Published in final edited form as:

J Exp Psychol Learn Mem Cogn. 2014 November ; 40(6): 1610–1623. doi:10.1037/xlm0000017.

Conjunction Illusions and Conjunction Fallacies in Episodic Memory

C. J. Brainerd,

Department of Human Development, Cornell University

R. E. Holliday,

School of Psychology, University of Leicester

K. Nakamura, and

Department of Human Development, Cornell University

V. F. Reyna

Department of Human Development, Cornell University

Abstract

Recent research on the overdistribution principle implies that episodic memory is infected by conjunction illusions. These are instances in which an item that was presented in a single context (e.g., List 1) is falsely remembered as having been presented in multiple contexts (e.g., List 1 and List 2). Robust conjunction illusions were detected in source-monitoring designs in which conjunctive probes (Was *bagpipe* presented on List 1 and List 2?) were added to the traditional nonconjunctive probes (Was *bagpipe* presented on List 1?). In Experiment 1, the levels of those illusions were comparable to what would be expected on the basis of results from prior overdistribution experiments. In Experiments 2 and 3, conjunction illusions were neither eliminated nor reduced by a manipulation that should have had such effects if the illusions are by-products of subjective differences in retrieved memory support. Also, conjunction illusions sometimes rose to the level of conjunction fallacies: In certain conditions, subjects thought that items were more likely to have occurred in all the presentation contexts than in any single context, which is impossible. Two general approaches to explaining overdistribution, representational accounts and retrieval accounts, are considered.

Keywords

episodic overdistribution; conjunction illusions; memory disjunction fallacies; dual-trace theories

According to the overdistribution principle, episodic memory takes too broad a view of the scope of experience by mapping the events of our lives with too many episodic states—including, dramatically, states that are logically or experientially incompatible with one another. False-memory experiments illustrate memory for states that are logical incompatible, and source-monitoring experiments illustrate memory for states that are experientially incompatible. In false memory experiments, certain findings suggest that

depending upon the probe that is administered, a studied target can be remembered as being both a target and a distractor and that a distractor can be remembered as being both a distractor and a target (Brainerd, Reyna, & Aydin, 2010). In source-monitoring experiments, related findings suggest that depending upon the probe that is administered, a target that only occurred in Context A can be remembered as occurring in both Context A and Context B and that a target that only occurred in Context B can be remembered as occurring in both Context B and Context A (Brainerd, Reyna, Holliday, & Nakamura, 2012).

The psychological significance of overdistribution is that our memory for events' episodic status seems to be an emergent property of testing, whereas it is customarily viewed as a property of prior experience. By this, we mean that prior to the administration of a probe, some events that objectively belong to a single episodic state on the basis of prior experience have the potential to be remembered as belonging to other states as well, even ones that are logically or experientially incompatible with their actual state. This is a memory analogue of the superposition phenomenon in quantum probability (e.g., Busemeyer, Wang, & Townsend, 2006) because it appears that prior to the administration of a probe, memory for some events is distributed over a set of mutually incompatible episodic states; that all of those states have nonzero probabilities (Brainerd, Wang, Reyna, 2013). Such events' remembered states are determined by how their states are described in memory probes. In false memory experiments, some targets will be remembered as targets when probes ask whether they are targets but, alternatively, as related distractors when probes ask whether they are related distractors, and conversely, some related distractors will be remembered as related distractors when probes ask whether they are related distractors but, alternatively, as targets when probes ask whether they are targets (Brainerd & Reyna, 2008). In source-monitoring experiments, some Context A targets will be remembered as Context A when probes ask whether they occurred in that context but, alternatively, as Context B when probes ask whether they occurred in that context, and conversely for Context B targets (Brainerd et al., 2012).

The aim of the experiments that are reported in this article was to investigate an important uncertainty in the empirical basis of the overdistribution principle. As discussed in the first section, below, overdistribution is currently measured via a subadditivity effect—specifically, the fact that the sum of the probabilities of remembering a test item as belonging to *each* of a series of incompatible episodic states (e.g., *Bagpipe* was on the study list. *Bagpipe* was not on the list but is related to a word on the list.) is greater than the probability of remembering it as belonging to *any* of those states (e.g., *Bagpipe* was either on the study list or is related to a word on the list.) The overdistribution principle posits that subadditivity is caused by items being over-remembered on nondisjunctive probes, relative to disjunctive (either-or) probes. The uncertainty is that there is a complementary explanation of subadditivity which assumes, instead, that items are being under-remembered on disjunctive probes, relative to nondisjunctive ones. As we show in the first section, however, the overdistribution principle predicts another set of phenomena that the complementary explanation does not predict: memory conjunction illusions. Memory conjunction illusions are circumstances in which uncertainties about whether items are over-

remembered on one type of probe versus under-remembered on another type do not arise because memory for multiple incompatible states is measured with a single probe.

