

Running Head: OVER-DISTRIBUTION

Over-Distribution in Source Memory

C. J. Brainerd and V. F. Reyna

Cornell University

R. E. Holliday

University of Leicester

K. Nakamura

Cornell University

Abstract

In semantic false memory, conventional false memory responses are confounded with a second type of error, over-distribution, in which items are attributed to logically contradictory episodic states. Over-distribution errors have proved to be more common than false memories when the two are disentangled. We investigated whether over-distribution is prevalent in another classic false memory paradigm: source monitoring. It is. As in semantic false memory, false memory errors (source misattributions) were predominantly over-distribution errors, but unlike semantic false memory, over-distribution also accounted for more than half of true memory responses (correct source attributions). Good experimental control of over-distribution was achieved via a series of manipulations that were designed to affect either recollection of contextual details or item memory (concreteness, frequency, list-order, number of presentation contexts, and individual differences in verbatim memory). The model that was used to analyze the data (conjoint process dissociation) specifies that over-distribution is directly proportional to item memory and inversely proportional to recollection. Results for recollection-slanted and item-slanted manipulations were consistent with that prediction. Moreover, item memory was not a necessary precondition for recollection of contextual details, and indeed, conditions that produced higher levels of item memory produced lower levels of recollection.

Keywords: over-distribution, false memory, source monitoring, item memory, recollection, word frequency

We report evidence that a new form of memory distortion that has been identified with semantic false memory tasks occurs in another classic paradigm, source monitoring. In this form of distortion, called over-distribution, memory attribution is promiscuous with respect to the perceived episodic scope of experience. Specifically, retrieval makes past experience seem broader than it is by attributing events to too many contexts, even to the point of attributing them to logically contradictory contexts. Importantly, over-distribution cuts across the types of responses that researchers customarily regard as instances of false memory (false alarms to distractors that preserve the meaning of targets, such as *Coke* when *Pepsi*, *7-Up*, and *Sprite* are targets). In prior research (Brainerd & Reyna, 2008; Brainerd, Reyna, & Aydin, 2010), over-distribution proved to be so ubiquitous that it accounted for more than half of the responses that would otherwise have been classified as false memories. Over-distribution also cuts across true memories (hits to targets such as *Pepsi* or *7-Up*), so that memory for actual experience is less true than it seems.

In that prior research, we focused on logically contradictory instances of over-distribution wherein retrieval treats meaning-preserving distractors as having been both presented and not presented on a study list and does likewise with targets. Initially, we reviewed over 100 sets of previously published data whose designs allowed over-distribution errors to be separated from false memories (Brainerd & Reyna, 2008). Overall, 18% of targets and 13% of meaning-preserving distractors were remembered as having been both presented and not presented on study lists. Later, we reported some experiments that generated 88 new data sets in which over-distribution and false memory could be separated and compared under controlled conditions (Brainerd, Reyna, & Aydin, 2010). Overall, 14% of targets and 19% of meaning-preserving distractors were remembered as having been both presented and not presented. Only 9% of those same distractors produced false memories, so that over-distribution swamped false memory. This was even true with tasks that have been found to produce exceptionally high levels of false memory in conventional designs, such as the Deese/Roediger/McDermott illusion (Deese, 1959; Roediger & McDermott, 1995).

As discussed in these articles, over-distribution cannot be disentangled from false memory with traditional designs, in which subjects only make old/new judgments about test cues. This can be seen in Table 1, where the operational distinction between over-distribution and false memory is illustrated for the

distractor *Coke* when subjects must accept/reject three statements about it: (a) It's a target. (b) It's new but related to a target. (c) It's either a target or new but related to a target. False memory for *Coke* means that subjects remember it as having been on the study list—as being a target rather than a related distractor. If so, the first and third statements will be accepted, and the second will be rejected, which is the pattern in the first column of Table 1. Over-distribution, on the other hand, means that *Coke* is remembered as being a target and a related distractor, although this is contradictory. If so, all three statements will be accepted. This is the pattern in the second column of Table 1.

Table 1 also illustrates that the reason that over-distribution cannot be disentangled from false memory when subjects only make old/new judgments about meaning-preserving distractors is that, assuming appropriate corrections for response bias, acceptance of “It's a target” does not guarantee rejection of “It's new but related to a target.” Thus, when *Coke* is presented as a cue for an old/new judgment in a standard design, an “old” response may mean false memory, or it may mean over-distribution. The procedure, in Table 1, which separates these possibilities, is a memory analogue of a methodology that is used to study disjunction fallacies in probability judgment (e.g., Brenner & Rottenstreich, 1999; Fox & Tversky, 1998; Rottenstreich & Tversky, 1997; Sloman, Rottenstreich, Wisniewski, Hadjichristidis, & Fox, 2004; Tversky & Koehler, 1994), and surprisingly, theoretical distinctions that are used to explain disjunction fallacies (e.g., Barrouillet, in press; Reyna & Brainerd, in press) also predict over-distribution (see below). The procedure in question, conjoint recognition, follows the customary design of semantic false memory experiments, with one addition that occurs on memory tests (see Brainerd, Reyna, & Mojardin, 1999; Lampinen, Odegard, Blackshear, & Togliani, 2005; Odegard & Lampinen, 2005; Singer & Remillard, 2008; Stahl & Klauer, 2008, 2009). The three standard types of test cues (targets, related distractors, unrelated distractors) are factorially crossed with the three types of probes in Table 1, which are denoted T (for “target”), R (for “related distractor”), and TUR (for “target or related distractor”), respectively. Only one probe is administered per cue to individual subjects, in order to avoid repeated testing effects, but across subjects, all three probes are administered for each cue.

Thus, independent estimates of three quantities are available for *each test cue*: (a) $P(T)$, the probability that it is judged to be a target; (b) $P(R)$, the probability that it is judged to be a related distractor; and (c) $P(TUR)$, the probability that it is judged to be either a target or a related distractor. (For

simplicity, we assume that each quantity has been corrected for response bias, using unrelated distractor data.) Episodic over-distribution can now be separated from false memory by relying on the familiar rule of probability theory that for any two events T and R, the sum of their probabilities equals the probability of their disjunction plus the probability of their conjunction; that is, $P(T) + P(R) = P(T \cup R) + P(T \cap R)$. As a matter of logic, $P(T \cap R)$ must be zero because, of course, a cue cannot be both present in and absent from a set of studied targets. As a matter of memory, however, it is conceivable that subjects may accept T or R probes because the retrieved information about a cue is consistent with *both episodic states*, rather than just one. If so, $P(T \cap R) > 0$. It then follows from the above rule that the relation between empirical estimates of the quantities $P(T)$, $P(R)$, and $P(T \cup R)$ will be *subadditive*—that is, $P(T) + P(R) > P(T \cup R)$ —because according to the rule, $P(T) + P(R) - P(T \cap R) = P(T \cup R)$ and we know that $P(T \cap R) > 0$. Hence, for a related distractor such as *Coke*, the quantity $P(T \cap R) = P(T) + P(R) - P(T \cup R)$ indexes the level of over-distribution, while the quantities $P(T \cap \sim R) = P(T) - P(T \cap R)$ and $P(\sim T \cap R) = P(R) - P(T \cap R)$ index the levels of true and false memory, respectively. For targets such as *Pepsi* or *7-Up*, the quantities $P(T \cap R)$, $P(T \cap \sim R)$, and $P(\sim T \cap R)$ index over-distribution, false memory, and true memory, respectively.

Over-distribution is paradoxical inasmuch as it violates the law of the excluded middle. Nevertheless, Brainerd and Reyna (2008) showed that it is predicted by dual-trace distinctions that have been used to explain both semantic false memory and disjunction fallacies in probability judgment (see Reyna & Brainerd, 1995). In dual-trace accounts, related distractors (and targets) provoke the retrieval of two types of episodic representations—namely, verbatim traces of the presentation of specific targets (e.g. *Pepsi*, *7-Up*), which support rejection of T probes and acceptance of R and TUR probes, and gist traces of targets' semantic content (e.g., soft drink). Depending on their strength, gist traces can support either false memory (acceptance of T and TUR probes, coupled with rejection of R probes) or over-distribution (acceptance of all three probes). When such traces are particularly strong, they produce an ersatz recollective phenomenology in which distractor cues' prior "presentations" flash in mind's eye and echo in the mind's ear (e.g., Heaps & Nash, 2001; Stahl & Klauer, 2009). This phenomenology is consistent with such cues being targets but not related distractors and, hence, is a basis for false memory. Weaker gist traces produce a generic phenomenology, usually called semantic similarity, which is consistent with both episodic states and, hence, is a basis for over-distribution. Because target cues

provoke retrieval of the same two types of traces, these distinctions also predict over-distribution for targets (Brainerd & Reyna, 2008).

Brainerd et al. (2010) formalized these distinctions in a mathematical model. Across their experiments, over-distribution was measured in two ways: (a) with bias-corrected values of the quantity $P(T \cap R)$ and (b) with the parameters of the model. Over-distribution was detected with both indexes in all experiments, but the more sensitive modeling instrument delivered stronger evidence. The dual-trace account was tested with a series of manipulations, whose logic was that the manipulations should increase the chances that probe acceptances would be based on gist retrieval. Examples include immediate versus delayed testing and emotional valence of study materials and test cues. As expected, over-distribution increased as a function of manipulations that favored gist retrieval.

The Present Experiments

Predicting Over-Distribution of Source Memories

Semantic false memory is one the two most commonly studied varieties of false memory, the other being source monitoring errors (for a review, see Brainerd & Reyna, 2005). The latter differ from the former in that (a) each target is presented in one of two or more distinct contexts (say, on List 1 rather on List 2) and (b) a false memory consists of remembering a *target* as having been presented in the wrong context (e.g., on List 2 rather than on List 1). The influential Loftus (1975; Loftus, Miller, & Burns, 1978) misinformation paradigm, for instance, is a special case of this procedure (Johnson, Hashtroudi, & Lindsay, 1993).

The aim of our research was to evaluate the prediction that over-distribution ought to occur in source monitoring as well as in semantic false memory. In that connection, Brainerd and Reyna (2008) showed that such a prediction falls out of the theoretical distinctions that lie behind Jacoby's (1991) process-dissociation model, which was originally implemented in a two-list source-monitoring task. In that task, subjects study two word lists in which items are presented in distinctive formats (e.g., as visual anagrams on List 1 or as spoken words on List 2), with each item appearing on only one list. Subjects then respond to a test list composed of List 1 targets, List 2 targets, and unrelated distractors under either inclusion or exclusion instructions. Inclusion instructions tell subjects to accept any word that was presented on either list and to reject distractors. Exclusion instructions tell subjects to accept words that were heard (List 2)

but to reject words that were seen (List 1), along with distractors. Instructions also emphasize that target words were seen or heard but not both.

The process dissociation model was defined over acceptance probabilities for List 1 items, as follows:

$$p_{I,L1} = R + (1-R)I \text{ and} \quad (1)$$

$$p_{E,L1} = (1-R)I, \quad (2)$$

where $p_{I,L1}$ is the List 1 inclusion hit probability, $p_{E,L1}$ is the List 1 exclusion false alarm probability, R is the probability that a List 1 word's presentation can be recollected, and I is the probability that a List 1 word whose presentation cannot be recollected is so familiar that it is accepted anyway. The second parameter is denoted as I , rather than as F (Jacoby's original notation), because as Buchner, Erdfelder, Steffens, and Martensen (1997) demonstrated, this parameter measures a process that is called *item memory* in the source-monitoring literature; that is, being able to remember that a word was presented without being able to recollect contextual details that specify which list it was presented on. The quantities $p_{I,L1}$ and $p_{E,L1}$ do not provide sufficient degrees of freedom to separate over-distribution from true and false memory for source, but Brainerd and Reyna (2008) showed that separation can be effected by adding a third test condition, which is called mirror exclusion. In the mirror exclusion condition, instructions tell subjects to accept words that were seen (List 1) but to reject words that were heard (List 2), along with distractors. Acceptances of List 1 words are hits, and the model's expression for this hit rate is

$$p_{M,L1} = R + (1-R)I. \quad (3)$$

We now have independent estimates of three quantities that parallel those that measure over-distribution in semantic false memory: (a) $P(L1 \cup L2 | L1) = p_{I,L1}$ is the probability that a List 1 item is accepted as having been on List 1 or List 2; (b) $P(L2 | L1) = p_{E,L1}$ is the probability that a List 1 item is accepted as having been on List 2, which subjects know also means that it was not on List 1; and (c) $P(L1 | L1) = p_{M,L1}$ is the probability that a List 1 item is accepted as having been on List 1, which subjects know also means that it was not on List 2. The over-distribution probability $P(L1 \cap L2 | L1)$ is obtained as before, via the rule $P(L1 | L1) + P(L2 | L1) = P(L1 \cup L2 | L1) + P(L1 \cap L2 | L1)$. Note that as in semantic false memory, it is objectively true that $P(L1 \cap L2 | L1) = 0$ because List 1 words never appear on List 2, and

subjects are instructed on that fact. However, the process dissociation model predicts $P(L1 \cap L2 | L1) > 0$. On this point, some algebraic manipulation of Equations 1-3 reveals that $P(L1 \cap L2 | L1) = p_{E,L1} + p_{M,L1} - p_{I,L1} = (1-R)I$. In other words, the over-distribution probability is just the probability that subjects have item memory for words whose presentation details cannot be recollected. The proposed mechanism whereby item memory produces over-distribution runs as follows (Brainerd & Reyna, 2008). When subjects are confronted with a source probe for a cue that they remember studying (e.g., *bagpipe*) but for which contextual details of its presentation cannot be recollected, the only residual source information is that which is stated in probe itself (e.g., List 1). That information is therefore relied upon, in lieu of recollection of contextual details, as a basis for source judgment. However, this means that probes such as “*Bagpipe* was on List 1” and “*Bagpipe* was on List 2” will both be accepted, resulting in over-distribution.

Although the process dissociation model predicts over-distribution of source memories, it has an important limitation that is revealed by the extended, three-condition design. Strictly speaking, it does not allow for the possibility of false memory for source; that is, remembering List 1 items as having been presented on List 2 but not on List 1 (see also, Reyna & Titcomb, 1997). To see why, consider that the generic expression for this false memory probability would be $P(\sim L1 \cap L2 | L1) = P(L2 | L1) - P(L1 \cap L2 | L1)$. Under the model, $P(L2 | L1) = R(1-I)$ and $P(L1 \cap L2 | L1) = R(1-I)$, so that $P(L2 | L1) - P(L1 \cap L2 | L1) = 0$. Brainerd and Reyna (2008) showed this limitation can be removed by combining the process dissociation model with the conjoint recognition model to yield a hybrid model, which will be termed conjoint process dissociation (CPD) to distinguish it from the parent models. The conjoint recognition model contains a third parameter, E , which indexes false memory for targets (see Brainerd, Reyna, Wright, & Mojardin, 2003). When the two models are integrated, the resulting expressions for the three conditions of the extended design are

$$p_{I,L1} = R_1 + (1-R_1)E_2 + (1-R_1)(1-E_2) I_1, \quad (4)$$

$$p_{E,L1} = (1-R_1)E_2 + (1-R_1)(1-E_2)I_1, \text{ and} \quad (5)$$

$$p_{M,L1} = R_1 + (1-R_1)(1-E_2)I_1, \quad (6)$$

where R_1 is the probability that a List 1 cue prompts retrieval of List 1 contextual details, E_2 is the probability that a List 1 cue prompts retrieval of List 2 contextual details, and I_1 is the probability that a List 1 cue prompts item memory.

Thus, CPD, which was used in our research, allows for all four of the possible forms of source memory for a List 1 item in this design: true memory, $L1 \cap \sim L2 \mid L1$; false memory, $\sim L1 \cap L2 \mid L1$; over-distribution, $L1 \cap L2 \mid L1$; and forgetting, $\sim L1 \cap \sim L2 \mid L1$. When the three CPD parameters are estimated for experimental data, exact probabilities of each type of source memory can be predicted because, under the model, $P(L1 \cap \sim L2 \mid L1) = R_1$, $P(\sim L1 \cap L2 \mid L1) = (1-R_1)E_2$, $P(L1 \cap L2 \mid L1) = (1-R_1)(1-E_2)I_1$, and $P(\sim L1 \cap \sim L2 \mid L1) = (1-R_1)(1-E_2)(1-I_1)$. Any three of these conditional probabilities fully determines the fourth, of course, because they must sum to unity.

Over-Distribution from Source Guessing?

We also investigated a second model that can accommodate over-distribution in some circumstances—namely, Batchelder and Riefer's (1990) source-memory model.¹ As an approach to over-distribution, this model differs from CPD in two key respects. First, it does not *predict* over-distribution, as CPD does, but rather, it permits this pattern under some values of its parameters. Under other values, the source-memory model permits the other two logically possible patterns: under-distribution (i.e., $[P(L1 \mid L1) + P(L2 \mid L1)] < P(L1 \cup L2 \mid L1)$) and additivity (i.e., $[P(L1 \mid L1) + P(L2 \mid L1)] = P(L1 \cup L2 \mid L1)$). Second, the source-memory model localizes the cause of over-distribution in a different type of process than CPD, that process being source guessing in the absence of item memory.

