the immediate test, the average value of β_V (.19) was much smaller than the average value of the other two bias parameters (.39). Unlike the immediate test, however, the average value of β_G (.36) was smaller than the average value of β_{VG} (.42). Within-condition likelihood ratio tests showed that the β_V versus β_G difference was reliable for 9 of the 12 sets of parameter estimates in the upper half of Table 7, that the β_V versus β_{VG} difference was reliable for all 12 data sets, and that the β_G versus β_{VG} difference was reliable for 8 of the 12 data sets. With respect to the other two memory processes, there was a single dominant finding—namely that phantom recollection ($M =$ .06) and similarity judgement ($M = .09$) had both dropped to very low levels after a week.

(2) Tested narratives. Turning to the parameter estimates in lower half of Table 7, the recollection rejection findings for previously tested narratives were dramatically different from those for untested narratives. To begin, note the much larger estimates of R that appear in the first data column. When narratives were previously tested, the mean value of R ($M = .47$) indicates that 1 week after presentation, participants were still able to perform recollection rejection at very high levels, indeed at levels that were comparable to the immediate test ($M =$ .49). Moreover, with respect to the other three findings that were of interest on the immediate tests, the effects on recollection rejection of all of the verbatim manipulations from the study phase were still readily detectable. Thus, for narratives that had been presented once, the average value of R was greater for the more recently presented members of narrative pairs (Ms = .48 versus .35). Between-condition likelihood ratio tests showed that this recency effect was reliable for TIO and TIN distractors, but not for TPN distractors. In addition, for members of narrative pairs that had been presented first, repetition had the same effect as before: The average value of R was greater for narratives that had been presented twice than for narratives that had been presented once (Ms = .60 versus .35). Within-condition likelihood ratio tests showed that this effect was reliable for three of the four TIN versus TPN comparisons and that it was also reliable for three of the four TIN versus TIO comparisons.

Considering findings for other parameters, as in all previous analyses the response-bias parameters varied as function of instructional condition. As before, the average value of β_V (.18) was much smaller than the average value of the other two bias parameters (.27). Within-condition likelihood ratio tests indicated that the β_V versus β_G difference was reliable for 8 of the 12 sets of parameter estimates in the lower half of Table 7 and that the β_V versus β_{VG} difference was reliable for 7 of the 12 sets of parameter estimates. Unlike the corresponding results for previously untested narratives, the average values of β_G and β_{VG} (.27 versus .26) did not differ. A final important finding is that prior narrative testing had a beneficial effect on response bias in the two instruction conditions that inflate bias, G and VG. Whereas the mean bias estimate for these conditions was .27, recall that it was .39 for previously untested narratives. This reduction in response bias as a consequence of prior testing was apparent for both β_G (.27 versus .36) and β_{VG} (.26 versus .42). Between-condition likelihood ratio tests showed that the reduction in β_G was reliable for 5 of the 12 comparisons of tested versus untested narratives in Table 7, and the corresponding reduction in β_{VG} was reliable for 10 of these 12 comparisons. These findings are of considerable theoretical significance because certain models of false memory, such as the source-monitoring framework, predict that prior memory tests will *elevate* response bias on delayed tests (see Reyna, 2000).

With respect to the other two memory processes, the dominant finding was different than for previously untested narratives: The average values of phantom recollection ($M = .26$) and similarity judgement ($M = .14$) were both well above floor and, hence, higher than the corresponding averages for previously untested narratives. Further, although the mean value of S_D was noticeably lower than on the immediate test (.14 versus .26), the mean value of P was only slightly smaller than on the immediate test (.26 versus .30). (For S_D, between-condition likelihood ratio tests showed that four of the immediate versus delayed comparisons for S_D were reliable.) Thus, a previous recognition test produced significant inoculation of similarity judgement and virtually complete inoculation of phantom recollection. Finally, phantom recollection and similarity judgement both reacted to some of the manipulations. Unlike the immediate test, phantom recollection was higher for repeated narratives than for narratives that were only presented once

(Ms = .36 versus .16), but it was not affected by recency of narrative presentation. However, like the immediate test, S_D was affected by recency, with similarity judgement occurring at higher levels for the first narrative in each pair than for the second (Ms = .20 versus .08).