In the remainder of the paper, we report some experiments that evaluated the conjunction illusion prediction. The first implemented a design that resembled an earlier subadditivity experiment (Brainerd et al., 2012), with a view toward comparing the levels of overdistribution that are estimated by the traditional subadditivity method versus the alternative conjunction illusion method. The remaining experiments extended the basic phenomenon of memory conjunction illusions to other types of designs. Those experiments included a manipulation (inter-list repetition of items) that was intended to affect an underlying process that might be responsible for memory conjunction illusions. They were also designed to detect an extreme form of such illusions, memory conjunction fallacies.

Subadditivity and Overdistribution

To date, research on overdistribution has relied on two staple paradigms of the false memory literature—namely, false memory for occurrence and false memory for source. In occurrence designs (e.g., Greene & Tussing, 2001; Lampinen, Copeland, & Neuschatz, 2001; Miller, Guerin, & Wolford, 2011; Miller & Wolford, 1999; Tussing & Greene, 1999), subjects encode a set of target materials, followed by an old/new recognition test on which the items are targets (*auto, couch*), related distractors (*car, sofa*; the customary false memory indexes), and unrelated distractors (*baby, river*; the customary bias indexes). In source designs (e.g., Dennis et al., 2008; Hicks, & Starns, 2006; Kurilla & Westerman, 2010), subjects encode target materials in two (or more) distinct contexts—for instance, List 1 versus List 2—and the test items are targets from the first context, targets from the second context, and unrelated distractors. Subjects make old/new decisions about all three types of items, followed by source judgments (e.g., List 1 or List 2?) for any that are classified as old.

To measure overdistribution with either paradigm, these traditional tests are replaced by conjoint recognition tests. There are three such tests, which are factorially manipulated over the three types of items. For occurrence designs, they are T = accept only targets, R = accept only related distractors, and $T \cup R$ = accept targets or related distractors. For source designs, they are L_1 = accept only List 1 targets, L_2 = accept only List 2 targets, and $L_1 \cup L_2$ = accept targets from either list. The index of overdistribution is a subadditive relation among the response probabilities. That relation holds between acceptance rates for disjunctive memory probes ($T \cup R$ and $L_1 \cup L_2$) versus acceptance rates for non-disjunctive ones, and it is quantified with metrics that are used to detect disjunction fallacies in probability judgment (Tversky & Koehler, 1994; Rottenstreich, Brenner, & Sood, 1999; Rottenstreich & Tversky, 1997).

To see how this works, consider the bias-corrected acceptance probabilities for the three types of conjoint recognition tests; denoted $p(T)$, $p(R)$, and $p(T \cup R)$, respectively, or $p(L_1)$, $p(L_2)$, and $p(L_1 \cup L_2)$, respectively. [The type of bias correction (e.g., one-high threshold, two-high threshold, d') is irrelevant for purposes of this illustration.] The former are the probabilities in occurrence experiments that subjects remember any given target or related

distractor as being a target, as being a related distractor, or as being one or the other, whereas the latter are the probabilities in source experiments that subjects remember any given target as having been presented on List 1, or presented on List 2, or presented on one or the other. Now, consider a fourth acceptance probability, $p(T \cap R)$ for occurrence or $p(L_1 \cap L_2)$ for source, which is the probability that an event is remembered as being *both* a target and a related distractor or as having been presented on *both* List 1 and List 2. These are the overdistribution probabilities for the two paradigms. Of course, $p(T \cap R) = 0$ as a matter of logic, and $p(L_1 \cap L_2) = 0$ as a matter of experience (no target appears on both lists). Nevertheless, memory may treat some items as being both targets and related distractors, so that $p(T \cap R) > 0$ empirically, or it may treat some items that were presented in only one context as being presented in both, so that $p(L_1 \cap L_2) > 0$ empirically. Indeed, various findings in the false memory literature (e.g., Multhaup, De Leonardi, & Johnson, 1999; Zaragoza, Mitchell, Payment, & Drivdahl, 2011) and certain theoretical distinctions (e.g., Holliday & Hayes, 2002; Lampinen, Meier, Arnal, & Leding, 2005) point to the conclusion that memory can be in contradictory minds about the episodic status of some items.