Relative to the memory tasks that were described above, in which subjects merely make yes/no source judgments about test cues, Batchelder and Riefer (1990) defined their model over a task in which subjects first make old/new recognition judgments about cues and then make source judgments about them (List 1 or List 2?). For that type of task and for a given source, such as List 1, the model contains a total of six free parameters (denoted D_1 , d_1 , D_N , g , a , and b). For the simpler task that we described, only five parameters are required (D_1 , d_1 , D_N , g , and b). Their process definitions are provided in Table 2. These parameters can be used to generate parallel expressions for Equations 4-6 that will accommodate over-distribution under some parameter values. Again assuming appropriate corrections for response bias (which involve the parameters D_N and b), the parallel expressions of the source-memory model involve three of the parameters in Table 2: D_1 , the probability that subjects have source memory for List 1 cues; d_1 , the probability that subjects can recollect that List 1 cues for which they have item memory were presented on List 1; and g , the probability that subjects guess that List 1 cues were presented on

whichever list is mentioned in the test probe (List 1? List 2?) when they do not have item memory for those cues. The parallel expressions for Equations 4-6, then, are:

$$p_{I,L1} = D_1, \quad (7)$$

$$p_{E,L1} = D_1(1-d_1)g, \text{ and} \quad (8)$$

$$p_{M,L1} = D_1d_1 + D_1(1-d_1)g. \quad (9)$$

Verbally, Equation 7 says that the response to inclusion probes will be “yes” for all List 1 cues for which subjects have item memory, Equation 8 says the response to exclusion probes will be “yes” when subjects guess that List 2 is the source of List 1 cues for which they have item memory but not source recollection; and Equation 9 says that the response to mirror exclusion probes will be “yes” (a) for List 1 cues for which subjects have item memory and source recollection and (b) when subjects guess that List 1 is the source of List 1 cues for which they have item memory but not source recollection. In these expressions, whether or not over-distribution occurs depends on the numerical value of the source guessing parameter g . According to these expressions, $P(L1UL2 | L1) = D_1$ and $P(L1 | L1) + P(L2 | L1) = D_1[d_1 + 2g(1-d_1)]$. Thus, regardless of the values of D_1 and d_1 , over-distribution will occur if $g > .5$ because, in that event, $D_1[d_1 + 2g(1-d_1)]$ must be larger than D_1 . However, if $g < .5$, the opposite pattern, under-distribution, will occur because $D_1[d_1 + 2g(1-d_1)]$ must be smaller than D_1 . Finally, if $g = .5$, the source guessing parameter vanishes from Equations 8 and 9 (because $2g = 1$), and the relation between $P(L1 | L1) + P(L2 | L1)$ and $P(L1UL2 | L1)$ will be additive because $P(L1 | L1) + P(L2 | L1) = D_1[d_1 + 1 - d_1] = D_1$.

At the level of retrieval processes, the difference between the CPD and source-memory model accounts of over-distribution is this. On the one hand, CPD localizes the cause of over-distribution within item memory, positing that over-distribution will occur for cues for which subjects have item memory when subjects cannot recollect the cues' sources. On the other hand, the source-memory model localizes over-distribution within a guessing process, positing that for cues for which subjects do *not* have item memory, over-distribution occurs when subjects tend to guess the sources that are stipulated in exclusion and mirror exclusion probes at high rates (i.e., more than half of the time). If subjects do not guess stipulated sources at high rates, however, under-distribution rather than over-distribution will result. The relative fits of these models were compared in all of the experiments that we report.

Overview of Experiments

We investigated over-distribution of source memories in the extended process dissociation design. One objective was to determine whether there is compelling evidence that over-distribution occurs in source monitoring as well as in semantic false memory. An equally important aim was to determine whether over-distribution can be brought under experimental control by theoretically-motivated manipulations. A third objective was to determine whether the CPD model yields good fits to the data of the extended process dissociation design, which would permit direct measurement of underlying theoretical processes.

In all of the present experiments, subjects studied mutually exclusive lists of words that were presented in physically distinctive contexts (different colors, fonts, and temporal positions), which served as the basis for source discrimination. Within and between experiments, certain factors were manipulated that the CPD model and prior experimentation suggest would be apt to influence over-distribution of targets to mutually incompatible contexts. With respect to the CPD model, remember that the over-distribution probability for a List 1 target is $(1-R_1)(1-E_2)I_1$ —so that at the level of retrieval processes, over-distribution is directly proportional to item memory, inversely proportional to recollection of the contextual details of List 1 (e.g., yellow-script-first), and inversely proportional to recollection of the contextual details of List 2 (e.g., pink-Broadway-second). In prior reviews of the source monitoring literature, Reyna and associates (Reyna, 1995; Reyna & Lloyd, 1997; Reyna & Titcomb, 1997) concluded that there is considerable evidence that manipulations that affect verbatim memory for the surface form of target presentations affect recollection of contextual details, which should affect levels of over-distribution under the CPD model. We included examples of such manipulations in all of our experiments.

Those manipulations are word concreteness, individual differences in recollective ability, and temporal order of presentation contexts. The prediction in each instance was that the R parameter of CPD should be affected. Concerning concreteness, it has long been understood that recognition accuracy is better for concrete words than for abstract ones (e.g., Gorman, 1961; Paivio, 1969). With reference to the R parameter, experimentation spanning several years points to the conclusion that it is easier to link concrete words to contextual details of their presentation than it is to link abstract words to such details (for a recent review, see Madan, Glaholt, & Caplan, 2010). Further, in source-monitoring

research, recollection of the contextual details of item presentations has been found to increase as word concreteness increases (e.g., Vogt, & Broder, A, 2007). With respect to individual differences, research on semantic false memory has identified clinical populations that display reduced verbatim memory for targets, coupled with spared gist memory for their semantic content (e.g., Budson et al., 2006). In research with nonclinical subjects, we have compared the performance of individuals drawn from different universities on word-list tasks that measured both verbatim memory for target presentations and gist memory for semantic content. We identified populations that display different levels of verbatim memory for word lists, coupled with similar levels of memory for the semantic content of those same lists (Brainerd, Yang, Toglia, Reyna, & Stahl, 2008). In the present research, we compared levels of over-distribution of source memories in subjects who were drawn from two populations (University P and University Q) that are known to differ in this way, on the hypothesis that over-distribution would be more likely to be detected in one than the other. Concerning temporal order of presentation contexts, levels of over-distribution can be compared for the earlier (List 1) versus later (List 2) presentation context in the extended process dissociation design. It is well known in the false memory literature that verbatim memory for targets is sensitive to retroactive interference from subsequently presented material (Barnhardt, Choi, Gerken, & Smith, 2006; Payne, Elie, Blackwell, & Neuschatz, 1996). Thus, in source monitoring, recollection of presentation details should be better for later (List 2) than for earlier (List 1) contexts, which means more over-distribution for earlier contexts.

We also included a fourth manipulation, word frequency, that was intended to increase item memory and, therefore, to increase over-distribution. A classic finding (e.g., Hall, 1979) that has often been studied is that old/new recognition is better for low- than for high-frequency words. Some researchers have explained this finding on the ground that low-frequency words induce higher levels of familiarity on recognition tests (e.g., Glanzer & Adams, 1990; Glanzer, Kim, Hilford, & Adams, 1999), others on the ground that low-frequency words induce higher levels of recollection (e.g., Arndt & Reder, 2002; Park, Arndt, & Reder, 2006). In source-monitoring research, the effects of word frequency have not been extensively investigated, but it has been proposed that recollection of the contextual details of item presentations increases as word frequency decreases (Marsh, Cook, & Hicks, 2006). Contrary to that proposal, we discovered in a pilot study that preceded the present experiments that low-frequency words

increased item memory, relative to high-frequency words, while having little effect on recollection. In that study, 48 undergraduates studied two lists of words, each composed of 20 high-frequency and 20 low-frequency items. To ensure that the lists were distinctive, the words on List 1 were presented in a different color and font than the words on List 2. The subjects then responded to a 120-item source recognition test composed of 30 List 1 cues (15 high-frequency, 15-low frequency), 30 List 2 cues (15 high-frequency, 15-low frequency), and 60 distractor cues (30 high-frequency, 30-low frequency). These test cues were factorially crossed with three yes/no source probes: Presented on List 1? Presented on List 2? Presented on List 1 or List 2? The results are displayed in Figure 1. In Panel A, bias-corrected acceptance probabilities are exhibited for correct probes (which asked whether a cue was presented on the list on which it in fact was presented) and for wrong probes (which asked whether a cue was presented on the list on which it was *not* presented). It can be seen that correct probes were accepted at a higher rate for low- than for high-frequency words, which might seem to suggest better recollection of source details for low-frequency words. However, wrong probes were also accepted at a higher for low-frequency words, which suggests the opposite. Estimates of the CPD model's parameters for these data are exhibited in Panel B. It can be seen that item memory was notably stronger for low- than for high-frequency words, but frequency had little effect on recollection of either correct or incorrect contextual details. Therefore, in order to have a variable with an item-memory slant, word frequency was manipulated in all of our experiments.

Beyond the concreteness, individual differences, list order, and frequency variables, we studied a fifth variable that is discussed in greater detail in Experiments 3 and 4: number of presentation contexts (lists). In those experiments, we extended the standard two-list design to three lists, *while holding the number of presented targets constant*. The logic of this manipulation runs as follows. Although overall memory load, in the sense of how many items must be remembered, remains constant, increasing the number of contexts in which they are presented increases the number of opportunities for over-distribution. In a two-list design, there is only one way to over distribute a List 1 (or List 2) target—namely, if retrieval attributes it to both lists. Thus, over-distribution is indexed by the quantity $P(L1 \cap L2 | L1)$. In a three-list design, however, there are three ways to over distribute the same target—namely, if retrieval attributes it to lists 1 and 2, to lists 1 and 3, and to all three lists. Now, over-distribution

is the sum of three quantities, $P(L1 \cap L2 \cap \sim L3 | L1) + P(L1 \cap \sim L2 \cap L3 | L1) + P(L1 \cap L2 \cap L3 | L1)$. As we show later, if the values of the CPD model's parameters are roughly the same with two and three lists, as they might be considering that the number of presented targets is the same, over-distribution will be greater with three lists because $P(L1 \cap L2 \cap \sim L3 | L1) + P(L1 \cap \sim L2 \cap L3 | L1) + P(L1 \cap L2 \cap L3 | L1)$ will be greater than $P(L1 \cap L2 | L1)$.

Experiments 1 and 2

The extended process dissociation design was implemented in these experiments. In both, the subjects studied two non-overlapping lists that were presented in visually distinctive contexts and then responded to recognition tests on which some test cues were items that had been presented on List 1, some were items that had been presented on List 2, and some were distractors that were unrelated to presented items. To measure over-distribution, the three types of episodic descriptions (presented on List 1, presented on List 2, presented on List 1 or List 2) were factorially manipulated over the three types of test cues. To measure the effects of frequency and concreteness on over-distribution, these variables were factorially manipulated over targets and distractors in a 2 (frequency: high versus low) X 2 (concreteness: concrete versus abstract) design. To measure the effects of temporal position, during the study phase all of the words in one of the contexts were presented before any of the words in the other context. Finally, to measure the effects of individual differences, Experiment 1 was conducted with subjects from University P, and Experiment 2 was conducted with subjects from University Q.

Method

Subjects

The subjects in Experiment 1 were 59 introductory psychology students from University P, who participated in the experiment to fulfill a course requirement. The subjects in Experiment 2 were 60 introductory psychology students from University Q, who participated in the experiment to fulfill a course requirement. The experiments were run in parallel, at the same time of year, and the respective subject samples did not differ reliably in mean age or gender composition. University P is a public institution, and University Q is a private institution. The mean SAT score of students attending University P is roughly 200 points higher than the mean SAT score of students attending University Q. In our prior

research, when students from these universities have performed word list tasks, their gist memories have been comparable, but verbatim memory has been superior in University Q students.

Materials

A pool of 256 nouns was created, using the Kucera and Francis (1967) frequency norms and the Toglia and Battig (1978) concreteness norms. The pool was composed of four groups of words, each containing 64 items: (a) high-frequency/concrete nouns (e.g., *ocean, shoe*), (b) high-frequency/abstract nouns (e.g., *idea, moment*), (c) low-frequency/concrete nouns (e.g., *barnacle, hurdle*), and low-frequency/abstract nouns (e.g., *gist, rating*). Of the 128 high-frequency nouns, except for 8 words, all had frequencies of A or AA in printed text, and the remaining 8 had frequencies above 35 per million. For the 128 low-frequency nouns, the mean frequency was 1.78 per million, and no word had a frequency higher than 7 per million. Of the 128 concrete nouns, the mean concreteness rating (7-point scale) was 6.48, and no word had a rating lower than 5.66. Of the 128 abstract nouns, the mean concreteness rating (7-point scale) was 2.20, and no word had a rating higher than 3.23. The study list-contexts that were administered to individual subjects were constructed by randomly sampling 24 words items from each of the four Frequency X Concreteness sub-pools. The cue words on the test lists that were administered to individual subjects consisted of these 96 presented words, plus 96 distractors. The distractors were obtained by sampling a further 24 words from each of the Frequency X Concreteness sub-pools.

Each subject was exposed to two lists of words, with each list being presented in a distinctive visual context that was generated by presenting all of the words on each list in one of several distinctive fonts (e.g., Algerian, Broadway, script) against one of several background colors (e.g., yellow, white, pink). Thus, each list-context was defined by a combination of font and background color, such as List 1 = words presented in Algerian font against a yellow background and List 2 = words presented in Broadway font against a white background. All of the words in one visual context were presented before any of the words in the other context, so that temporal order was another list-specific cue.

During the study phase, a total of 108 words were presented, with 54 being presented in each list (order/font/color combination). Each list began with a three-word opening buffer and ended with a three-word closing buffer. These words were fillers that did not appear on the later memory test. The remainder of the list consisted of the 48 focal words, which were presented in random order and were

subdivided as follows: 12 high-frequency/concrete nouns, 12 high-frequency/abstract nouns, 12 low-frequency/concrete nouns, and 12 low-frequency/abstract nouns. Thus, over the 2 lists, subjects were exposed to a total of 24 exemplars of each of the 4 Frequency X Concreteness combinations.

During the test phase, 192 probes were presented in random order, with each probe consisting of a cue word and an episodic description of the word. There were two types of cue words, 96 presented items (24 high-frequency/concrete, 24 high-frequency/abstract, 24 low-frequency/concrete, 24 low-frequency/abstract) and 96 distractors (24 high-frequency/concrete, 24 high-frequency/abstract, 24 low-frequency/concrete, 24 low-frequency/abstract). There were three episodic descriptions: presented on List 1, presented on List 2, and presented on List 1 or List 2. These descriptions were manipulated factorially over the cues words, so that 64 probes asked whether the cue word had been presented on List 1 (32 presented items, 8 exemplars of each Frequency X Concreteness combination, and 32 distractors, 8 exemplars of each Frequency X Concreteness combination), 64 probes asked whether the cue word had been presented on List 2 (32 presented items, 8 exemplars of each Frequency X Concreteness combination, and 32 distractors, 8 exemplars of each Frequency X Concreteness combination), and 64 probes asked whether the cue word had been presented on List 1 or List 2 (32 presented items, 8 exemplars of each Frequency X Concreteness combination, and 32 distractors, 8 exemplars of each Frequency X Concreteness combination). The subjects accepted or rejected each probe, accordingly as they thought that the episodic description of the cue was true or false.

Procedure

At the start of the experiment, each subject received general memory instructions. He or she was informed that two different lists of words would be presented, one after the other, followed by a memory test. The two lists were then presented on a computer screen. Words were presented at a 2 sec rate, centered on the computer screen and printed in 72-point bold type. There was a 15 sec pause between lists. After the second list had been presented, the subject received instructions for the upcoming memory test. The instructions stated that half the test cues would be presented words and half would be distractors. The three types of episodic descriptions were defined and illustrated, and example probes with answers were provided, so that the subject understood how to respond to them. Subjects were instructed to accept a probe if they thought that the episodic description was true for the indicated cue

word and to reject it otherwise. The instructions stressed that the two lists did not overlap; that none of the words that were presented on one of the lists had been presented on the other and therefore if subjects could clearly recollect the appearance of a word in one context, it could not have appeared in the other. The 192 test probes were then presented in random order, with the subject responding in a self-paced manner.

Results and Discussion

The findings on episodic over-distribution are reported in two waves. First, we report some analyses of variance (ANOVAs) using the over-distribution metric that was mentioned earlier-- specifically, the statistic $P(L1 \cap L2 | L1) = P(L1 | L1) + P(L2 | L1) - P(L1 \cup L2 | L1)$ for words presented on List1 and the corresponding statistic $P(L1 \cap L2 | L2) = P(L1 | L2) + P(L2 | L2) - P(L1 \cup L2 | L2)$ for words presented on List 2. If these statistics have positive values in a given condition, over-distribution occurred in that condition. Second, we report findings that made use of the CPD model. As we know, this model contains the parameters R , I , and E , and it also contains bias parameters (see Appendix). In each condition, the amounts of true memory, false memory, over-distribution, and forgetting are provided by the quantities R , $(1-R)E$, $(1-R)(1-E)I$, and $(1-R)(1-E)(1-I)$, which are estimated separately for List 1 and List 2.