True-memory items. For targets, parameter estimates for the three memory parameters and the three bias parameters appear in Table 8. Estimates for previously untested narratives appear in the upper half of this table, while estimates for previously tested narratives appear in the lower half.

Taking the data for previously untested targets first, it will be recalled that on the immediate test, hits were dominated by verbatim-based judgements of identity. The mean value of I was .60, which means, under FTT's interpretation of this parameter, that 60% of target probes were accepted on the basis of exact surface matches between probes and their verbatim traces. There was, on the one hand, a precipitous decline in such judgements over the 1-week interval, with the mean value of I dropping by more than half (to .26). On the other hand, hits were still dominated identity judgements because S_T also dropped considerably (to .19). A second finding that was present on the immediate test is that the average value of the identity parameter was higher than the average value of the recollection rejection, which is in the nature of a validity test of FTT's assumptions that both involve retrieval of verbatim traces. The relation also held on the 1-week test (.26 versus .16). A third finding that was present on the immediate test is that the average level of similarity judgement for targets was much higher than the corresponding level of similarity judgement for gist-consistent distractors, which is congruent with the notion that although both types of probes match target meanings, the match will be higher for targets. This discrepancy between rates of similarity judgement for targets and distractors was still present on the delayed test (Ms = .16 versus .09), but it was much reduced relative to the immediate test (Ms = .51 versus .19).

We move now to the data in the lower half of Table 8, delayed target performance for narratives that had been tested 1 week earlier. As was the case for gist-consistent distractors, there is an overriding finding that dominates other results: A previous memory test had a powerful inoculating effect on true-memory processes. With respect to identity judgement, the mean value of I (.49) was nearly twice as large as the corresponding value for untested narratives (.26) and only slightly smaller than the corresponding value for the immediate test (.60). In the latter connection, a series of between-condition comparisons of estimates of the I parameter for previously tested narratives showed that (a) it was reliably larger than the corresponding value for untested narratives in seven of eight comparisons, but (b) it was reliability smaller than the corresponding value on the immediate test for only in five of eight comparisons. This inoculation effect also extended to the other two memory processes, erroneous recollection rejection and similarity judgement: The mean value of S_T was .40 (as compared to .09 for previously untested narratives) and the mean value of E was .23 (as a compared to .06 for untested narratives). Two important features of these data are that the value for S_T means that, after 1 week, hits were still dominated by identity judgement, and that the value of E actually increased over the 1-week delay (from .09 to .23). The latter result means that participants were far more willing to incorrectly reject a target on the basis of verbatim memory for other targets after time had passed, which is reasonable if verbatim memory is generally depleted after a week. The next finding was that the average value of the identity parameter continued to be higher than the average value of the recollection rejection parameter, but the difference was now small. The implication is that the verbatim inoculation influence of a prior memory test is proportionately greater for gist-consistent distractors than for targets. That finding, too, is sensible given that, initially, such distractors are poorer retrieval cues for verbatim traces: Logically, practice at retrieving verbatim traces on the immediate test ought to have more beneficial consequences for probes for which verbatim access is more difficult. Last, the third finding we have mentioned for targets, that similarity judgement is higher for targets than for gist-consistent distractors, was again present. For previously tested narratives, the pattern of similarity judgement being higher for targets than for gist-consistent distractors continued to be present. The mean values of S_T and S_D were, .40 and .14, respectively, which again is congruent with the notion that the level of meaning match between gist traces and probes is higher for targets than for gist-consistent distractors. A series of within-condition comparisons of the

1-week recognition data for previously tested narratives showed that S_T was reliability greater than S_D in 21 of the possible comparisons.