Given values of $p(T)$, $p(R)$, and $p(T \cup R)$ or of $p(L_1)$, $p(L_2)$, and $p(L_1 \cup L_2)$ from data, $p(T \cap R)$ and $p(L_1 \cap L_2)$ are easily calculated via subadditivity statistics that Tversky and Koehler (1994) introduced to measure disjunction fallacies in probability judgment. The axioms of classical probability theory constrain probabilities to be additive—so that $p(T \cap R)$ is completely determined by $p(T)$, $p(R)$, and $p(T \cup R)$ because $p(T) + p(R) = p(T \cup R) + p(T \cap R)$, and similarly, $p(L_1 \cap L_2)$ is completely determined by $p(L_1)$, $p(L_2)$, and $p(L_1 \cup L_2)$. Thus, whenever $p(T)$ and $p(R)$ are subadditive with respect to $p(T \cup R)$ —that is, whenever $p(T) + p(R) > p(T \cup R)$ —it must be the case that $p(T \cap R) > 0$, and whenever $p(L_1)$ and $p(L_2)$ are subadditive with respect to $p(L_1 \cup L_2)$, it must be the case that $p(L_1 \cap L_2) > 0$. Given empirical values of $p(T)$, $p(R)$, and $p(T \cup R)$ or of $p(L_1)$, $p(L_2)$, and $p(L_1 \cup L_2)$, then, the estimate of $p(T \cap R)$ is just $p(T) + p(R) - p(T \cup R)$, and the estimate of $p(L_1 \cap L_2)$ is just $p(L_1) + p(L_2) - p(L_1 \cup L_2)$. Equivalently, the ratios $[p(T) + p(R)] \div p(T \cup R)$ and $[p(L_1) + p(L_2)] \div p(L_1 \cup L_2)$, which Tversky and Koehler called unpacking ratios, will be > 1 .

It has been found with such measures that nondisjunctive memory probes are subadditive with respect to disjunctive ones, in both occurrence and source designs. Illustrative data from occurrence experiments appear in Panel A of Figure 1, and illustrative data from source experiments appear in Panel B. In the occurrence data, it can be seen that the level of subadditivity is roughly 20% overall, with the level being greater on immediate than on delayed tests. In the source data, the level of subadditivity is more than one-third overall, and it is greater with three contexts than with two. The overdistribution principle is an explanation of those patterns which, based on classical probability theory, interprets them to mean that some items are remembered as being both targets and distractors (i.e., $p(T \cap R) > 0$), and some items are remembered as having been presented on both List 1 and List 2 (i.e., $p(L_1 \cap L_2) > 0$). As noted earlier, a key psychological implication of this principle is that memory for such items' episodic status is not a direct consequence of prior experience but, rather, is an emergent property of testing.

However, the overdistribution principle's conclusion that $p(T \cap R) > 0$ follows from $p(T) + p(R) - p(T \cup R)$ and $p(L_1 \cap L_2) > 0$ follows from $p(L_1) + p(L_2) - p(L_1 \cup L_2)$ is subject to a measurement uncertainty that leads to a different explanation of subadditivity. We saw that the overdistribution principle's conclusion rests on the rule that probabilities are additive, but perhaps that rule does not hold for memory performance. Certainly, the possibility that the probabilities of other types of cognitive events violate additivity has been discussed in other literatures, with inductive inference (e.g., Dubois, & Prade, 1988), similarity judgment (e.g., Pothos, Busemeyer, & Trueblood, 2013), and probability judgment (e.g., Tversky & Koehler, 1994) being illustrations. Nonadditive models of probability judgment are of special interest because they explicitly postulate *subadditivity*—that, like memory performance, nondisjunctive probability estimates (What is the probability of dying in a car accident? What is the probability of dying in an airplane crash?) are subadditive with respect to equivalent disjunctive estimates (What is the probability of dying in a car accident or in an airplane crash?).

Moreover, nonadditive models of probability judgment supply process explanations of subadditivity that do *not* assume that subjective probabilities of incompatible conjunctions (What is the probability of dying in a car accident and in a plane crash?) are greater than zero. In particular, support theory (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997) provides an influential memory account of subadditivity of probability estimates that features the notion of packed versus unpacked retrieval cues. According to that explanation, disjunctive estimates (e.g., dying in car accident or an airplane crash) are less effective cues for accessing pertinent memory support because they are packed descriptions of the same events that nondisjunctive estimates describe in unpacked form. The underlying assumption, for which there is independent evidence (e.g., Biswas, Keller, & Burman, 2012; Redden & Frederick, 2011), is that disjunctive estimates have to be mentally decomposed before retrieval can occur, whereas nondisjunctive ones do not. Owing to that obstacle, the total amount of memory support that accrues for two (or more) nondisjunctive probability estimates is greater than the amount that accrues for an equivalent disjunctive estimate. It is easy to see that a parallel account could be formulated for memory probes: Because they are packed descriptions, $T \cup R$ and $L_1 \cup L_2$ may simply be poorer retrieval cues for the episodic traces that are accessed by their nondisjunctive counterparts.