Results for the Over-Distribution Metric

The memory probes were constructed by factorially manipulating three types of episodic descriptions (presented on List 1, presented on List 2, presented on List 1 or List 2) over three types of cues (List 1 targets, List 2 targets, distractors). Thus, there were 9 acceptance probabilities (3 descriptions X 3 cues) for each of the 4 types of items on List 1 (2 levels of concreteness X 2 levels of frequency) and 9 acceptance probabilities for the same 4 types of items on List 2. For each list, we used the distractor acceptance probabilities to correct the acceptance probabilities for presented items for response bias, using the two-high threshold correction (Pr) of signal detection theory (Snodgrass & Corwin, 1988). Those corrected acceptance probabilities are displayed by condition in the upper half of Table 3, along with the corresponding over-distribution statistics, $P(L1 \cap L2 | L1)$ and $P(L1 \cap L2 | L2)$, for each condition.

Inspection of the results for Experiment 1 suggests five findings of importance. First, over-distribution was clearly in evidence as six of the eight values of this statistic in Table 3 are greater than zero. Second, frequency had the expected effect. The average value of the over-distribution statistic for low-frequency items was more than three times the value for high-frequency items (.24 versus .07). (We shall see below that this is because low-frequency items elevate the l parameter of CPD.) Third, the effects of concreteness and list were in the expected directions (more over-distribution for abstract items and for List 1) but were smaller than the effects of frequency. Fourth, concreteness interacted with list and list interacted with frequency: The concreteness effect was large and in the expected direction for List 2 but was negligible for List 1, and the list effect was large and in the expected direction for high-frequency items but was negligible for low-frequency items. Inspection of the corresponding results for Experiment 2 in Table 3 reveals the same patterns, plus a fifth finding: Levels of over-distribution were higher in Experiment 1 than in Experiment 2, as expected.

We computed a 2 (frequency: high versus low) X 2 (concreteness: concrete versus abstract) X 2 (list: 1 versus 2) repeated measures ANOVA for each experiment, using the subjects' scores on the over-distribution statistics $P(L1 \cap L2 | L1)$ and $P(L1 \cap L2 | L2)$ as dependent variables. In Experiment 1, there was a main effect for frequency, $F(1, 59) = 13.18$, $MSE = 4.01$, $\eta^2 = .18$, such that over-distribution was more pronounced for low- than for high-frequency items. (The confidence level is .05 for all significance tests reported in this paper.) There was neither a concreteness main effect nor a list main effect. However, with respect to concreteness, there was a Concreteness X List interaction, $F(1, 59) = 4.11$, $MSE = .12$, $\eta^2 = .07$, such that abstract items produced more over-distribution than concrete items on List 2 but not on List 1. With respect to list, there was a List X Frequency interaction, $F(1, 59) = 15.12$, $MSE = .12$, $\eta^2 = .20$, such that List 1 produced more over-distribution than List 2 with high-frequency items but not with low-frequency items. Last, in order to determine whether the observed levels of over-distribution were reliable, we computed one-sample t tests that compared the mean level of over-distribution in each of the 8 conditions to a predicted value of zero. Six of the $t(59)$ statistics were reliable: all four List 1 tests, plus the low-frequency/concrete and low-frequency/abstract tests for List 2.

Turning to the results of Experiment 2, averaging over the 8 conditions of the design (2 list-contexts X 2 levels of frequency X 2 levels of concreteness), it can be seen in Table 3 that the mean level of over-

distribution was lower than in Experiment 1, as expected. When we repeated the above ANOVA, using values of the over-distribution statistics for Experiment 2, there was a main effect for word-frequency, $F(1, 58) = 4.85$, $MSE = .30$, $\eta^2 = .08$, and a main effect for concreteness, $F(1, 58) = 5.44$, $MSE = .35$, $\eta^2 = .09$, both of which were in the predicted direction (i.e., more over-distribution for low-frequency items and abstract items). Last, in order to determine whether the observed levels of over-distribution were reliable, we computed one-sample t tests that compared the mean level of over-distribution in each of the 8 conditions of the design (2 lists X 2 levels of frequency x 2 levels of concreteness) to a predicted value of zero. Two of the $t(58)$ statistics were reliable: List 1/low-frequency/abstract, and List 2/low-frequency/abstract.

To summarize, using the over-distribution metrics $P(L1 \cap L2 | L1)$ and $P(L1 \cap L2 | L2)$, there was statistical support for the notion that retrieval treats some items as having been presented on both List 1 and List 2, notwithstanding that none has and that subjects are well aware of this fact. Importantly, the two word-content manipulations affected observed levels of over-distribution, as did the individual differences variable (University P versus University Q). Thus, although the list manipulation did not produce consistent effects, a good degree of experimental control of over-distribution was achieved with the word content and individual differences manipulations.

Model-Based Results

Although above-zero values of $P(L1 \cap L2 | L1)$ and $P(L1 \cap L2 | L2)$ are sufficient conditions for over-distribution, these statistics have two limitations. First, they are difference scores, and there are well-known reliability problems with such scores—specifically, that they cumulate the measurement error of the component scores that are used to compute them and, hence, are less reliable than the component scores, which lowers sensitivity to treatment effects (e.g., Cronbach, 1970). Hence, some of the conditions of Experiments 1 and 2 that did not exhibit over-distribution might do so with more reliable measures. Second, although $P(L1 \cap L2 | L1)$ and $P(L1 \cap L2 | L2)$ provide estimates of over-distribution, they do not identify the underlying processes that are responsible for variations in over-distribution across experiments and conditions. These limitations can be removed by estimating the parameters of the CPD model (Appendix, Equations A1-A6).

Model fits. Before doing that, however, we conducted model fit analyses to answer two general questions. The first was the obvious one, which is simply whether the CPD model delivered statistically tolerable fits to the data. The second question was concerned with the relative fits of the CPD model and the source-memory model that was discussed earlier. We saw earlier that although the CDP model predicts over-distribution, Batchelder and Riefer's (1990) source-memory model can accommodate this phenomenon under certain values of its parameters—explicitly, when the source guessing parameter has high values ($> .5$). The two-list version of the source-memory model that parallels the CPD model is presented in the Appendix (Equations A22-A27). What was at stake in the comparative fit analyses is the theoretical basis of over-distribution. The phenomenon is better interpreted as a consequence of item memory if the CPD fits are superior, but it is better interpreted as a consequence of source guessing if CPD fits are poorer.

With respect to the first question, there are five parameters in the CPD model (R , E , l , b , and b_{1U2}), and the experimental design provides six free empirical probabilities with which to estimate them ($p_{i,L1}$, $p_{E,L1}$, $p_{M,L1}$, $p_{i,\emptyset}$, $p_{E,\emptyset}$, and $p_{M,\emptyset}$). Thus, the fit test for this model, which asks whether it is adequate to account for the empirical probabilities, is a $G^2(1)$ statistic that is asymptotically distributed as χ^2 with a critical value of 3.84 to reject the null hypothesis of fit for any experimental condition at the .05 level. Because there are 8 conditions in each experiment, the experimentwise test of the null hypothesis that the CPD model is adequate to account for the data is a $G^2(8)$ statistic that is asymptotically distributed as χ^2 with a critical value of 30.72 to reject the null hypothesis at the .05 level. When this test was computed, the null hypothesis could not be rejected for either experiment, the values of the $G^2(8)$ statistic being 8.27 for Experiment 1 and 25.60 for Experiment 2. Hence, the experimentwise fit tests showed that the CPD model provided a statistically acceptable account of the data of both experiments.

With respect to the second question, it can be seen in the Appendix that the source-memory model also uses five parameters (D_1 , d_1 , g , D_N , and b) to account for the same empirical probabilities, though the parameters have different theoretical interpretations than those of the CPD model. Therefore, the fit test for the source-memory model, like the fit test for CPD, is a $G^2(1)$ statistic with a critical value of 3.84 to reject the null hypothesis of fit for any experimental condition, and also like CPD, the experimentwise test of the null hypothesis that the source-memory model is adequate to account for the data is a $G^2(8)$

statistic with a critical value of 30.72. When that test was computed, the null hypothesis was rejected in both experiments, the values of the $G^2(8)$ statistic being 52.40 for Experiment 1 and 156.31 for Experiment 2. In short, although the CPD model was acceptable in both experiments, the source-memory model was rejected in both experiments. Therefore, the latter model is not considered further with respect to these two experiments, though we reconsider it in Experiments 3 and 4.

Parametric tests of over-distribution. We return now to a point that was mentioned at the start of this section—namely, the CPD model's parameters provide more sensitive tests of over-distribution than $P(L1 \cap L2 | L1)$ and $P(L1 \cap L2 | L2)$ do. The relevant data appear in Tables 4 and 5. Maximum likelihood estimates of the model's parameters are reported in the upper half of Table 4, for the 8 conditions of each experiment. Here, the three memory parameters are defined as before: R = the probability that an item's presentation on its corresponding list is correctly recollected; E = the probability that an item's presentation on the wrong list is erroneously recollected; and I = the probability of item memory. In addition, there are bias parameters, b and b_{1U2} , that measure the probability of accepting a target that cannot be remembered in any these three ways when the episodic description is either L1? or L2? (parameter b) or L1UL2? (parameter b_{1U2}). Thus, R , E , and I are parameters that measure the degree to which manipulations affect the rates of true memory, false memory, and over-distribution, respectively.

The parameter estimates in Table 4 were used to compute the proportions of presented items that occupied the true memory, false memory, over-distribution, and forgetting states. Those data appear in Table 5. To understand the relation between the results in the Tables 4 and 5, it must be borne in mind that: (a) R = the proportion of items that occupied the true memory state; (b) $(1-R)E$ = the proportion of items that occupied the false memory state, so that the incidence of false memory is *directly proportional* to E but *inversely proportional* to R ; and (c) $(1-R)(1-E)I$ = the proportion of items that occupied the over-distribution state, so that the incidence of over-distribution is *directly proportional* to I but *inversely proportional* to R and E . As previously discussed, this means that (d) any manipulation that increases R drives true memory up but drives false memory and over-distribution down, (e) any manipulation that increases E drives false memory up but over-distribution down, and (f) any manipulation that increases I drives over-distribution up.

Using the CPD model and the data in Tables 4 and 5, an experimentwise parametric test of the null hypothesis that over-distribution *did not occur* can be computed for each experiment. This test asks whether the over-distribution state is required to account for the data of an experiment or whether just three episodic states (true memory, false memory, and forgetting) are sufficient. If over-distribution is unnecessary, the model has too many parameters because, by that assumption, $l = 0$. When that constraint is imposed on the model's equations, a test of the null hypothesis that the over-distribution state is unnecessary can be computed for any of the conditions of either experiment, which is a $G^2(2)$ statistic that is asymptotically distributed as χ^2 with a critical value of 5.99 for rejection of the null hypothesis at the .05 level. Because there were 8 conditions in each experiment, the experimentwise test of the null hypothesis that the over-distribution state is unnecessary is a $G^2(16)$ statistic that is asymptotically distributed as χ^2 with a critical value of 47.92. When that test was computed, the null hypothesis was rejected for both experiments, $G^2(16) = 294.89$ and 143.90 .

A final point to consider is that although this null hypothesis was rejected on an experimentwise basis in both experiments, this does not mean that the over-distribution state is required to account for the data of each of the conditions of an experiment, as suggested by the earlier analyses of the $P(L1 \cap L2 | L1)$ and $P(L1 \cap L2 | L2)$ statistics. Therefore, we examined the $G^2(2)$ tests for each condition. We found that this over-distribution state was necessary for account for the data of all 8 conditions in both experiments. Hence, the model-based analyses provided stronger evidence of over-distribution than the $P(L1 \cap L2 | L1)$ and $P(L1 \cap L2 | L2)$ statistics.

Detailed parametric results. In Table 4, note that the value of the parameter that produces over-distribution, l , was quite substantial in these experiments ($M = .31$) and that, in Table 5, nearly a quarter of the items in these conditions occupied the over-distribution state ($M = .22$). Notice, too, that in those same conditions, as in prior studies of over-distribution in semantic false memory (Brainerd & Reyna, 2008), over-distribution was a more common form of memory distortion than false memory, with the overall estimated level of the latter ($M = .11$) being half of the former.

1. General findings. How did the proportion of items occupying the over-distribution state (Table 5) vary as a function of the experimental manipulations? First, over-distribution was highest for low-frequency abstract items, with more than a third of such items being over distributed, across the two

experiments. Second, the key effect of low-frequency items, relative to high-frequency ones, was to more than double the level of over-distribution—from .13 to .30, averaging across all conditions of the two experiments. In contrast, frequency did not have a consistent effect on levels of false memory. Third, the key effect of concrete items, relative to abstract ones, was to drive up the level of true memory—from .16 to .27, averaging across all conditions of the two experiments. Concreteness had no effect on false memory, and its effects on over-distribution depended on frequency: Averaging across the two experiments, with low-frequency items but not high-frequency items, concrete items drove over-distribution down (from .36 to .24).

Fourth, with respect to the individual differences manipulation, we mentioned that our prior work indicates that University P subjects perform less well than University Q subjects on measures of verbatim memory for word lists, which led us to expect that University P subjects would display lower levels of true source memory and, therefore, higher levels of the over-distribution. That expectation was borne out for List 1, though levels of true memory and levels of over-distribution were more comparable for the two groups on List 2. On List 1, the level of true memory was higher in Experiment 2 than in Experiment 1 (grand M s = .26 and .20), and the proportion of over distributed items was greater in Experiment 1 than in Experiment 2 (grand M s = .25 and .18).

2. Parametric explanations of findings. Now, we identify the process loci of each of these four effects by conducting significance tests of the values of R , E , and I in Table 4. In models of this sort, significance tests of between-condition differences in parameter values are likelihood ratio statistics (see Riefer & Batchelder, 1988). In such tests, the joint likelihood of the data of two conditions when all model parameters are free to vary is compared to the joint likelihood when one or more parameters are required to have the same value in both conditions. Twice the negative natural log of the latter likelihood divided by the former likelihood is a G^2 statistic with degrees of freedom equal to the number of parameters that are required to have equal values. In the tests that we conducted, except where indicated, the value of a single memory parameter (R , E , or I) was always being compared between conditions, so that the G^2 tests had one degree of freedom (with a .05 critical value of 3.84). We only report results for comparisons that produced null hypothesis rejections.

With respect to the first finding mentioned above (that over-distribution levels were highest with low-frequency/abstract items), the explanation lies in the R and I parameters: For both lists in both experiments, the value of R was smaller for low-frequency/abstract items than for the other three item types (grand $M_s = .13$ and $.26$), whereas the value of I was larger for these items (grand $M_s = .45$ and $.26$). All but one of the $G^2(1)$ tests for item-type differences in R produced null hypothesis rejections, and all eight of the $G^2(1)$ tests for item-type differences in I produced null hypothesis rejections. (Remember here that because over-distribution levels are inversely proportional to the value of R , even when I does not differ between item types, over-distribution levels will differ if R values differ.) Turning to the second finding mentioned above (that frequency had consistent effects on over-distribution, though not on true or false memory), with the parameter values in Table 4, it is possible to compute a total of 8 $G^2(1)$ tests of how differences in frequency affected any parameter, with concreteness held constant. When these eight tests were computed for I , each produced a null hypothesis rejection.

Concerning the third finding (concrete items drive true memory levels up), the parametric results showed that the value of R was larger for concrete than for abstract items ($M_s = .28$ and $.18$), as predicted. This difference was also responsible for the fact that over-distribution levels were higher for low-frequency abstract items than for low-frequency concrete ones because the value of I was *the same* for the two types of items abstract items ($M_s = .38$ and $.38$). Thus, the reason that the level of over-distribution is higher for low-frequency/abstract items than for low-frequency/concrete items, is not because abstract items directly increase item memory. Instead, this is an indirect effect that results from the fact that abstract items decrease recollection of presentation details, to which the level of over-distribution is inversely proportional. Across the two experiments, it was possible to compute a total of 8 $G^2(1)$ tests of how differences in concreteness affected R , and five of the eight tests showed that R was reliably larger for concrete items.

Finally, the fourth finding mentioned above was that over-distribution levels on List 1 were noticeably higher for University P subjects (Experiment 1) than for University Q subjects (Experiment 2), as expected. As predicted, the locus of this effect was the parameter R , which had a larger mean value in Experiment 2 than Experiment 1. When between-experiment significance tests were computed for this parameter on a condition-by-condition basis, four of the eight $G^2(1)$ produced a null hypothesis rejection.

In each instance, the R estimate in Experiment 2 was reliably larger than the corresponding estimate in Experiment 1.

Discussion

These experiments detected the same over-distribution phenomenon with a standard source-monitoring procedure that has been previously identified with a very different paradigm, semantic false memory. There were also quantitative similarities inasmuch as over-distribution was more prevalent than false memory, as has been found in semantic false memory. The grand means for over-distribution and false memory in Table 5 are .22 and .11, whereas the corresponding grand means in Brainerd et al (2010) were .17 and .09.