DISCUSSION

Because recollection rejection is not the only cognitive operation that, in theory, controls participants' responses to false-memory items, model-based measurement is required to disentangle its influences from those of other processes. Conjoint recognition is one of two extant techniques that accomplish this separation. Previous applications of conjoint recognition have focused on word-list paradigms whose features (e.g., the presentation of long lists of unrelated items) would not be expected to produce robust verbatim memory for most targets and, hence, should not produce particularly high levels of recollection rejection. The fact that estimates of recollection rejection have sometimes been quite low in such situations (e.g., Lampinen et al., 2005; Rotello, 2001) might lead some to question whether participants routinely rely on this operation, even in situations in which it is highly appropriate (cf. Gallo, 2004). Consequently, a primary aim of this research was to extend model-based measurement of recollection rejection to a paradigm that, on the face of it, would seem to ensure ready accessibility of verbatim traces of target information. Under current theoretical conceptions of recollection rejection (Brainerd & Reyna, 2005; Reyna & Lloyd, 1997), this should translate into high estimated values of parameters that measure this operation. This prediction was confirmed in comparisons involving the three different types of gist-consistent distractors (TPN, TIO, and TPN). On immediate recognition tests, the mean estimate of the R parameter revealed that participants performed successful recollection rejection of half of these distractor probes, a success rate that was only moderately lower than the corresponding level of verbatim-based acceptance of targets (60%, on average).

The other primary aim of this research was to generate further, direct evidence on the current theoretical interpretation of recollection rejection by studying manipulations that should affect storage and retrieval of verbatim traces of narrative statements. We included five manipulations of this sort in the design, the idea being that the parameter R should react to those manipulations in predictable ways. It did. The specific manipulations were recency of narrative presentation, narrative repetition, type of gist-consistent distractor, testing delay, and repeated testing. In all cases, the effects on R were consistent with theoretical expectations. On the immediate tests, recollection rejection was more likely for the more recently presented member of narrative pairs; it was more likely for narratives that had been presented twice rather than once; and it was more likely for distractors whose surface forms were more distinct from those of targets. These effects were also detected on delayed tests, 1 week after narrative presentation. Delay itself was expected to have a marked influence on recollection rejection, owing to many prior findings which suggest that verbatim traces become rapidly inaccessible as time passes—and it did. However, it was the last manipulation, repeated testing, that had by far the largest effects on recollection rejection.

Of course, it has long been known that as long as memory tests do not contain suggestive information, they can redound to the benefit of accuracy on later tests (e.g., Runquist, 1983). An explanation that has been proposed by various authors (e.g., Brainerd & Mojardin, 1998; Payne, Elie, Blackwell, & Neuschatz, 1996; Reyna & Lloyd, 1997; Warren & Lane, 1995) is that accessing verbatim traces on prior memory tests makes verbatim traces more accessible on subsequent tests. Comparing estimates of R (and the identity parameter for targets, I) on delayed tests for previously tested versus untested probes provided direct tests of this explanation. The results were clear. The average estimate of R on the delayed tests for previously tested narratives was three times the corresponding estimate for untested narratives, and remarkably, there was no reliable decline in R over the 1-week delay for narratives that had received immediate tests. In short, a prior memory test provided virtually complete inoculation against declines in recollection rejection over a 1-week interval. This conclusion is strengthened by the fact that a parallel pattern was observed for the target verbatim memory, I: The average estimate of I on the delayed tests for previously tested narratives was twice that for untested narratives, and the decline in I over the 1-week delay for previously tested narratives, though reliable, was small. Such findings are consistent with the verbatim-inoculation hypothesis.

These findings are also fairly unusual in the literature because, interestingly, although the verbatim-inoculation hypothesis has been discussed in several papers, we know of only one other experiment in which rates of recollection rejection on delayed tests were compared for previously tested and untested materials, using either the conjoint-recognition model or Lampinen et al.'s (2005) phantom ROC model. In that experiment (Experiment 2 of Brainerd et al., 2001), participants were administered 1-week delayed tests for long lists of words that had been studied 1 week earlier, with some words having been tested during the initial session and others not having been tested. Unlike the present experiment, the data did not reveal high levels of recollection rejection on the delayed test for semantically related distractors that had been previously tested ($M = .12$), and it also did not reveal that these levels were much higher than the corresponding ones for previously untested distractors ($M = .06$). As mentioned earlier, however, the level of recollection rejection on the immediate tests ($M = .16$) was much lower than in the present experiment, and hence there was little room for the inoculating influence of prior memory tests to be expressed. Despite these points of difference, there is one important similarity between the present findings and those of Brainerd et al.: In both instances, levels of recollection rejection on delayed and immediate tests were virtually the same for previously tested items ($M\text{s} = .47$ and .49 in the present experiment versus .12 and .16 in Experiment 2 of Brainerd et al.).

As the data were congruent with the ostensible importance of verbatim memory to recollection rejection, we conclude this paper by considering three theoretical issues that bear on this notion: (a) whether the variability in verbatim memory that was apparently responsible for variability in recollection rejection was chiefly a matter of differences in the sheer availability of verbatim traces in storage or differences in their degree of accessibility on memory tests; (b) whether there is any support for dual-process interpretations of recollection rejection; and (c) whether recollection rejection was indeed occurring on an item-by-item basis or whether a verbatim-based metacognitive strategy was operating. The crux of the first issue is whether, in this narrative task, the observed variations in recollection rejection can be tied primarily to differences in the formation and preservation of verbatim traces, or to differences in participants' ability to access those traces on recognition tests. On the whole, our findings favour a retrieval locus over a storage locus. Three findings, in particular, seem to argue for such a conclusion. First, there is one manipulation, type of gist-consistent distractor, whose effects could not have been due to differences in storage or preservation verbatim traces: When the availability of verbatim traces in storage was held constant within particular treatment conditions, recollection rejection rates were nevertheless higher for TIN distractors than for the other two types of false-memory items. Second, the manipulation that had the largest effects on recollection rejection was a retrieval manipulation—namely repeated testing. Although it has sometimes been proposed that a memory test can have beneficial effects on the quality of stored traces by redintegrating them once they have undergone some forgetting (Reyna & Kiernan, 1995; Reyna & Titcomb, 1997), that proposal would not seem to apply here because the initial memory tests for each narrative were administered less than a minute following its presentation, before there had been much if any opportunity for forgetting. Thus, the most sensible interpretation is that the powerful effect of the immediate tests on delayed recollection rejection lies in the practice that those tests provided at accessing verbatim traces when responding to recognition probes. Third, there was one manipulation in the experiment, narrative repetition, that could have unambiguously affected the availability of verbatim traces in storage because it provided additional opportunities to store the same verbatim traces. However, although narrative repetition increased recollection rejection, the effect was small and was limited to the first narrative in each pair. Further, although repetition of first narratives increased recollection rejection, it did not increase the corresponding verbatim memory parameter for targets (I), which are especially good retrieval cues for verbatim traces. Hence, given that repetition did not affect I, even its small effects on R might be retrieval effects (e.g., the second presentation of the first narrative might stimulate retrieval of verbatim traces of the first presentation). In sum, the overall pattern of results suggests that baseline availability of verbatim traces of narrative statements in storage was excellent and that variations in recollection rejection were chiefly differences in the accessibility of such traces on memory tests.