And yet, the mechanisms that are responsible for over-distribution are presumably different in the two paradigms. In semantic false memory, the items on study lists repeatedly instantiate certain meanings (e.g., medicine, furniture), so that the retrieval of gist traces of those meanings is the obvious route to over-distribution (Brainerd & Reyna, 2008). For instance, if a test cue such as *doctor* retrieves a gist trace of a meaning such as medicine, that information is consistent both with the cue being a target and being a meaning-preserving distractor. Hence, *doctor* may be accepted both when a probe states that it is a target and when a probe states that it is a related distractor, although that is an impossible conjunction. This cannot be the route to over-distribution in source monitoring because study lists do not repeatedly instantiate meanings and distractors are not meaning-preserving (Reyna & Titcomb, 1997). Now, item memory without recollection is the logical route to over-distribution: If a test cue such as *bagpipe* retrieves a memory of its presentation that is not accompanied by contextual details, that information is congruent with *bagpipe* having appeared on both lists. Thus, it may be accepted when the probe question is L1? and when the probe question is L2?, although, again, that is an impossible conjunction. These considerations argue that over-distribution is a basic distortion phenomenon that cuts across quite different memory processes.

Last, we briefly comment on the relation between true and false memory in these experiments. In the larger literature, the relation between standard measures of true and false memory, such as bias-corrected acceptance rates for targets and meaning-preserving distractors, has figured centrally, with the usual finding being that such measures are either uncorrelated or negatively correlated (for a review, see

Brainerd & Reyna, 2005). As we know, however, such standard measures confound true and false memory with over-distribution. In contrast, the R and E parameters in the present experiments are not confounded in this way. Inspection of the 16 paired values of these parameters in Table 5 suggests that they covary positively; that experimental variations that make memory more accurate also make it more inaccurate. When we computed the correlation between these paired values, it was indeed positive, $r = .47$ ($df = 15$, $p < .03$).

Experiments 3 and 4

The purpose of these two experiments was to determine whether higher levels of over-distribution could be generated, following the logic of the first two experiments, by extending the process-dissociation design one step further, from two lists to three. The motivating idea is that holding all other factors constant, over-distribution ought be more prevalent with three lists with than two simply because three lists provide more opportunities for this form of distortion to occur; that is, more distinct ways in which a target can be assigned to mutually incompatible episodic states. Returning to the example of an item that is presented on List 1, there is only way to over distribute it with the two list procedure—namely, by remembering it as having been presented on both lists, which is represented by the quantity $P(L1 \cap L2 | L1)$. Suppose that the same set of targets that was previously presented in two list-contexts is presented in three. Now, there are three ways of over distributing this same item. It can be remembered as having been presented on Lists 1 and 2, represented by the quantity $P(L1 \cap L2 \cap \sim L3 | L1)$, or on Lists 1 and 3, represented by the quantity $P(L1 \cap \sim L2 \cap L3 | L1)$, or on all three lists, represented by the quantity $P(L1 \cap L2 \cap L3 | L1)$. If the values of the memory and bias parameters of the CPD model are roughly the same as before, the sum of these three quantities will be greater than $P(L1 \cap L2 | L1)$.

This can be seen by considering how the CPD expressions in Equations 4-6 change with the addition of a third list-context. The revised expressions for a List 1 target are

$$p_{i,L1} = R_1 + (1-R_1)E_2 + (1-R_1)(1-E_2)E_3 + (1-R_1)(1-E_2)(1-E_3)I_1, \quad (7)$$

$$p_{E2,L1} = (1-R_1)E_2 + (1-R_1)(1-E_2)(1-E_3)I_1, \quad (8)$$

$$p_{E3,L1} = (1-R_1)(1-E_2)E_3 + (1-R_1)(1-E_2)(1-E_3)I_1, \text{ and} \quad (9)$$

$$p_{M,L1} = R_1 + (1-R_1)(1-E_2)(1-E_3)I_1. \quad (10)$$

The parameters R_1 , E_2 , and I_1 have the same meanings as before. E_3 is a new false memory parameter, for the additional List 3, which measures the probability that a List 1 cue provokes retrieval of contextual details of List 3. We now have four empirical quantities, which are sufficient to measure over-distribution: (a) $P(L1UL2UL3 | L1) = p_{i,L1}$ is the probability that a List 1 item is accepted as having been on List 1 or List 2 or List 3; (b) $P(L2 | L1) = p_{E2,L1}$ is the probability that a List 1 item is accepted as having been on List 2, which subjects know also means that it was not on List 1 or List 3; (c) $P(L3 | L1) = p_{E3,L1}$ is the probability that a List 1 item is accepted as having been on List 3, which subjects know also means that it was not on List 1 or List 2; and (d) $P(L1 | L1) = p_{M,L1}$ is the probability that a List 1 item is accepted as having been on List 1, which subjects know also means that it was not on List 2 or List 3. The over-distribution probability $O_D = P(L1 \cap L2 \cap L3 | L1) + P(L1 \cap L2 \cap \sim L3 | L1) + P(L1 \cap \sim L2 \cap L3 | L1)$ is obtained as before, via the rule $P(L1 | L1) + P(L2 | L1) + P(L3 | L1) = P(L1UL2UL3 | L1) + O_D$. Also as before, it is objectively true that $O_D = 0$ because List 1 words only appear on that list, and subjects are instructed on this fact. However, the CPD model predicts $O_D > 0$. In that connection, some algebraic manipulation of Equations 7-10 reveals that $O_D = p_{E2,L1} + p_{E3,L1} p_{M,L1} - p_{i,L1} = 2(1-R_1)(1-E_2)(1-E_3)I_1$, according to the model.

In other words, the over-distribution probability is twice the probability that subjects have item memory for words whose presentation details cannot be recollected. Because the memory load, with respect to the number of items that are presented, is the same in the two- and three-list designs, suppose that the values of R , E , and I parameters are the same as in the first two experiments. Those means are .25, .17, and .24. If we substitute these values in the expression $2(1-R_1)(1-E_2)(1-E_3)I_1$, letting $E_2 = E_3 = .17$, the predicted over-distribution probability is .24, which is roughly a 40% increase over the mean probability in the first two experiments.

Method

Subjects

The subjects in Experiment 3 were 63 introductory psychology students from University P, who participated in the experiment to fulfill a course requirement. The subjects in Experiment 2 were 57 introductory psychology students from University Q, who participated in the experiment to fulfill a course requirement. The experiments were run in parallel, at the same time of year, and the respective subject samples did not differ reliably in mean age or gender composition.

Materials and Procedure

The methodological details of these experiments were the same as those of Experiments 1 and 2, except for three modifications. The first and most important one is that there were now three presentation contexts. As before, the presentation context for each word list was distinguished by a unique combination of background color, font, and order (first, second, third) cues. Crucially, the total number of words that were presented over the three lists-contexts (108) was the same as the total number that had been presented over the two list-contexts in Experiments 1 and 2. Thus, the only increase in memory load was the addition of a single presentation context. Each list-context consisted of 36 words: an opening buffer of 2 words, 32 focal words (8 exemplars of each Frequency X Concreteness combination), and a closing buffer of two words.

The other two modifications involved the memory test: Because there were three presentation contexts, (a) there was a third nondisjunctive description (Presented on List 3?) and (b) the disjunctive description referred to three contexts rather than two (Presented on List 1 or List 2 or List 3?). Thus, although the 192 test cues were the same as in the first two experiments, 4 episodic descriptions (rather than 3) were factorially manipulated over these test cues (i.e., 48 cues per description rather than 64).

Results and Discussion

As before, findings on episodic over-distribution are reported in two waves. First, we report ANOVAs using the over-distribution metric $O_D = P(L1 \cap L2 \cap L3 | L1) + P(L1 \cap L2 \cap \sim L3 | L1) + P(L1 \cap \sim L2 \cap L3 | L1)$, with positive values indicating the presence of over-distribution. Second, we report findings that made use of the parameters of the three-list version of CPD model (see Appendix). The effects of the experimental manipulations on the R , E , and I parameters are considered, and parameters are used to estimate of levels true memory, false memory, over-distribution, and forgetting.

Results for the Over-Distribution Metric

The memory probes were constructed by factorially manipulating the four types of episodic descriptions over four types of cues (List 1 targets, List 2 targets, List 3 targets, and distractors). Consequently, there were 16 acceptance probabilities (4 descriptions X 4 cues) for each of the 4 types of items (2 levels of concreteness X 2 levels of frequency) for each of the three list-contexts. For each list-context, we used the distractor acceptance probabilities to correct the acceptance probabilities for

presented items for response bias, using the two-high threshold correction (P_T), as before. The corrected probabilities are displayed by condition in the bottom half of Table 4, along with values of the over-distribution probability, O_D .

Inspection of the results for Experiments 3 and 4 reveals seven findings that are worthy of note. First, over-distribution was strongly in evidence because 23 of the 24 values of O_D are greater than zero. Second, in comparison to the first two experiments, those values are large in absolute terms. The grand mean, over the 24 values for the two experiments, was .63, which is roughly five times the grand mean in the first two experiments. Third, word frequency had the same effect on over-distribution as in the first two experiments. The average value of O_D was roughly 50% greater for low- than for high-frequency items (.74 versus .52). Fourth, although concreteness had the anticipated effect (more over-distribution for abstract items), the effect only occurred for the last list. Fifth, unlike the first two experiments, there was a large list-order effect on over-distribution of the serial position variety. Averaging over the two experiments, the O_D statistic was largest for List 1, as previously predicted, smallest for list 2, and intermediate for list 3. Sixth, over-distribution was more marked in Experiment 3 than in Experiment 4, as expected (grand M s of .88 and .38, respectively). The seventh finding, which is counterintuitive, occurred in Experiment 3—namely, that the over-distribution metric sometimes exceeded unity. Here, it is important to note that a key property of the CPD model is that it permits such violations of classical probability because its expressions for this metric are $2(1-R_1)(1-E_2)(1-E_3)I_1$ for List 1, $2(1-R_2)(1-E_1)(1-E_3)I_2$ for List 2, and $2(1-R_3)(1-E_1)(1-E_2)I_3$ for List 3. It is easy to see that the upper bound of these expressions, as the R and E parameters approach 0 and the I parameter approaches 1, is 2, rather than 1.

We computed a 2 (frequency: high versus low) X 2 (concreteness: concrete versus abstract) X 3 (list order: 1 versus 2 versus 3) repeated measures ANOVA for each experiment, with the subjects' O_D scores as the dependent variable. In Experiment 3, there were large main effects for frequency, $F(1, 62) = 26.60$, $MSE = .61$, $\eta^2 = .30$, and list-order, $F(2, 62) = 27.67$, $MSE = .38$, $\eta^2 = .31$. Over-distribution was more pronounced for low- than for high-frequency items, as in the first two experiments, but unlike the first two experiments, list order strongly affected over-distribution, with the effect being of the serial position variety (List 1 > List 3 > List 2). There was also a List-Order X Concreteness interaction, $F(2, 62) = 3.27$, $MSE = .48$, $\eta^2 = .05$, such that over-distribution was more marked for abstract than for concrete items on

List 3 only. Last, in order to determine whether the observed levels of over-distribution were reliable, we computed one-sample t tests that compared the mean O_D score in each of the 12 conditions of the design (3 lists X 2 levels of frequency x 2 levels of concreteness) to a predicted value of zero. All 12 of the $t(62)$ statistics were reliable.

Turning to Experiment 4, the 2 X 2 X 3 ANOVA of O_D scores again produced large main effects for frequency, $F(1, 56) = 12.00$, $MSE = .65$, $\eta^2 = .18$, and list-order, $F(2, 56) = 15.62$, $MSE = .51$, $\eta^2 = .22$, with both effects taking the same form as in Experiment 3. Also like Experiment 3, there was a List-Order X Concreteness interaction, $F(2, 63) = 3.73$, $MSE = .45$, $\eta^2 = .07$, such that over-distribution was more marked for abstract than for concrete items on List 3 only. In order to determine whether the observed levels of over-distribution were reliable, we again computed one-sample t tests that compared the mean O_D score in each of the 12 conditions to a predicted value of zero. Eleven of the $t(56)$ statistics were reliable, the List 2/high-frequency/abstract test being the only one that was not.

In brief, increasing the number of presentation contexts minimally, from two to three, while keeping the number of targets fixed, produced a marked increase in the over-distribution phenomenon, with the overall average value of the O_D metric increasing five-fold. Further, the effects of the frequency and concreteness manipulations replicated their effects in the first two experiments, and the effects of the list-order manipulation were in the predicted direction. Hence, the level of experimental control over the over-distribution phenomenon that was achieved in the first two experiments was not sacrificed by increasing the number of presentation contexts, and indeed, it improved because list-order effects were stronger.

Model-Based Results

The CPD model for the present experiments (Appendix, Equations A7-A14) is slightly more complex than the model for the first two experiments because, with three presentation contexts, there are two false memory states for the items that are presented in each context. For example, on the memory test, an item presented on List 1 can be in the true memory state, the over-distribution state, or the forgetting state, as in the first two experiments, but it can also be in one of two false memory states (presented on List 2 or presented on List 3). An item presented on List 2 or List 3 can likewise be in one of two false memory states. Thus, whereas the model for two contexts (Experiments 1 and 2) has a single false memory parameter for each list, the model for three contexts (Experiments 3 and 4) has two

false memory parameters per list—specifically, E_2 and E_3 for List 1, E_1 and E_3 for List 2, and E_1 and E_2 for List 3. Moreover, E_1 can have different values for List 2 versus List 3, E_2 can have different values for List 1 versus List 3, and E_3 can have different values for List 1 versus List 2.

As in Experiments 1 and 2, we report the model analyses in three steps. First, fit results are reported for the CPD model and for the alternative source-memory model. Second, parametric tests of the over-distribution phenomenon are presented. Third, detailed results for the true memory, false memory, and item memory parameters are considered.

Model fits. The fit analyses focused on the same two questions as before: Does the CPD model deliver statistically tolerable fits to the data? What are the comparative levels of fit for the CPD model versus the source-memory model? The three-list version of the source-memory model that parallels the three-list CPD model (Equations A7-A14) is presented in the Appendix (Equations A22-A27). As before, the comparative fit analyses address the theoretical basis of over-distribution. Over-distribution is better interpreted as a consequence of item memory if the CPD fits are superior, but it is better interpreted as a consequence of source guessing if CPD fits are poorer.

With respect to the first question, there are six parameters in the CPD model for List 1 (R , E_2 , E_3 , l , b , and b_{1U2}) and for the other two lists, and the experimental design provides eight free empirical probabilities with which to estimate them ($p_{I,L1}$, $p_{E2,L1}$, $p_{E3,L1}$, $p_{M,L1}$, $p_{I,\emptyset}$, $p_{E2,\emptyset}$, $p_{32,\emptyset}$, and $p_{M,\emptyset}$). Thus, the fit test for this model, which asks whether it is adequate to account for the empirical probabilities, is a $G^2(2)$ statistic that is asymptotically distributed as χ^2 with a critical value of 5.99 to reject the null hypothesis of fit for any experimental condition at the .05 level. Because there are 8 conditions in each experiment, the experimentwise test of the null hypothesis that the CPD model is adequate to account for the data is a $G^2(24)$ statistic that is asymptotically distributed as χ^2 with a critical value of 71.88 to reject the null hypothesis at the .05 level. When this test was computed, the null hypothesis could not be rejected for either experiment, the values of the $G^2(24)$ statistic being 36.85 for Experiment 1 and 36.28 for Experiment 2. As in the first two experiments, then, the experimentwise fit tests showed that the CPD model provided a statistically acceptable account of the data.

With respect to the second question, it can be seen in the Appendix that the source-memory model uses five parameters (D_1 , d_1 , g , D_N , and b) to account for the same empirical probabilities, so that the fit

test for the source-memory model is a $G^2(3)$ statistic with a critical value of 7.82 to reject the null hypothesis of fit for any experimental condition. Because there are 12 conditions in each experiment, the experimentwise test of the null hypothesis that the source-memory model is adequate to account for the data is a $G^2(36)$ statistic with a critical value of 93.84. When that test was computed, the null hypothesis was rejected in both experiments, the values of the $G^2(36)$ statistic being 157.18 for Experiment 1 and 152.77 for Experiment 2. Therefore, the comparative fit picture was the same as in the first two experiments: The CPD model was acceptable, the source-memory model was rejected, and consequently, the source-memory model is not considered further.

Parametric tests of over-distribution. The model-based data appear in the bottom halves of Tables 4 and 5. Maximum likelihood estimates of the parameters are reported in Table 4, for the 12 conditions of each experiment. As before, the R , E , and I parameters measure the degree to which manipulations directly affect true memory, false memory, and item memory, respectively. As in the first two experiments, values of these parameters were used to estimate the proportions of items that occupied the true memory, false memory, over-distribution, and forgetting states in each condition of each experiment, with those data appearing in Table 5. With respect to the relationship between the results in Tables 4 and 5, remember that (a) the proportions of items that occupy any false memory state are inversely proportional to the true memory parameter, R , but directly proportional to the false memory parameter for that state and that (b) the proportions of items that occupy the over-distribution state for any list are inversely proportional to R and to the two false memory parameters for that list but directly proportional to the item memory parameter, I . Thus, even if a manipulation does not affect I , any manipulation that drives down R or drives down the false memory parameters will drive up the proportion of items that are over distributed.

We conducted the same types of parametric tests of over-distribution as in the first two experiments. These tests asked whether the CPD model could account for the data with only three episodic states (true memory, false memory, and forgetting) or whether the fourth state (over-distribution) was also required. For each condition of each experiment, the relevant test is a $G^2(3)$ statistic (see Appendix) that is asymptotically distributed as χ^2 and evaluates the null hypothesis that only three states are required to account for the data. As there were 12 conditions in each experiment, the

experimentwise test of this null hypothesis had a critical value of 93.84. When this test was conducted, the null hypothesis was rejected for both experiments, the values of the test statistic being 383.51 (Experiment 3) and 247.39 (Experiment 4).

As we saw in the first two experiments, however, the fact that this test produced an experimentwise null hypothesis rejection in both experiments does not mean that a fourth episodic state is necessary to account for the data of each of all 12 conditions of each experiment. Hence, as before, the individual $G^2(3)$ statistics for each condition were examined. We found that all 24 tests exceeded the critical value of 7.84, so that over-distribution was present at statistically reliable levels in all conditions.

Detailed parametric results. Turning to the detailed parametric findings, we note two results of general significance before passing on to other results. First, as predicted, it can be seen in Table 5 that when subjects studied the same items in three contexts rather than two, the proportion of items that were over distributed increased greatly. Second, it can be seen in Table 4 that when subjects studied the same items in three contexts rather than two, the value of the item memory parameter, l , increased by roughly 50%—from .30 overall in the first two experiments to .46 in these experiments. For any randomly selected target, then, subjects were better at remembering that that specific item had been presented when the same targets were presented in more rather than fewer distinctive contexts.

1. General findings. We begin by considering the results in the bottom half of Table 5—specifically, how the proportion of over distributed items varied as a function of the experimental manipulations. First, as predicted (and as observed in the first two experiments), over-distribution levels were highest for low-frequency/abstract items, with 40% of such items being over distributed, across the two experiments. Second, the primary effect of the frequency manipulation was on the level of over-distribution. Low-frequency items increased the average level from .25 to .37 overall, which parallels the frequency effect in the first two experiments. In contrast, also as in the first two experiments, frequency did not have consistent effects on either true or false memory levels. Third, it was again found that concrete items, relative to abstract ones, drove up levels of true memory. Fourth, as predicted (and as observed in the first two experiments), University P subjects' levels of true memory were lower than those of University Q subjects. Overall, the decrease in true memory levels for University P subjects, relative to University Q subjects (.09 versus .26), was more pronounced than in the first the two experiments.

The preceding patterns were present in the first two experiments, but two other results are new. The first is that there were large list-order effects, and further, those effects were specific to over-distribution. Averaging across the two experiments, over-distribution levels were highest for List 1, lowest for List 2, and intermediate for list 3 (grand $M_s = .40, .23, \text{ and } .30$, respectively). In contrast, the proportions of items that occupied the true and false memory states varied negligibly over the three lists. The other new result is concerned with forgetting. Surprisingly, the level of forgetting declined dramatically in Experiment 4, relative to Experiment 2, from .44 to .29, though this decline was not present in Experiment 3 relative to Experiment 1 (where the respective forgetting probabilities were .46 and .44). As can be seen in Table 5, the reductions in forgetting were due to increases in over-distribution in Experiment 4, rather than to increases in true or false memory.

2. Parametric explanations of findings. We identified the process loci of the preceding effects via significance tests of the values of the R , E_1 , E_2 , E_3 , and I parameters in the lower half of Table 4. As previously discussed, significance tests of between-condition differences in parameter values are likelihood ratio statistics, with degrees of freedom equal to the difference in the number of parameters that are estimated in the numerator versus the denominator of such a ratio. As also discussed, the likelihood ratio tests that we computed always had one degree of freedom because each compared the value of a single parameter (R , E_1 , E_2 , E_3 , or I) between two conditions that varied on one of the experimental manipulations. As in Experiments 1 and 2, we only report parameter differences for which likelihood ratio tests produced null hypothesis rejections.

With respect to the first finding mentioned above (that over-distribution levels were highest with low-frequency/abstract items), the explanation in the first two experiments lay in the R and I parameters. (The average value of R was smallest and the average value of I was largest for low-frequency/abstract items.) In these experiments, these parameters were again responsible but only one was responsible within individual experiments. In Experiment 3, the average value of I was larger for low-frequency/abstract items than for the other three types of items (.52 versus .38), whereas in Experiment 4, the average value of R was smaller for low-frequency/abstract items than for the other three types of items (.15 versus .29). In Experiment 3, 6 of the $G^2(1)$ tests for between-condition differences in I produced null hypothesis

rejections, and in Experiment 4, 5 of the $G^2(1)$ tests for between-condition differences in R produced null hypothesis rejections.

Turning to the second finding mentioned above (that frequency had consistent effects on over-distribution but not on true or false memory), with the parameter values in Table 4, it is possible to compute 12 $G^2(1)$ tests of how differences in frequency affected any parameter with concreteness held constant, 6 per experiment. When these 12 tests were computed for I , 7 of them produced a null hypothesis rejection, with the overall means for this parameter being .52 for low-frequency items and .39 for high-frequency items. When these 12 tests were computed for R , only 2 produced a null hypothesis rejection, and those differences were not in the same direction (i.e., R was larger for high-frequency items in one test but larger for low-frequency items in the other test). In short, the overall tendency of low-frequency items to elevate over-distribution was an I -specific effect, so that frequency affected over-distribution directly (by affecting I) rather than indirectly (by affecting the true or false memory parameters).

Concerning the third finding mentioned above (that concrete items drove over-distribution levels down for List 3), this was once again an R -specific effect (an indirect one), as predicted. On List 3, the overall means of this parameter were .21 and .08 for concrete and abstract items, respectively. Three of the four $G^2(1)$ tests for R , on List 3, were reliable. In contrast, the overall means for I were quite similar for concrete versus abstract items, and the differences for specific comparisons were not always in the same direction. The same was true for the three false memory parameters. Hence, as before, the reason that abstract items increased the level of over-distribution was *not* because such items increased over-distribution directly, but because abstract items decreased true memory, to which over-distribution is inversely proportional.

With respect to the fourth finding (that over-distribution levels were again higher for University P subjects than for University Q subjects), the parameter analyses showed that this, too, was an R -specific (indirect) effect. As in the first two experiments, a between-experiment omnibus likelihood ratio test was computed for each of the three lists to determine if some of the model's parameters differed for Experiment 3 versus Experiment 4. Each of these three tests was a $G^2(5)$ statistic, with a critical value of 11.07 to reject the null hypothesis that the two experiments produced the same parameter values (for a

given list). This test produced a null hypothesis rejection for each list. Next, we computed 12 $G^2(1)$ tests for each model parameter (2 levels of concreteness X 2 levels of frequency X 3 lists), which evaluated the null hypothesis that that parameter had the same value in the two experiments. With respect to R , 9 of the 12 tests exceeded the critical value of 3.84. Averaging over the 12 list conditions of each experiment, the mean of this parameter was nearly 3 times larger in Experiment 4 than in Experiment 3 (grand $M_s = .25$ versus $.09$). In contrast, differences in the false memory parameters and the I parameter for Experiment 3 versus Experiment 4 were small. Because there are 2 E parameters for each list, there were 24 $G^2(1)$ tests of how false memory parameters are affected by differences in subject populations, with frequency, concreteness, and list held constant. When these tests were computed, only 4 produced null hypothesis rejections, and the parameter differences in those tests were not all in the same direction. Last, with respect to the 12 $G^2(1)$ tests of how the I parameter is affected by differences in subject populations, 4 produced null hypothesis rejections, and the parameter differences in those tests were not all in the same direction. The crucial point here is that because the proportion of over distributed items is inversely proportional to R , as well as directly proportional to I , the large differences in the true memory parameter between University P versus University Q subjects were responsible for the much higher levels of over-distribution in the former subjects.

With respect to the fifth finding (that there were marked list-order effects on over-distribution), the parameter analyses showed that like item frequency, this was chiefly an I -specific (direct) effect. Remember that the nature of this effect is that over-distribution was much higher for List 1 than for List 2 or List 3, as predicted, and that List 3 over-distribution was somewhat higher than List 2 over-distribution. The primacy portion of this effect (List 1 versus other lists) was entirely I -specific. Averaging over the two experiments, the mean value of this parameter ($.57$) was larger than the mean values for Lists 2 and 3 ($.39$ and $.42$, respectively). In each experiment, it is possible to compute 4 $G^2(1)$ tests of the null hypothesis that I had the same value for List 1 and List 2 and a further 4 $G^2(1)$ tests of the null hypothesis that I had the same value for List 1 and List 3, for a total of 16 tests in all (8 per experiment). When those tests were computed, 14 produced a null hypothesis rejection and all list differences in I were in the same direction (I was larger for List 1). With respect to List 2 versus List 3, we saw that the smaller difference in the level of over-distribution for List 3 was only reliable in Experiment 3. That difference was both an I

and an R effect. Specifically, when $G^2(1)$ tests were computed that compared the values of each of the these parameters for List 3 versus List 2, two of the four I tests showed that this parameter was reliably larger for List 3, and two of the four R tests showed that this parameter was reliably smaller for List 3. In sum, then, list-order effects on levels of over-distribution were overwhelmingly due to differences in I , the largest differences (List 1 versus the other lists) being pure- I effects, with a slight contribution from R in the case of List 3 versus List 2.

Regarding the sixth finding (that forgetting levels were lower with three contexts than with two in Experiment 4), note, first, that forgetting levels will decrease if *any* of the three memory parameters increase because, according to the model, the probability that a presented item will be forgotten is given by the expression $(1-R)(1-E_i)(1-E_j)(1-I)$, where i and j denote the two contexts in which that item was *not* presented. Thus, identifying the loci of the declines in forgetting entailed comparing the parameter values in Experiment 4 to those in Experiment 2 on a list-by-list basis. To reduce the number of parameter comparisons, we conducted a preliminary analysis of the forgetting proportions for the individual List X Frequency X Concreteness combinations of Experiment 2 versus Experiment 4, in order to determine which ones produced reliable differences in forgetting rates. Because items were presented in three contexts in Experiment 4 but in only two in Experiment 2, a decision had to be taken as to which of the two lists from the latter experiments should be compared to those in the former experiments. Naturally, List 1 was compared across experiments, and we also compared the last list across experiments (i.e., List 2 in Experiment 2 to List 3 in Experiment 4). Using exact probability tests and the .05 level of confidence, the results were as follows. List 1 produced reliable differences in forgetting rates between Experiment 2 and Experiment 4 for three of these combinations (low-frequency/concrete was the exception). The last list (List 2 in the first two experiments versus List 3 in the present experiments) produced reliable differences in forgetting rates for three of the four Frequency X Concreteness combinations (high-frequency/abstract was the exception) between Experiment 2 and Experiment 4.

The next step was to determine the parametric locus of each of these between-experiment differences in forgetting rates. That was done as follows. For List 1 in Experiment 2 versus Experiment 4, for each of the Frequency X Concreteness combinations that produced reliable differences in forgetting rates, we computed a series of four $G^2(3)$ tests, one apiece for R , I , and the two false memory

parameters, that compared the average values of those parameters across the three combinations. The key finding was that the average value of I increased between Experiment 2 and Experiment 4, from .31 to .53. For the last-list comparison of Experiment 2 versus Experiment 4, for each of the Frequency X Concreteness combinations that produced reliable differences in forgetting rates, we computed a series of four $G^2(3)$ tests, one apiece for R , I , and the two false memory parameters, that compared the average values of those parameters across the three combinations. We found that once again, the decline in forgetting between experiments was I -specific. Consequently, there was a single, simple reason for the decreases in forgetting rates that resulted from increasing the number of presentation contexts. Three presentation contexts increased the value of the item memory parameter, relative to two presentation contexts.

Finally, we also computed a full series of between-experiment $G^2(4)$ tests, one apiece for R , I , and the two false memory parameters, for List 1 in Experiment 1 versus Experiment 3 and Experiment 2 versus Experiment 4 and also for the last list in Experiment 1 versus Experiment 3 and Experiment 2 versus Experiment 4. Those tests compared the average values of the memory parameters for the two subject populations when there were two versus three presentation contexts. The purpose, beyond the forgetting analyses that were just reported, was simply to identify the overall process differences that result from studying the same items in a smaller versus a larger number of contexts. The analyses converged on the conclusion that increasing the number of presentation contexts enhances item memory. For List 1, the $G^2(4)$ test produced a null hypothesis rejection for the I parameter for Experiment 1 versus Experiment 3 (grand means = .39 versus .51) and for Experiment 3 versus Experiment 4 (grand means = .31 versus .62). Likewise, for the last list, the $G^2(4)$ test produced a null hypothesis rejection for the I parameter for Experiment 1 versus Experiment 3 (grand means = .26 versus .39) and for Experiment 3 versus Experiment 4 (grand means = .26 versus .45). In contrast, the $G^2(4)$ tests for the two E parameters produced no reliable differences as a function of number of presentation contexts, and the $G^2(4)$ tests for the R parameter produced only two: For University P subjects, R values were higher for two versus three lists on List 1 (grand means = .18 versus .08) and for the last list (grand means = .23 versus .08). The indicated conclusion, then, is that as long as the number of targets is held constant,

increasing the number of presentation contexts has little effect on recollection of contextual details, but it consistently elevates item memory.

Discussion

As the number of presentation contexts increases, so does the number of distinct ways in which an item can be over distributed. Even if other factors are held constant, over-distribution should increase, then, merely by increasing the number of presentation contexts. It did. When the *same set* of targets was presented using three lists rather than two, the over-distribution metric rose dramatically in University P subjects and in University Q subjects. We mentioned that if values of the *R*, *E*, and *I* parameters from the first two experiments were not affected by switching from two lists to three, the CPD model predicts roughly a 40% increase in over-distribution. The observed increases were much greater because merely allocating the same words to three list-contexts affected one of the parameters, *I*, substantially.

As we know, the model posits that levels of over-distribution are inversely related to the *R* and *E* parameters but directly related to *I*. The latter increased from an overall mean of .31 in the first two experiments to .46 in the present experiments. The psychological meaning of this is that if an item was not in the true memory state or the false memory state, there was roughly a 50-50 chance that it would be over distributed, whereas there was roughly a one-third chance in the first two experiments. There were also mean declines in the *R* and *E* parameters, which increased over-distribution, but they were fractions of the increase in *I*. Thus, the key process-level effect of increasing the number of lists on which targets are presented, was to substantially increase item memory, as distinct from true and false memory for presentation context.

Beyond this, the picture of how over-distribution varies as function of theoretically-motivated manipulations agreed with the picture in the first two experiments with respect to the frequency, concreteness, and individual differences variables. As before, over-distribution levels were higher for infrequent words, for abstract words, and for University P subjects, and the process-level reasons were the same as before: Frequency affects *I* (item memory), while concreteness and the subject manipulation affect *R* (recollection). In addition, list-order, which did not have substantial effects in the first two experiments, contributed to over-distribution. List 1 should exhibit the highest level of over-distribution. This effect was observed, and the cause was a primacy effect for item memory. Averaging across

experiments and conditions, the mean value of I was .56 on List 1, which dropped to .39 and .42 on Lists 2 and 3, respectively.

Finally, we return to the question of the relation between levels of true and false memory, which has been prominent in the larger false memory literature. As mentioned, past research on this question is limited by the fact that standard methods of measuring true and false memory, such as biased-corrected hit and false alarm rates, do not disentangle the contributions of true and false memory from those of over-distribution. Thus, it is unclear whether prior findings reflect actual relations between true and false memory or over-distribution confounds. CPD's measures of true and false memory, the R and E parameters, are not confounded with over-distribution, and we found that they were positively correlated in the first two experiments—which is consistent with the hypothesis that in this basic type of source-monitoring procedure, true and false memory both depend on recollection of details of the presentation contexts. When we computed the correlation between R and E , there was once again a strong positive relation, $r = .64$ ($df = 23$, $p < .001$).

General Discussion

We began with the question of whether over-distribution, which has recently been found to be a prime source of distortion in semantic false memory, can also be detected in source-monitoring designs. Such designs were chosen for study because source monitoring is one of the most common procedures for investigating false memory and because the underlying basis for over-distribution ought to be different than in semantic false memory. In semantic false memory, gist memory for meanings that are repeatedly instantiated by study materials (e.g., soft drink) should foment over-distribution, while verbatim memory for the targets that instantiated those meanings should suppress it: If the cue *Pepsi* retrieves a gist memory of “soft drink” from the study phase, that is consistent with the cue being a target and a related distractor, but if it retrieves a verbatim memory of the presentation of *Pepsi* on the study list, that is consistent with the cue being a target and not a related distractor. To explore this interpretation, Brainerd et al. (2010) manipulated some traditional gist and verbatim variables from the false memory literature, finding that the former increased over-distribution and the latter decreased it. Another key result, across many data sets, is that over-distribution has accounted for a far larger proportion of the responses that

are customarily interpreted as false memories (false alarms to related distractors) than false memory has (Brainerd & Reyna, 2008; Brainerd et al., 2010).