The second issue is whether the data provide any support for dual-process interpretations of recollection rejection. In FTT, it is assumed that the memory representations that are responsible for recollection rejection of distractors versus acceptance on the basis of similarity judgements are dissociated, both when they are initially stored and when they are subsequently retrieved (e.g., Brainerd et al., 2003; Reyna & Brainerd, 1995). However, a one-process interpretation is also possible: When a semantically related distractor is presented, recollection rejection might occur when weak support for the probe accumulates from memory, but similarity judgement might occur when much stronger support *of the same sort* accumulates from memory. A standard method of evaluating such dual-process versus one-process interpretations of the parameters memory models is to compute Dunn and Kirsner's (1988) reversed association test (for examples, see Brainerd, Wright, Reyna, & Payne, 2002). This test can rule out one-process interpretations for experimental designs in which there are more than two treatment conditions, so that there are more than two possible pairings of the values of the parameters of interest (recollection rejection and similarity judgement in this case). As Dunn and Kirsner showed, under a one-process interpretation, simple monotonic-increasing or monotonic-decreasing curves must result when values of one parameter, such as S_D, are plotted against increasing values of another parameter, such as R. A one-process interpretation can be rejected if nonmonotonic curves are obtained (i.e., the value of S_D sometimes increases and sometimes decreases as the value of R increases).

With the design of this experiment, there are three different data sets in which the reversed association test can be used to pit the dual-process and one-process hypotheses against each other: the data of the immediate session, the data of the delayed session for previously untested narratives, and the data of the delayed session for previously tested narratives. Within each of these data sets, the combination of the order-of-narrative-presentation factor (first versus second) and the narrative-repetition factor (once versus twice) yields four distinct treatment combinations for which values of R and S_D can be separately paired and plotted. To increase the reliability of the parameter estimates and the power of reversed associate test, we computed mean estimates of R and S_D for the three gist-consistent distractors (TPN, TIO, TIN) for each of these treatment combinations, and plotted the pairs of mean estimates. The results are displayed in Figure 1, with the reversed association tests for the immediate session, the delayed session for previously untested narratives, and the delayed session for previously tested narratives appearing in panels A, B, and C, respectively. There, it can be seen that the data ran consistently against the one-process hypothesis. In all three plots, the relation was nonmonotonic, with S_D sometimes increasing and sometimes decreasing as R increased.

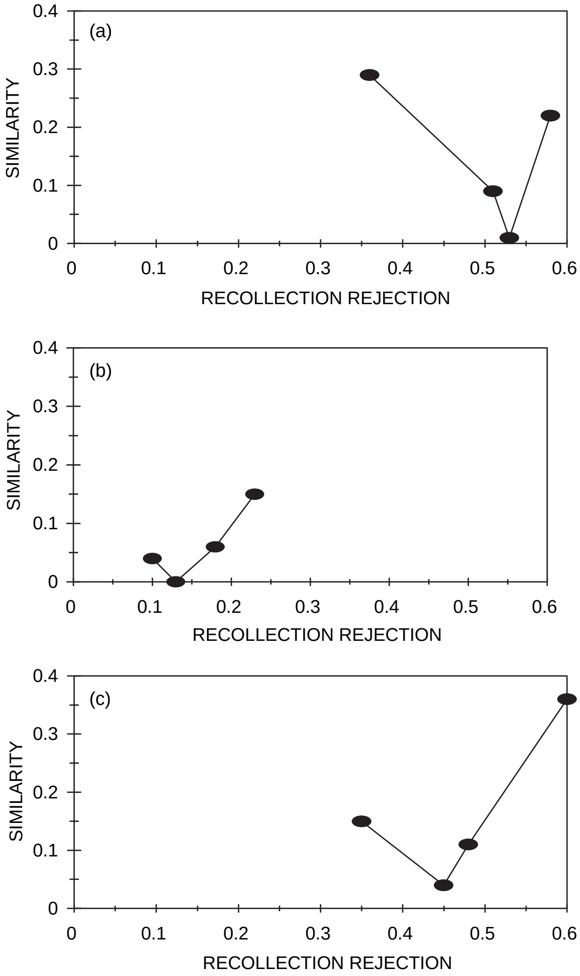

Figure 1. Relations between average probabilities of recollection rejection and similarity on the immediate test (Panel A), on the 1-week test for previously untested narratives (Panel B), and on the 1-week test for previously tested narratives (Panel C).