The gist route to over-distribution is not available in simple source-monitoring designs, such as the ones that we used, because study lists do not repeatedly instantiate meanings. However, an alternative mechanism, in the form of item memory, is available that could produce over-distribution, and hence, the presence of over-distribution in source-monitoring designs is an empirical question. If item memory produces over-distribution, theoretical predictions are possible as to how to control it experimentally. In particular, manipulations that inflate item memory should increase it, and manipulations that inflate recollection of the details of presentation contexts should decrease it. We incorporated some manipulations that could have such effects.

There were four outcomes of principal significance. First, over-distribution was ubiquitous in source monitoring, so we now know that it is not synonymous with a single process mechanism. Second, as in semantic false memory, over-distribution is a more common form of distortion than false memory. Referring to Table 5, across our experiments the proportion of items that were over distributed was two-and-a-half times the proportion that were falsely remembered as having been presented on the wrong list. Third, unlike semantic false memory, over-distribution was slightly more common in these experiments than true memory, the overall proportions being .26 and .20, respectively. Fourth, over-distribution was readily controlled by manipulations that should influence item memory or recollection of presentation details. We anticipated that recollection would be better for concrete than for abstract words, for University Q subjects than for University P subjects, and for later as opposed to earlier lists. Those manipulations decreased over-distribution as expected, though list-order also affected item memory. We anticipated that item memory would be better for infrequent words and for larger numbers of presentation contexts, and in the event, both manipulations increased over-distribution. For convenience, a summary of how the manipulations in our experiments affected over-distribution and how they affected the processes that are measured by the CPD model is provided in Table 6.

In sum, accumulated data invite the conclusions that over-distribution is a core form of distortion that does not depend on any one process mechanism and that it is quite amenable to experimental control by theoretically-motivated manipulations. In the remainder of this section, we briefly consider two

issues that bear on the larger significance of over-distribution—namely, what the phenomenon portends for our theoretical understanding of false memory and what the relation between recollection of contextual details and item memory portends for our theoretical understanding of over-distribution. We also consider a possible explanation of the item memory locus of word frequency effects.

Implications for Theories of False Memory

Over-distribution poses a basic question for theories of false memory: Can it be explained with the same process assumptions that are used to account for false memory? It would be preferable, of course, if additional assumptions did not have to be imported into existing theories and, instead, their current assumptions encompassed over-distribution as well as false memory. We consider this matter separately for semantic false memory and source monitoring.

Semantic false memory. The modal accounts of semantic false memory posit a semantic mechanism of some sort that foments false memory and a nonsemantic mechanism of some sort that suppresses it (for a review, see Brainerd & Reyna, 2005). One of those accounts, fuzzy-trace theory, can accommodate over-distribution for both related distractors and targets (Brainerd & Reyna, 2008). With respect to related distractors, the theory posits that processing gist traces of salient semantic content stimulates two distinct phenomenologies, phantom recollection and semantic similarity, which support false memory and over-distribution respectively. Phantom recollection is illusory vivid reinstatement of the contextual details of a distractor's prior "presentation," which is consistent with it being a target and not a distractor (false memory). Semantic similarity is the subjective impression that salient distractor meaning was very recently encountered, which is consistent with it being a target and a meaning-preserving distractor (over-distribution). Experimentation has shown that phantom recollection is a more difficult phenomenology to induce than semantic similarity because it requires both strong gist memories (e.g., relevant meanings have been instantiated by many targets) and distractors that are very good exemplars of those meanings.

Target cues (e.g., *Pepsi*), in addition to retrieving verbatim traces of their prior presentation, can retrieve verbatim traces of semantically-related targets (e.g., of the presentation of *7-Up*) or gist traces of their meaning content. If gist traces are retrieved, the type of remembering will depend on whether the associated phenomenology is semantic similarity or recollection. The former is consistent with the cue

being a target and a related distractor (over-distribution), and latter is consistent with it being a target and not a related distractor (true memory).

Summing up, the distinctions between verbatim and gist retrieval that have been used to explain semantic false memory can accommodate over-distribution of both related distractors and targets without additional assumptions. Moreover, the fact that over-distribution is more common than false memory with related distractors falls out of the fact that phantom recollective phenomenology has been found to be more difficult to produce than semantic similarity.

Source monitoring. As we have seen, gist memory for semantic content is not a viable basis for false memory or over-distribution in source monitoring because salient meanings are not repeatedly instantiated during the study phase and meaning-preserving distractors are not administered during the test phase.² Instead, the traditional theoretical concepts in source monitoring are item memory and source memory, where the former refers to remembering that a certain target was presented without being able to remember in which context, and the latter refers to remembering both that a certain target was presented and in which context (e.g., Raj & Bell, 2010). Thus, the two types of memory might be more accurately termed item memory with and without recollection of contextual details. Beginning with the Batchelder and Riefer (1990) model, this distinction has been implemented in virtually all mathematical models of source memory (for a recent review, see Klauer & Kellen, 2010). The CPD model is an exception inasmuch as it allows targets to provoke recollection of contextual details with or without item memory.

That feature of the CPD model has a theoretical advantage, too, because, as we saw, the traditional item memory/source memory distinction fails to encompass both false memory and over-distribution. If a target cue from Context A produces source memory, the cue will be remembered as originating Context A and not Context B (true memory). If it produces item memory, the subject must guess the cue's source, it could be remembered as originating in both Context A and Context B (over-distribution). However, there is no theoretical basis for remembering this cue as originating in Context B but not in Context A (false memory). In the CPD model, a target cue can produce recollection of details of the correct presentation context or recollection of details of a wrong presentation context. The former is measured by the

parameter R and produces true memory, while the latter is measured by the parameter E and produces false memory.

Thus, although the usual item memory/source memory distinction does not simultaneously encompass false memory and over-distribution, the CPD model does. Moreover, the model generates further predictions about how true memory, false memory, and over-distribution should be related to one another, which can be tested with its parameters because they provide unconfounded estimates of those processes. There are two obvious predictions. First, true and false memory ought to be positively related over experimental conditions because both are recollective processes. Second, over-distribution ought to be negatively related to both true and false memory because it is nonrecollective.

It will be recalled that although University P subjects (Experiments 1 and 3) and University Q subjects (Experiments 2 and 4) participated in identical experimental conditions, they differed in specific memory abilities that are measured by the CPD model. Therefore, the preceding predictions were tested separately for subjects from these two populations. More explicitly, the data of Experiments 1 and 3 were pooled and correlations were computed among R , E , and I estimates for University P subjects, and the data of Experiments 2 and 4 were pooled and correlations were computed among R , E , and I estimates for University Q subjects. For University P subjects, R and E estimates were positively correlated ($r = .66$, $df = 15$, $p < .001$), and the same was true for University Q subjects ($r = .42$, $df = 15$, $p < .04$). With respect to the relation between I and the other two parameters, because R and E both measure recollection of contextual details and are positively correlated, we correlated I with the total probability of recollection. That probability is given by $R + (1-R)E$ in each condition of Experiments 1 and 2, and it is given by $R + (1-R)E_i + (1-R)(1-E_i)E_j$ in each condition of Experiments 3 and 4. The relation between I and total recollection probability is shown in Figure 2 for University P subjects (panel A) and University Q subjects (Panel B), where it can be seen that the relation is negative in both instances. In sum, across the list and content manipulations that were used in our experiments, (a) to the extent that a target prompted recollection of correct contextual details, it also prompted recollection of incorrect details, and (b) to the extent that a target cue prompted memory for a particular item, it failed to produce recollection of contextual details.

In addition, it was found that in both University P and University Q subjects that estimates of the CPD model's bias parameters were negatively correlated with estimates of I ($r_s = -.43$ and $-.57$, respectively, $dfs = 15$, $ps < .04$ and $.004$) but were uncorrelated with estimates of R and E . Thus, list and content manipulations that increased item memory for individual targets (and therefore increased over-distribution) caused subjects to adopt more conservative decision criteria. However, decision criteria were not affected when manipulations enhanced recollection of correct or incorrect contextual details.

Recollecting Without Remembering

Next, we focus attention on a question that bears closely on theoretical mechanisms for over-distribution: Is it possible for subjects to recollect contextual details of target cues' prior presentations without having coincident memories of the targets themselves? Traditional models of source memory say no. As Klauer and Kellen (2010) discussed in a recent review, such models posit that when a test cue is presented for a target that appeared in Context A, memory will be in one of three states, "it is remembered that the item is old and from Source A ... it is remembered that the item is old, but memory for the source is absent ... it is not remembered that the item is old" (p. 467). In the CPD model, on the other hand, it is not required that cue-induced recollection of contextual details be accompanied by item memory for the target itself; contextual details of prior presentations can be recollected without remembering that cues are old. To clarify this feature of the model, consider the earlier expression (Equation 4) for the probability of accepting a cue for a List 1 target when the subject is asked if the cue appeared on List 1 or List 2, $p_{i,L1} = R_1 + (1-R_1)E_2 + (1-R_1)(1-E_2)I_1$. It is easy to see that this equation is a simplification of the full expression for this probability, which is $p_{i,L1} = R_1I_1 + R_1(1-I_1) + (1-R_1)E_2I_1 + (1-R_1)E_2(1-I_1) + (1-R_1)(1-E_2)I_1$. The second and fourth terms are probabilities of recollecting contextual details in the absence of item memory. Now, consider the earlier expressions (Equations 5 and 6) for the probability of accepting the same cue when subjects are asked if it appeared on List 2 or if it appeared on either List 1, $p_{E,L1} = (1-R_1)E_2 + (1-R_1)(1-E_2)I_1$ and $p_{M,L1} = R_1 + (1-R_1)E_2 + (1-R_1)(1-E_2)I_1$. These equations, too, are simplifications of the full expressions for these probabilities, which are $p_{E,L1} = (1-R_1)E_2I_1 + (1-R_1)E_2(1-I_1) + (1-R_1)(1-E_2)I_1$ and $p_{M,L1} = R_1I_1 + R_1(1-I_1) + (1-R_1)(1-E_2)I_1$. The second term in each expression indexes recollection without item memory. Thus, in the CPD model, the unconventional circumstance that

a cue can prompt recollection of contextual details even if it fails to prompt item memory of itself is the normal state of affairs.

The remaining question is whether our experiments generated any data that supply differential support for one or the other view of the relation between recollection and item memory. There are two such lines of evidence: (a) relations among parameters over experimental conditions and (b) comparative model fits. Concerning a, if retrieving recollective support for a cue requires that subjects also have item memory for the target, manipulations that increase item memory should also increase recollection, for the simple reason that the former is a necessary condition for the latter. For the same reason, third variables that covary positively or negatively with item memory ought to covary in the same manner with recollection. As we just saw, however, neither prediction is borne out in the parameter correlations for either University P subjects or University Q subjects. With respect to the relation between recollection and item memory, the data showed that, actually, they were negatively correlated over the conditions of our experiments. It is difficult to see how such a pattern could be reconciled with the hypothesis that item memory is necessary for recollection. We also saw that, contrary to this hypothesis, response bias produced a single dissociation between item memory and recollection. Specifically, variations in bias were associated with variations in item memory (bias increased as item memory decreased) but were unrelated to variations in recollection of either correct or incorrect contextual details.

Turning to the other line of evidence, comparative model fits, the CPD model delivered statistically acceptable approximations to the data of all of the present experiments. However, it is possible that a model that implements the assumption that item memory is required for recollection would also deliver acceptable fits, in which event it would be preferable by reason of containing fewer free parameters. This assumption is easily incorporated into the CDP model's expressions for the two- and three-list experiments, after which it can be refitted to the data of each experiment. To illustrate how the assumption is incorporated into the expressions, revised expressions for the two-list experiments are exhibited in the Appendix (Equations 15-20). After the assumption was incorporated, the CPD model was refitted to the data of each condition of each experiment.

The goodness-of-fit results for that analysis are displayed in Table 7, where it can be seen that the assumption that recollection of contextual details requires item memory fared poorly. Across the various

list conditions of these experiments, the value of the G^2 statistic exceeded the critical value that is required to reject the null hypothesis that the revised model fits the data by a wide margin. The only exception to this rule was the data of List 1 in Experiment 1. However, inspection of the parameter estimates for those data (Table 4) revealed that they do not actually provide a valid comparison of the revised model to the original model, owing to floor effects for recollection: The grand mean of the R_1 , E_2 , and E_3 parameters was only .05. Thus, in all list conditions in which levels of recollection were not near-floor, the hypothesis that recollection requires item memory was rejected at high levels of confidence.

That recollection can occur without item memory is surprising from the perspective of conventional ideas about the relation between these processes, but it is consistent with some source-memory experiments reported by Starns, Hicks, Brown, and Benjamin (2008) and with studies of a phenomenon that has been termed recognition without identification (e.g., Kurilla & Westerman, 2010). In the former experiments, the authors relaxed a standard design feature of source-memory research in which source judgments are only requested for cues that have been accepted as old on a prior recognition test. That feature prevents measurement of source recollection without item memory because source judgments are not requested for a cue unless item memory has already been expressed for that cue. Starns et al. requested source judgments for all cues, regardless of recognition judgments. Accurate source discrimination for unrecognized cues was found, and in most instances, the accuracy of source discrimination did not differ for unrecognized versus recognized cues. Obviously, this pattern echoes the CPD model's conception of the relation between item memory and recollection of contextual details. Turning to studies of recognition without identification, in this line of research recognition refers to source recognition, rather than old/new recognition, and identification refers to item identification. Targets are presented in two or more contexts, followed by separate tests for item identification and source recognition (see Arndt, Lee, & Flora, 2008; Ceci, Fitneva, & Williams, 2010; Cleary & Greene, 2000, 2001, 2004; Peynircioglu, 1990). The focus is then on source recognition for a subset of targets that fail the item identification test, the key outcome being that source recognition is above chance for such targets. To illustrate, Kurilla and Westerman reported some experiments in which subjects studied two lists and the same word fragments (e.g., $r_i_r_p$ for *raindrop*) were used as retrieval cues for both item identification tests (fragment completion) and source recognition tests (which asked whether the

fragments' corresponding targets had appeared on List 1 or List 2). Pooling across experiments, the mean probability of accurate source identification for fragments that failed the item identification test was well above chance and did not differ from the corresponding probability for fragments that passed the item identification test.

Beyond the obvious similarity between such results and our own findings, the paradigm that was used to generate our findings, together with the CPD model, allow the underlying processes that cause recollection without remembering to be separately measured and allow the phenomenon to be surgically controlled via manipulations that only affect specific processes. Indeed, the parameter estimates in Table 4 provide preliminary data on process-specific manipulations that elevate recollection without remembering. For example, concrete targets had this effect because they elevated recollection parameters while not affecting item memory parameters. High-frequency targets had the same effect, but because they lowered item memory parameters while not affecting recollection parameters. Also, there were individual differences in recollection without remembering because some subjects (University Q) had higher recollection parameter values than other subjects (University P), though they did not differ in their parameter values for item memory.

Word Frequency and Item Memory

Word-frequency effects have long been foci of experiments on old/new recognition, a well-known finding being that recognition is more accurate for low- than for high-frequency words (e.g., Hall, 1979). That result has been variously explained as being due to a familiarity advantage for low-frequency words (e.g., Glanzer, Kim, Hilford, & Adams, 1999) or to a recollection advantage for low-frequency words (e.g., Arndt & Reder, 2002). In source-monitoring research, on the other hand, the effects of word frequency have not been extensively investigated. Relying on the recollection hypothesis about recognition, however, Marsh, Cook, and Hicks (2006) proposed that frequency affects source memory rather than item memory and that source judgments will be more accurate for low- than for high-frequency words. Owing to the scarcity of data, however, it remains an open question whether word frequency effects are localized within source or item memory.

Our research generated considerable evidence on this question, all of which converged on item memory as the locus of frequency effects. Across the four experiments and the pilot study, two key

findings emerged. First, estimates of the item-memory parameter (I) were substantially larger for low- than for high-frequency words. Across these data sets, the grand mean of this parameter was .57 for low-frequency words versus .38 for high-frequency words. The second finding is model-free and points to the same conclusion. If item memory is elevated for low-frequency words, subjects should be (a) more likely to accept probes that name the correct source for low- than for high-frequency targets *but also* (b) more likely to accept probes that name the incorrect source for low- than for high-frequency targets. That pattern was clearly apparent. The relevant data come from probes that asked about individual list-contexts (List 1? List 2? List 3?), rather than disjunctive probes. For all the data sets, we computed bias-corrected target acceptance probabilities for such probes when they named a cue's correct source and when they named an incorrect source. When the correct source was named, the acceptance probability was higher for low- than for high-frequency words (grand means = .87 and .66, respectively), but when an incorrect source was named, the acceptance probability was also higher for low- than for high-frequency words (grand means = .38 and .25 respectively).

How is the item memory locus of word frequency to be explained? The explanation that we propose is predicated on a particular conception of the content that supports item memory—namely, memory for recency. When subjects study word lists, recency is a dimension of information that they are known to preserve—one that is relatively independent of other preserved dimensions, such as number of repetitions (for a recent review, see Hintzman, 2011). For instance, targets that are presented later in a given list or targets that are presented on later as opposed to earlier lists are judged to be more recent than targets that are presented earlier in a given list or on earlier as opposed to later lists. In our conception, memory for recency is a key component of item memory. To see why, consider again that the types of targets that are treated as instances of item memory in source monitoring research are ones for which subjects do not recollect contextual details of their presentations but that subjects are nevertheless confident were presented. Memory for recency is an obvious way that subjects can be confident that items are targets in the absence of vivid recollective phenomenology: When items provoke feelings of having been encountered during the past few minutes, that is presumably because the items were presented on study lists.

A finding that is consistent with this recency view of item memory is the fact that in our research, increasing the number of presentation contexts elevated item memory. Because presentation contexts were successive, the distinctive features of each list (color/font) were distinctive cues to recency (e.g., words printed in Broadway font were presented before words printed in script.) To the extent that memory for recency is aided by distinctive signposts that segment the temporal stream (see Huttenlocher, Hedges, & Prohaska, 1992), increasing the number of such signposts should improve forms of performance that depend on memory for recency, item memory in this instance.

How does the recency interpretation of item memory fit with the item memory locus of frequency effects? If that interpretation is correct, independent evidence should show that word frequency affects memory for recency in the same manner that it affects item memory. It does. The effects of word frequency on judgments of recency have been studied—for example by Chalmers (2002) and Hintzman (2003). If item memory depends on memory for recency, low-frequency targets should provoke stronger feelings of recency and better recency discrimination than high-frequency targets. This is the pattern that has been reported (e.g., Chalmers, 2002).

Concluding Comment

The psychological significance of over-distribution is that false memory is not as false and true memory is not as true as conventional designs lead us to suppose. Episodic attributions that are objectively false (“*bagpipe* was on List 2”) may not represent false memories, and attributions that are objectively true (“*ocean* was on List 1”) may not represent true memories—the reason being that retrieval can assign the same events to additional episodic states with contrasting truth values. To decide between these possibilities, a specific type of further information is required—specifically, whether events that are attributed to false episodic states are also attributed to true ones (“*bagpipe* was on List 1”) and whether events that are attributed to true episodic states are also attributed to false ones (“*ocean* was on List 2”). Those data tell a surprising tale. On the false memory side, the types of responses that have traditionally been regarded as false memories are found to be predominantly over-distribution errors. On the true memory side, the type of retrieval that separates true memory from over-distribution, recollection of contextual details, does not seem to require that the events themselves be remembered.

Last, it would be natural to entertain the hypothesis that over-distribution might be predicted by the notion of a continuum of memory strength/confidence that is so familiar from research on recognition memory, especially from signal detection models of recognition. Thus, for a Source A target, one could think, in the usual signal detection manner, of a distribution of strength/confidence values that the item belongs to Source A and of another distribution of strength/confidence values that the item belongs Source B. Under this conception, there would be some average probability of accepting Source A probes and also some average (but smaller) probability of accepting Source B probes. Although this may seem like over-distribution (because probes for both sources will be accepted for some items), it is not: Over-distribution refers to the relation between *the sum* of these two acceptance probabilities and the probability of accepting the disjunctive probe “Source A or Source B?”—specifically, that the relation will be subadditive. When Brainerd and Reyna (2008) analyzed the strength/confidence conception, they found that it did not predict subadditivity.

References

- Arndt, J., Lee, K., & Flora, D. B. (2008). Recognition without identification for words, pseudowords and nonwords. *Journal of Memory and Language, 59*, 346-360.
- Arndt, J., & Reder, L. M. (2002). Word frequency and receiver operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 830-842.
- Barnhardt, T. M., Choi, H., Gerkens, D. R., & Smith, S. M. (2006). Output position and word relatedness effects in a DRM paradigm: Support for a dual-retrieval process theory of free recall and false memories. *Journal of Memory and Language, 55*, 213–231.
- Barrouillet, P. (in press). Dual-process theories of reasoning: The test of development. *Developmental Review*.
- Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*, 548-564.
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial processing tree modeling. *Psychonomic Bulletin & Review, 6*, 57-86.
- Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. New York: Oxford University Press.
- Brainerd, C. J., & Reyna, V. F. (2008). Episodic over-distribution: A signature effect of recollection without familiarity. *Journal of Memory and Language, 58*, 765-786.
- Brainerd, C. J., Reyna, V. F., & Aydin, C. (2010). Disjunction fallacies in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 711–735.
- Brainerd, C. J., Reyna, V. F., & Mojardin, A. (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., Yang, Y., Toglia, M. P., Reyna, V. F., & Stahl, C. (2008, November). *Emotion and false memory: The Cornell/Cortland norms*. Paper presented at the annual meeting of the Psychonomic Society, Chicago, IL.

- Brenner, L., & Rottenstreich, Y. (1999). Focus, repacking, and the judgment of disjunctive hypotheses. *Journal of Behavioral Decision Making*, *12*, 141–148.
- Buchner, A., Erdfelder, E., Steffens, M. C., Martensen, H. (1997). The nature of memory processes underlying recognition judgments in the process dissociation procedure. *Memory & Cognition*, *25*, 508-517.
- Budson, A. E., Todman, R. W., Chong, H., Adams, E. H., Kensinger, E. A., Krangel, T. S., & Wright, C. I. (2006). False recognition of emotional word lists in aging and Alzheimer disease. *Cognitive and Behavioral Neurology*, *19*, 71-78.
- Chalmers, K. A. (2002). Word frequency effects in episodic judgments of recency and frequency. *Australian Journal of Psychology*, *54*, 49.
- Ceci, S. J., Fitneva, S. A., & Williams, W. M. (2010). Representational constraints on the development of memory and metamemory: A developmental-representational theory. *Psychological Review*, *117*, 464-495.
- Cleary, A. M., & Greene, R. L. (2000). Recognition without identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1063-1069.
- Cleary, A. M., & Greene, R. L. (2001). Memory for unidentified items: Evidence for the use of letter information in familiarity. *Memory & Cognition*, *29*, 540-545,
- Cleary, A. M., & Greene, R. L. (2004). True and false memory in the absence of perceptual identification. *Memory*, *12*, 231-236.
- Cronbach, L. H. (1970). How should be measure change – Or should we? *Psychological Bulletin*, *74*, 68-80.
- Deese, J. (1959). On the prediction of occurrence of certain verbal intrusions in free recall. *Journal of Experimental Psychology*, *58*, 17–22.
- Fox, C. R., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science*, *44*, 879–895.
- Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 5-16.
- Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in

- recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 500-513.
- Gorman, A. N. (1961). Recognition memory for names as a function of abstractness and frequency. *Journal of Experimental Psychology*, 39, 950–958.
- Hall, J. F. (1979). Recognition as a function of word frequency. *American Journal of Psychology*, 92, 497–505.
- Heaps, C. M., & Nash, M. (2001). Comparing recollective experience in true and false autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 920-930.
- Hicks, J. L., & Marsh, R. L. (2001). False recognition occurs more frequently during source identification than during old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 375-383.
- Hicks, J. L., & Starns, J. J. (2006). Remembering source evidence from associatively related items: Explanations from a global matching model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 32, 1164-1173.
- Hintzman, D. L. (2003). Judgments of recency and their relation to recognition memory. *Memory & Cognition*, 31, 26-34.
- Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the “coordinates of truth.” *Perspectives on Psychological Science*, 6, 253-271.
- Hu, X. (1998). General processing tree (Version 1.0) [Computer software]. Memphis, TN: University of Memphis.
- Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, 32, 421-445.
- Huttenlocher, J., Hedges, L. V., & Prohaska, V. (1992). Memory for day of the week: a 5 + 2 cycle. *Journal of Experimental Psychology: General*, 121, 313-325.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513-541.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28.

- Klauer, K. C., & Kellen, D. (2010). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review*, *17*, 465-478.
- Kucera, H., & Francis, W. (1967). Computational analysis of present day American English. Providence, RI: Brown University Press.
- Kurilla, B. P., & Westerman, D. L. (2010). Source memory for unidentified stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 398-410.
- Lampinen, J. M., Odegard, T. N., Blackshear, E., & Toglia, M. P. (2005). Phantom ROC. In F. Columbus (Ed.), *Progress in experimental psychology research* (pp. 235–267). Hauppauge, NY: Nova.
- Loftus, E. F. (1975). Leading questions and eyewitness report. *Cognitive Psychology*, *7*, 560-572.
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 19-31.
- Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, *63*, 46-63.
- Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). The effect of context variability on source memory. *Memory & Cognition*, *34*, 1578-1586.
- Odegard, T. N., & Lampinen, J. M. (2005). Recollection rejection: Gist cuing of verbatim memory. *Memory & Cognition*, *33*, 1422–1430.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*, 241-259.
- Park, H., Arndt, J., & Reder, L. M. (2006). A contextual interference account of distinctiveness effects in recognition. *Memory & Cognition*, *34*, 743-751.
- 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.
- Peynircioglu, Z. F. (1990). A feeling-of-recognition without identification. *Journal of Memory and Language*, *29*, 493-500.
- Raj, V., & Bell, M. A. (2010). Cognitive processes supporting episodic memory formation in childhood: The role of source memory, binding, and executive functioning. *Developmental Review*, *30*, 384-

402.

- Reyna, V. F. (1995). Interference effects in memory and reasoning: A fuzzy-trace theory analysis. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 29-59). San Diego, CA: Academic Press.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1-75.
- Reyna, V. F., & Brainerd, C. J. (in press). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review*.
- Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9*, 95-123.
- Reyna, V. F., & Titcomb, A. (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* (pp. 157-174). Hillsdale, NJ: Erlbaum.
- Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95*, 318–339.
- Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented on lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.
- Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review, 101*, 547-567.
- Singer, M., & Remillard, G. (2008). Veridical and false memory for text: A multi-process analysis. *Journal of Memory and Language, 59*, 18–35.
- Sloman, S., Rottenstreich, Y., Wisniewski, E., Hadjichristidis, C., & Fox, C. R. (2004). Typical versus atypical unpacking and superadditive probability judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 573–582.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34-50.
- Stahl, C., & Klauer, K. C. (2008). Validation of a simplified conjoint recognition paradigm for the

- measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 570–588.
- Stahl, C., & Klauer, K. C. (2009). Measuring phantom recollection in the simplified conjoint recognition paradigm. *Journal of Memory and Language*, 60, 180–193.
- Starns, J. J., Hicks, J. J., Brown, N. L., & Martin, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition*, 36, 1-8.
- Toglia, M. P., & Battig, W. F. (1978). Handbook of semantic word norms. Hillsdale, NJ: Lawrence Erlbaum.
- Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, 101, 547–567.
- Vogt, V., & Broder, A. (2007). Independent retrieval of source dimensions: An extension of results by Starns and Hicks (2005) and a comment on the ACSIM measure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 443-450.

Appendix

The CPD model expresses performance on memory probes (cues + episodic descriptions) as functions of three memory-state parameters and two bias parameters. Consider an experiment in which subjects encode events in two physically distinctive contexts, List 1 and List 2. On a subsequent memory test, three types of cues (items presented on List 1, items presented in List 2, and distractors) are factorially crossed with three types of episodic descriptions (Presented on List 1? Presented on List 2? Presented on List 1 or List 2?). Thus, there are nine distinct types of probes. For an item that was presented List 1, let $p_{M,L1}$ be the probability of accepting a probe that describes such an item as having been presented on List 1, let $p_{E,L1}$ be the probability of accepting a probe that describes such an item as having been presented on List 2, and let $p_{I,L1}$ be the probability of accepting a probe that describes such an item as having been presented on either List 1 or List 2. For distractors, let $p_{M,\emptyset}$ be the probability of accepting a probe that describes such an item as having been presented on List 1, let $p_{E,\emptyset}$ be the probability of accepting a probe that describes such an item as having been presented on List 2, and let $p_{I,\emptyset}$ be the probability of accepting a probe that describes such an item as having been presented on List 1 or List 2.

For a target that was presented on List 1, R_1 is the probability that it is in the true memory state (provokes recollection of List 1 contextual details), E_1 is the probability that it is in the false memory state (provokes recollection of List 2 contextual details), and I_1 is the probability that it is in the over-distribution state (provokes item memory without recollection of contextual details). For a distractor, b is the probability that it is accepted when a probe asks if was presented on List 1 and when a probe asks if it was presented on List 2, and b_{1U2} is the probability that it is accepted when a probe asks if it was presented on either List 1 or List 2. The model's expressions for the above empirical probabilities, then, are

$$p_{I,L1} = R_1 + (1-R_1)E_2 + (1-R_1)(1-E_2)I_1 + (1-R_1)(1-E_2)(1-I_1)b_{1U2}, \quad (\text{A1})$$

$$p_{E,L1} = (1-R_1)E_2 + (1-R_1)(1-E_2)I_1 + (1-R_1)(1-E_2)(1-I_1)b, \quad (\text{A2})$$

$$p_{M,L1} = R_1 + (1-R_1)E_2 + (1-R_1)(1-E_2)I_1 + (1-R_1)(1-E_2)(1-I_1)b, \quad (\text{A3})$$

$$p_{I,\emptyset} = b_{1U2}, \quad (\text{A4})$$

$$p_{E,\emptyset} = b, \text{ and} \quad (\text{A5})$$

$$p_{M,\emptyset} = b. \quad (A6)$$

For application to experimental data, estimates can be obtained for the three memory parameters and the two bias parameters, goodness-of-fit tests can be conducted, and within- and between-condition significance tests of parameter values can be conducted. This is done by implementing Equations A1-A6 in a multinomial modeling program, such as GPT (Hu, 1998). Finally, for items that are presented on List 2 rather than List 1, a set of expressions that parallel Equations A1-A6 can be obviously written, from which parameter estimates can be obtained and goodness-of-fit tests and parameter significance tests can be conducted.

Next, consider a memory experiment in which subjects encode events in three physically distinctive contexts, List 1, List 2, and List 3. On a memory test that follows presentation of the target events, four types of cues (items presented on List 1, items presented on List 2, items presented on List 3, and distractors) are factorially crossed with four types of episodic descriptions (Presented on List 1? Presented on List 2? Presented on List 3? Presented in either List 1 or List 2 or List 3?). Thus, there are now 16 distinct probes, rather than 9. For an item that was presented in context A, there are four types of probes with associated empirical probabilities: $p_{i,L1}$, $p_{E2,L1}$, $p_{E3,L1}$, and $p_{M,L1}$. For distractors, there are also four types of probes with associated empirical probabilities: $p_{i,\emptyset}$, $p_{E2,\emptyset}$, $p_{E3,\emptyset}$, and $p_{M,\emptyset}$. The CPD model's expressions for these eight empirical probabilities are

$$p_{i,L1} = R_1 + (1-R_1)E_2 + (1-R_1)(1-E_2)E_3 + (1-R_1)(1-E_2)(1-E_3)I_1 + (1-R_1)(1-E_2)(1-E_3)(1-I_1)b_{1U2U3}, \quad (A7)$$

$$p_{E2,L1} = (1-R_1)E_2 + (1-R_1)(1-E_2)(1-E_3)I_1 + (1-R_1)(1-E_2)(1-E_3)(1-I_1)b, \quad (A8)$$

$$p_{E3,L1} = (1-R_1)(1-E_2)E_3 + (1-R_1)(1-E_2)(1-E_3)I_1 + (1-R_1)(1-E_2)(1-E_3)(1-I_1)b, \quad (A9)$$

$$p_{M,L1} = R_1 + (1-R_1)(1-E_2)(1-E_3)I_1 + (1-R_1)(1-E_2)(1-E_3)(1-I_1)b, \quad (A10)$$

$$p_{i,\emptyset} = b_{1U2U3}, \quad (A11)$$

$$p_{E2,\emptyset} = b, \quad (A12)$$

$$p_{E3,\emptyset} = b, \text{ and} \quad (A13)$$

$$p_{M,\emptyset} = b. \quad (A14)$$

The key difference between this model and the model for two contexts is that there are now two false memory parameters because a presented item can occupy either of two false memory states. E_2 is the probability that an item that was presented on List 1 is falsely remembered as having been

presented on List 2. E_3 is the probability that an item that was presented on List 1 is falsely remembered as having been presented on List 3. For application to experimental data, Equations A7-A14 (like Equations A1-A6) are simply implemented in a multinomial modeling program, which estimates parameters, conducts goodness-of-fit tests, and computes within- and between-condition significance tests of parameter values. Finally, for items that are presented on List 2 rather than on List 1 or List 3 and for items that are presented on List 3 rather than on List 1 or List 2, it is obvious that a set of expressions that parallel Equations A7-A10 can be written, from which parameter estimates can be obtained and goodness-of-fit tests and parameter significance tests can be conducted.