The final issue is whether, in this narrative paradigm, false-memory suppression was operating at the item level, as the notion of recollection rejection assumes, with false alarms to individual distractors being suppressed via the retrieval of verbatim traces of their corresponding targets, or whether verbatim-based suppression was operating primarily at a metacognitive level. According to the notion of metacognitive suppression (e.g., Ghetti, Qin, & Goodman, 2002), certain methods of presenting target materials create an expectation in participants that they will be able to vividly recollect the surface features of those materials on memory tests. That expectation, which has been termed the distinctiveness heuristic (Schacter, Israel, & Racine, 1999), causes participants to reject memory probes, whether targets or distractors, that are not accompanied by mental reinstatement of realistic presentation details. The empirical signature of the distinctiveness heuristic is that it operates across the board for all probes, not on an item-by-item basis for distractors only. To generate the distinctiveness heuristic, it is important that target presentation methods be manipulated between- rather than within-condition, so that participants can develop a general expectation of vivid phenomenology that can be applied across the board on memory tests. Examples of methods that seem to induce this expectation are presenting word lists as pictures (rather than as printed words), presenting word lists as printed words (rather than pronouncing them orally), and presenting word lists repeatedly (rather than presenting them only once)(e.g., Dodson, & Schacter, 2001, 2002a,b). Within- versus between-condition manipulation of such variables provides differential evidence as to whether false-memory suppression is only due to the distinctiveness heuristic, only due to recollection rejection, or due to a mixture of these processes: If a variable is manipulated both within- and between-condition, the first scenario is supported if only between-condition manipulation produces suppression; the second scenario is supported if within- and between-condition manipulation produce comparable amounts of suppression; and the third scenario is supported if between-condition manipulation produces greater suppression than within-condition manipulation (Brainerd & Reyna, 2005).

Available evidence suggests that in word-list tasks, both recollection rejection and the distinctiveness heuristic contribute to false-memory suppression (e.g, Arndt & Reder, 2003; Hege & Dodson, 2004). In the present narrative task, on the other hand, suppression seems to have been almost wholly due to recollection rejection. Two key findings should be considered in this connection. First, the design contained three variables that were expected to affect recollection rejection by influencing the accessibility of verbatim traces and that were manipulated within condition: recency of narrative presentation, type of gist-consistent distractor, and repeated testing. Because these variables could not create condition-wide differences in expectation of vivid phenomenology, their observed effects, while consistent with recollection rejection, cannot be explained via the operation of the distinctive heuristic. Second, the design also contained another variable, narrative repetition, that was expected to affect recollection rejection by influencing the accessibility of verbatim traces, but that was manipulated both within and between condition. This allowed us to evaluate, for repetition at least, which of the above three scenarios was supported. In the event, it was the second scenario (recollection rejection without the distinctiveness heuristic) that received support because although narrative repetition suppressed false alarms to gist-consistent distractors, suppression was comparable for within- versus between-condition manipulation. Obviously, this outcome is compatible with the overall picture of excellent verbatim memory for narrative statements, which is a precondition for robust recollection rejection.

REFERENCES

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language*, *48*, 1–15.

Brainerd, C. J., & Mojardin, A. H. (1998). Children's false memories for sentences: Long-term persistence and mere testing effects. *Child Development*, *69*, 1361–1377.

Brainerd, C. J., & Reyna, V. F. (1996). Mere memory testing creates false memories in children. *Developmental Psychology*, *32*, 467–476.

Brainerd, C. J., & Reyna, V. F. (2002). Recollection rejection: How children edit their false memories. *Developmental Psychology*, *38*, 156–172.

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. Oxford, UK: Oxford University Press.

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

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*, 762–784.

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

Brainerd, C. J., & Wright, R. (in press). Forward association, backward association, and the false-memory illusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 554–567.