The CPD model does not assume that item memory, as measured by the I parameter, is a necessary precondition for recollecting contextual details of either correct or incorrect lists, as measured by the R and E parameters, respectively. However, that constraint is easily imposed on the model. For instance, consider again an experiment in which subjects encode events in two physically distinctive contexts, List 1 and List 2. Keeping all the details of this experiment the same and preserving the notation in Equations A1-A6, if the assumption that item memory is necessary for recollection of contextual details is imposed on these expressions, the revised CPD model for List 1 becomes

$$p_{I,L1} = I_1 + (1-I_1)b_{1U2}, \quad (\text{A15})$$

$$p_{E,L1} = I_1(1-R_1) + (1-I_1)b, \quad (\text{A16})$$

$$p_{M,L1} = I_1 + (1-I_1)b, \quad (\text{A17})$$

$$p_{I,\emptyset} = b_{1U2}, \quad (\text{A19})$$

$$p_{E,\emptyset} = b, \text{ and} \quad (\text{A20})$$

$$p_{M,\emptyset} = b. \quad (\text{A21})$$

Statistical analysis of this revised model proceeds in the same manner as with the original model. That is, the revised model is implemented in a multinomial modeling program. Then, parameter estimates, goodness-of-fit tests, and within- and between-condition significance tests can all be generated for sample data.

Last, consider the alternative Batchelder and Riefer (1990) source-memory model for our experiments. The relevant parameters of this model are presented in Table 2. The model's expressions

for the two-list procedure of Experiments 1 and 2 (i.e., the expressions that parallel Equations A1-A6 of CPD) are:

$$p_{i,L1} = D_1 + (1-D_1)g, \quad (\text{A22})$$

$$p_{E,L1} = D_1(1-d_1)g, \quad (\text{A23})$$

$$p_{M,L1} = D_1d_1 + D_1(1-d_1)g, \quad (\text{A24})$$

$$p_{i,\emptyset} = (1-D_N)b \quad (\text{A25})$$

$$p_{E,\emptyset} = (1-D_N)b, \text{ and} \quad (\text{A26})$$

$$p_{M,\emptyset} = (1-D_N)b. \quad (\text{A27})$$

Application of this model to experimental data proceeds in the same manner as application of the CPD model: Estimates are obtained for the five parameters, goodness-of-fit tests are conducted, and within- and between-condition significance tests of parameter values are conducted by implementing Equations A22-A27 in a multinomial modeling program. It also a simple matter to construct the source-memory model for the three-list designs of Experiments 3 and 4 that corresponds to the CPD expressions in Equations A7-A14. The expressions for List 1, which parallel Equations A7-A14 for CPD, are:

$$p_{i,L1} = D_1 + (1-D_1)g, \quad (\text{A28})$$

$$p_{E2,L1} = D_1(1-d_1)g, \quad (\text{A29})$$

$$p_{E3,L1} = D_1(1-d_1)g, \quad (\text{A30})$$

$$p_{M,L1} = D_1d_1 + D_1(1-d_1)g, \quad (\text{A31})$$

$$p_{i,\emptyset} = (1-D_N)b, \quad (\text{A32})$$

$$p_{E2,\emptyset} = (1-D_N)b, \quad (\text{A33})$$

$$p_{E3,\emptyset} = (1-D_N)b, \text{ and} \quad (\text{A34})$$

$$p_{M,\emptyset} = (1-D_N)b. \quad (\text{A35})$$

Author Note

C. J. Brainerd, V. F. Reyna, and K. Nakamura, Department of Human Development, Cornell University. R. E. Holliday, Department of Psychology, University of Leicester.

Portions of the research were supported by a grant to the first and second authors from the National Institutes of Health (1RC1AG036915-01). The PC-based software that was used to conduct the modeling analyses that are reported in this paper is available from the first author upon request.

Correspondence concerning this article should be addressed to C. J. Brainerd, Department of Human Development, Cornell University, Ithaca, NY 14853. E-mail: cb299@cornell.edu

Footnote

¹We are grateful to K. C. Klauer for drawing this point to our attention.

²The proviso should be added that although source-monitoring designs need not have either of these features, they often do. As Reyna Lloyd (1997) discussed in a review of source monitoring research, it is often the case that there are salient meaning connections among the targets that are presented in Context A and among the targets that are presented in Context B, so that when these items are presented as test cues, subjects can retrieve gist traces of those meanings and use them to make source judgments. For instance, Context A might consist of four DRM lists presented in a male voice (say, the *cold*, *music*, *smoke*, and *window* lists), and Context B might consist of four other DRM lists presented in a female voice (say, the *chair*, *doctor*, *sleep*, and *trash* lists). Then, when subjects are asked if they heard *piano* in a male voice and if they heard *hospital* in a female voice, they can respond via the gist memories that musical meaning is associated with the male voice while medical meaning is associated with the female voice (see Hicks & Marsh, 2001; Hicks & Starns, 2006).

Table 1

Operational Distinction between Over-Distribution and Semantic False Memory

	Memory state		
	False memory	Over-distribution	Forgotten
<i>Coke</i> : target?	Yes	Yes	No
<i>Coke</i> : new but related to target?	No	Yes	No
<i>Coke</i> : target or new but related to target?	Yes	Yes	No

Note. *Coke* is a semantically-related distractor on a test list, following a study list on which *Pepsi*, *7-Up*, and *Sprite* were targets.

Table 2

Process Definitions of the Parameters of Batchelder and Riefer's (1990) Source-Memory Model, for Experiments 1-4

Parameter	Process definition
D_1	For words that were presented on List 1, this is the probability that subjects have item memory for those words.
d_1	For words that were presented on List 1 for which subjects have item memory, this is the probability that subjects can recollect that the words were presented on List 1.
g	For words that were presented on List 1 for which subjects do not have item memory, this is the probability that subjects guess that the words were presented on whichever list is mentioned in the test probe.
D_N	For words that were not presented on either list (distractors), this is the probability that subjects judge that the words are new.
b	For words that were not presented on either list (distractors) that subjects judge to be old, this is the probability that subjects judge that the words were presented on whichever list is mentioned in the test probe.

Table 3

Bias-Corrected Acceptance Probabilities and Over-Distribution Statistics for Experiments 1-4

List-context/statistic	Word content			
	Hi Frequency Concrete	Hi Frequency Abstract	Lo Frequency Concrete	Lo Frequency Abstract
Experiment 1				
List 1:				
$p(L1?)$.32	.22	.43	.21
$p(L2?)$.14	.20	.32	.42
$p(L1 \text{ or } L2?)$.34	.22	.51	.43
Over-distribution	.12	.20	.24	.20
List 2:				
$p(L1?)$.16	.06	.21	.21
$p(L2?)$.23	.22	.38	.42
$p(L1 \text{ or } L2?)$.45	.28	.42	.23
Over-distribution	-.06	.00	.17	.40
Experiment 2				
List 1:				
$p(L1?)$.25	.21	.37	.29
$p(L2?)$.12	.03	.34	.37
$p(L1 \text{ or } L2?)$.41	.11	.66	.50
Over-distribution	-.04	.13	.05	.16
List 2:				
$p(L1?)$.04	.07	.16	.14
$p(L2?)$.30	.23	.42	.41
$p(L1 \text{ or } L2?)$.39	.25	.49	.34
Over-distribution	-.05	-.05	.09	.11
Experiment 3				
List 1:				
$p(L1?)$.58	.60	.65	.65
$p(L2?)$.41	.43	.64	.72
$p(L3?)$.42	.68	.56	.65
$p(L1 \text{ or } L2 \text{ or } L3?)$.56	.59	.65	.75
Over-distribution	.85	1.12	1.20	1.27
List 2:				
$p(L1?)$.36	.41	.37	.51
$p(L2?)$.42	.45	.50	.60
$p(L3?)$.41	.43	.51	.44
$p(L1 \text{ or } L2 \text{ or } L3?)$.63	.75	.51	.68
Over-distribution	.56	.54	.87	.87
List 3:				
$p(L1?)$.31	.43	.39	.43
$p(L2?)$.50	.60	.55	.65
$p(L3?)$.36	.49	.55	.56
$p(L1 \text{ or } L2 \text{ or } L3?)$.64	.55	.65	.65
Over-distribution	.53	.97	.84	.97

Table 3—*continued*

Experiment 4				
List 1:				
$p(L1?)$.47	.38	.55	.34
$p(L2?)$.17	.00	.33	.45
$p(L3?)$.52	.20	.40	.35
$p(L1 \text{ or } L2 \text{ or } L3?)$.48	.14	.72	.52
Over-distribution	.68	.44	.56	.62
List 2:				
$p(L1?)$.05	.04	.25	.17
$p(L2?)$.32	.25	.49	.39
$p(L3?)$.25	.09	.30	.23
$p(L1 \text{ or } L2 \text{ or } L3?)$.51	.40	.65	.52
Over-distribution	.11	-.02	.39	.27
List 3:				
$p(L1?)$.14	.02	.14	.35
$p(L2?)$.23	.32	.31	.44
$p(L3?)$.39	.21	.53	.44
$p(L1 \text{ or } L2 \text{ or } L3?)$.58	.31	.64	.55
Over-distribution	.28	.24	.34	.68

Table 4

Estimates of the Parameters of the CPD Model for Experiments 1-4

List-context/parameter	Word content			
	Hi Frequency Concrete	Hi Frequency Abstract	Lo Frequency Concrete	Lo Frequency Abstract
Experiment 1				
List 1:				
<i>R</i>	.22	.19	.23	.07
<i>E</i>	.00	.12	.16	.12
<i>I</i>	.29	.26	.46	.53
<i>b</i>	.23	.37	.18	.30
<i>b_{1U2}</i>	.24	.42	.22	.28
List 2:				
<i>R</i>	.29	.25	.23	.13
<i>E</i>	.34	.22	.07	.00
<i>I</i>	.15	.17	.35	.38
<i>b</i>	.23	.42	.18	.30
<i>b_{1U2}</i>	.21	.33	.22	.28
Experiment 2				
List 1:				
<i>R</i>	.29	.23	.34	.14
<i>E</i>	.20	.00	.49	.19
<i>I</i>	.20	.10	.43	.50
<i>b</i>	.26	.39	.15	.25
<i>b_{1U2}</i>	.25	.47	.18	.23
List 2:				
<i>R</i>	.33	.22	.34	.19
<i>E</i>	.15	.19	.10	.00
<i>I</i>	.16	.20	.29	.37
<i>b</i>	.26	.38	.15	.26
<i>b_{1U2}</i>	.25	.48	.22	.22
Experiment 3				
List 1:				
<i>R</i>	.14	.11	.05	.01
<i>E₂</i>	.00	.00	.04	.08
<i>E₃</i>	.00	.13	.00	.01
<i>I</i>	.37	.43	.59	.63
<i>b</i>	.17	.28	.13	.22
<i>b_{1U2U3}</i>	.23	.33	.14	.23
List 2:				
<i>R</i>	.12	.17	.07	.14
<i>E₁</i>	.07	.15	.00	.09
<i>E₃</i>	.14	.21	.08	.00
<i>I</i>	.30	.31	.39	.45
<i>b</i>	.17	.28	.13	.22
<i>b_{1U2U3}</i>	.23	.34	.13	.23

Table 4—continued

List 3:				
<i>R</i>	.08	.00	.12	.07
<i>E</i> ₁	.04	.00	.00	.00
<i>E</i> ₂	.25	.10	.13	.17
<i>I</i>	.30	.32	.47	.48
<i>b</i>	.17	.28	.13	.22
<i>b</i> _{1U2U3}	.23	.32	.14	.22
Experiment 4				
List 1:				
<i>R</i>	.19	.39	.31	.05
<i>E</i> ₂	.00	.00	.11	.16
<i>E</i> ₃	.28	.30	.17	.07
<i>I</i>	.73	.49	.70	.57
<i>b</i>	.19	.32	.11	.19
<i>b</i> _{1U2U3}	.21	.41	.15	.24
List 2:				
<i>R</i>	.31	.27	.32	.27
<i>E</i> ₁	.00	.15	.13	.09
<i>E</i> ₃	.27	.20	.15	.17
<i>I</i>	.32	.35	.53	.44
<i>b</i>	.19	.32	.11	.19
<i>b</i> _{1U2U3}	.21	.41	.15	.24
List 3:				
<i>R</i>	.30	.11	.34	.12
<i>E</i> ₁	.06	.00	.00	.05
<i>E</i> ₂	.27	.24	.25	.13
<i>I</i>	.45	.34	.40	.62
<i>b</i>	.19	.32	.11	.19
<i>b</i> _{1U2U3}	.21	.41	.15	.24

Note. The values for the bias parameter *b* in each condition are the raw false alarm rates for the L1? and L2? judgments in Experiments 1 and 2 and for the L1?, L2?, and L3? judgments in Experiments 3 and 4. The values for the bias parameter *b*_{1U2} of Experiments 1 and 2 are the raw false alarm rates for L1UL2? judgments, while the values for the bias parameter *b*_{1U2U3} of Experiments 3 and 4 are the raw false alarm rates for L1UL2UL3?

Table 5

Proportions of True Memory, False Memory, Over-Distribution, and Forgetting Estimated by the CPD Model

List-context/memory state	Word content			
	Hi Frequency Concrete	Hi Frequency Abstract	Lo Frequency Concrete	Lo Frequency Abstract
Experiment 1				
List 1:				
True	.22	.19	.23	.07
False	.00	.10	.12	.11
Over-distribution	.23	.19	.31	.44
Forgetting	.55	.52	.34	.38
List 2:				
True	.29	.25	.23	.13
False	.24	.15	.05	.00
Over-distribution	.08	.11	.27	.33
Forgetting	.39	.49	.45	.54
Experiment 2				
List 1:				
True	.29	.23	.34	.14
False	.14	.00	.32	.16
Over-distribution	.09	.08	.19	.36
Forgetting	.48	.69	.15	.34
List 2:				
True	.33	.22	.34	.19
False	.10	.15	.07	.00
Over-distribution	.10	.13	.18	.30
Forgetting	.47	.40	.41	.51
Experiment 3				
List 1:				
True memory	.14	.11	.05	.01
False memory (list 2)	.00	.00	.04	.08
False memory (list 3)	.00	.12	.00	.01
Over-distribution	.32	.33	.54	.57
Forgetting	.52	.44	.37	.33
List 2:				
True memory	.12	.17	.07	.14
False memory (list 1)	.06	.13	.00	.12
False memory (list 3)	.12	.15	.07	.00
Over-distribution	.21	.17	.33	.35
Forgetting	.49	.38	.53	.39

Table 5—*continued*

List 3:					
True memory	.08	.00	.12	.07	
False memory (list 1)	.04	.00	.00	.00	
False memory (list 2)	.22	.10	.11	.16	
Over-distribution	.23	.29	.36	.37	
Forgetting	.43	.61	.41	.40	
Experiment 4					
List 1:					
True memory	.19	.39	.31	.05	
False memory (list 2)	.00	.00	.08	.15	
False memory (list 3)	.23	.18	.10	.06	
Over-distribution	.43	.21	.36	.42	
Forgetting	.15	.22	.25	.32	
List 2:					
True memory	.31	.27	.32	.27	
False memory (list 1)	.00	.11	.09	.07	
False memory (list 3)	.19	.12	.09	.11	
Over-distribution	.16	.17	.27	.24	
Forgetting	.34	.33	.23	.31	
List 3:					
True memory	.30	.11	.34	.12	
False memory (list 1)	.05	.00	.00	.04	
False memory (list 2)	.18	.21	.17	.11	
Over-distribution	.22	.23	.20	.45	
Forgetting	.25	.45	.29	.28	

Table 6

How the Manipulations in Experiments 1-4 Affected Over-Distribution and the Processes Measured by the CPD Model

Manipulation	Type of Effect
Frequency	Direct effect: Over-distribution levels were higher for low-frequency words, which increased parameter I .
Number of Contexts	Direct effect: Over-distribution levels were higher for three presentation contexts, which increased parameter I .
List-order	Direct effect: With three lists, over-distribution levels were higher for List 1 than for Lists 2 and 3. Parameter I was larger for List 1 than for Lists 2 and 3.
University P vs. Q	Indirect effect: Over-distribution levels were higher for University P subjects because the R parameter was smaller for those subjects.
Concreteness	Indirect effect: Over-distribution levels were higher for abstract words, which decreased parameter R .

Table 7

Observed Values of the G^2 Fit Statistic for the Standard Model of the Relation between Recollection and Item Memory

List	Experiment			
	1	2	3	4
List 1	393.13 ^a	471.33 ^a	21.20 ^b	79.16 ^b
List 2	398.32 ^a	415.49 ^a	62.87 ^b	56.27 ^b
List 3			108.73 ^b	69.10 ^b

^aThe critical value of the G^2 statistic to reject the null hypothesis that the model fits the data is 31.28.

^bThe critical value of the G^2 statistic to reject the null hypothesis that the model fits the data is 37.96.

Figure Captions

Figure 1. Results of the pilot study of source recognition of high- and low-frequency words. The bias-corrected acceptance probabilities for probes that ask whether a target cue appeared on the correct list (i.e., the list on which it was presented) and for probes that ask whether a target cue appeared on the wrong list (i.e., the list on which it was not presented) are displayed in Panel A. Estimates of the R (true recollection), E (erroneous recollection), and I (item memory) parameters for these same data are displayed in Panel B.

Figure 2. Relation between item memory and recollection of contextual details in University P subjects (Panel A) and University Q subjects (Panel B).