Brainerd, C. J., Wright, R., Reyna, V. F., & Payne, D. G. (2002). Dual-retrieval processes in recall. *Journal of Memory and Language*, *46*, 120–152.

Brown, V. R., & Gorfein, D. S. (2004). A new look at recognition in the Brown-Peterson distractor paradigm: Toward the application of new methodology to unsolved problems of recognition memory. *Memory & Cognition*, *32*, 674–685.

Budson, A. E., Sullivan, A. L., Mayer, E., Daffner, K. R., Black, P. M., & Schacter, D. L. (2002). Suppression of false recognition in Alzheimer's disease and in patients with frontal lobe lesions. *Brain*, *12*, 2750–2765, Pt. 125.

Dodson, C. S., & Schacter, D. S. (2001). "If I said it, I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155–166.

Dodson, C. S., & Schacter, D. L. (2002a). Aging and strategic retrieval processes: Reducing false memories with a distinctiveness heuristic. *Psychology and Aging*, *17*, 405–415.

Dodson, C. S., & Schacter, D. S. (2002b). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language*, *46*, 782–803.

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

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 120–128.

Gerkens, D. R., & Smith, S. M. (2004). Effects of perceptual modality on verbatim and gist memory. *Psychonomic Bulletin & Review*, *11*, 143–149.

Ghetti, S., Qin, J., & Goodman, G. S. (2002). False memories in children and adults: Age, distinctiveness, and subjective experience. *Developmental Psychology*, *38*, 705–718.

Hege, A. C. G., & Dodson, C. S. (2004). Why distinctive information reduces false memories: Evidence for both impoverished relational-encoding and distinctiveness heuristic accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 787–795.

Hu, X. (1995). *Statistical inference program for multinomial binary tree models* (Version 1) [Computer software]. Irvine, CA: University of California at Irvine.

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

Karibian, D. (2003). *Using fuzzy-trace theory to predict memory for word lists and sentences in children with learning disabilities*. Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of aging. *Cognitive Neuropsychology*, *16*, 399–415.

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

Lampinen, J. M., Odegard, T. N., Backshear, E., & Toglia, M. P. (2005). Phantom ROC. In F. Columbus (Ed.), *Progress in experimental psychology research*. Hauppauge, NY: Nova.

Lampinen, J. M., Odegard, T. N., & Neuschatz, J. S. (2004). Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 332–342.

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition*, *22*, 85–94.

Odegard, T. N., & Lampinen, J. M. (in press). Recollection rejection: Gist cuing of verbatim memory. *Memory & Cognition*.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*, 261–285.

Reyna, V. F. (2000). Fuzzy-trace theory and source monitoring: An evaluation of theory and false-memory data. *Learning and Individual Differences*, *12*, 163–175.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75.

Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, *30*, 178–191.

Reyna, V. F., & Kiernan, B. (1995). Children's memory and interpretation of psychological metaphors. *Metaphor and Symbolic Activity*, *10*, 309–331.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences*, *9*, 95–123.

Reyna, V. F., Mills, B., Estrada, S., & Brainerd, C. J. (in press). False memory in children: Data, theory, and legal implications. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *Handbook of eyewitness psychology*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Reyna, V. F., & Titcomb, A. L. (1997). Constraints on the suggestibility of eyewitness testimony: A fuzzy-trace theory analysis. In D. G. Payne & F. G. Conrad (Eds.), *A synthesis of basic and applied approaches to human memory*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Rotello, C. M. (2001). Recall processes in recognition memory. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 40, pp. 183–221). San Diego, CA: Academic Press.

Runquist, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, *11*, 641–650.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1–24.

Warren, A. R., & Lane, P. L. (1995). Effects of timing and type of questioning on eyewitness accuracy and suggestibility. In M. S. Zaragoza, J. R. Graham, N. N. Hall, R. Hirschman, & Y. S. Ben-Porath (Eds.), *Memory, suggestibility, and eyewitness testimony in children and adults* (pp. 44–60). Thousand Oaks, CA: Sage.