It is clear that other scenarios could be generated whose common thread would be that disjunctive memory probes are less effective retrieval cues than nondisjunctive probes. Perhaps the most obvious one is sheer familiarity. In everyday life, disjunctive queries (Did you drink a soda or eat a hamburger at lunch?) are rare in comparison to nondisjunctive ones (Did you drink a soda at lunch? Did you eat a hamburger at lunch?), and thus, we are far more practiced at the latter type of retrieval. Our point is not to argue for any particular process model of why disjunctive probes might be less effective retrieval cues than nondisjunctive probes. Rather, our point is merely that if this general line of reasoning happens to be correct, subadditivity may mean nothing more than that; it may not have larger implications about overdistribution of episodic memories.

Fortunately, there is a way around this dilemma because the overdistribution principle predicts another series of phenomena that do not fall out of this line of reasoning:

conjunction illusions. Suppose that memory is actually in two minds about some items' episodic status, so that in fact $p(T \cap R) > 0$ and $p(L_1 \cap L_2) > 0$. Concerns about the relative effectiveness of disjunctive versus nondisjunctive retrieval cues arise because these probabilities are not directly measured but, instead, are indirectly estimated by applying the additivity axiom of classical probability theory. In other words, the overdistribution interpretation assumes that subjects are experiencing conjunctive memory illusions—are actually remembering items as being both targets and distractors in occurrence designs or has having been presented in multiple contexts in source designs—but such illusions are not directly measured. That is simply because conjunctive memory probes do not figure in the procedure.

Of course, such probes could not be administered in occurrence designs because describing items as simultaneously being targets and related distractors would be oxymorons to subjects. In source designs, however, probes that ask if targets were presented in both contexts are not oxymorons. The incompatibility is a design feature rather than a logical inconsistency. Subjects need not be informed of that design feature, which would make conjunctive probes quite plausible. Because conjunctive probes provide direct measurements of $p(L_1 \cap L_2)$, they evidently answer the question of whether memory distributes items that were experienced in a single context to other contexts as well.

We report three experiments that attempted to measure conjunction illusions in source designs. To maximize the theoretical yield of the results, other key attributes of the designs—such as lists, content manipulations, and subject samples—paralleled those of an earlier series of experiments on subadditive relations between disjunctive and nondisjunctive probes (Brainerd et al., 2012). Except for the use of conjunctive probes, Experiment 1 was the same as one of those earlier experiments, so that conjunctive measurements of $p(L_1 \cap L_2)$ and the corresponding subadditivity estimates could be compared for several treatment conditions. The aim of those comparisons was to determine how well indirect subadditivity estimates calibrate direct conjunctive measurements.

As the results of the Experiment 1 showed that subjects were indeed experiencing conjunction illusions, Experiments 2 and 3 incorporated a manipulation that evaluated a possible process mechanism, which involves retrieved memory support—namely, that illusions are experienced for targets that retrieve the highest levels of subjective memory support. Experiments 2 and 3 continued to show that subjects were experiencing conjunction illusions. An instructive new finding was that in some conditions of these experiments, conjunction illusions became conjunction fallacies. On analogy to conjunction illusions in probability judgment (Tversky & Kahneman, 1983), subjects thought it was more likely that items had been presented in multiple contexts than in any of the individual contexts, which is impossible.

Experiment 1

As mentioned, this experiment was a conjunctive analogue of an earlier subadditivity experiment—specifically, the first experiment reported in Brainerd et al. (2012). In the earlier experiment, subjects studied target words in two contexts, List 1 and List 2, and

responded to a memory test on which three types of items (List 1 targets, List 2 targets, and distractors) were factorially crossed with three types of test probes (L_1 , L_2 , and $L_1 \cup L_2$). The current experiment followed the same design, except that each $L_1 \cup L_2$ probe was replaced by a corresponding $L_1 \cap L_2$ probe. Other features of the earlier experiment were preserved as well. In particular, two word-content variables (high versus low frequency and abstract versus concrete) that were manipulated over target and distractor items in the earlier experiment were manipulated in the same manner in this experiment.

Method

Subjects—In their Experiment 1, Brainerd et al. (2012) used subject samples from a university undergraduate psychology pool. In this experiment, the subjects were 59 introductory psychology students from the same pool who participated to fulfill a course requirement. The present experiment was conducted six months after the earlier one.

Materials and Procedure—In earlier research in which source-monitoring designs were used to measure subadditivity, word frequency and concreteness were manipulated over study and test lists. To maintain methodological comparability with that work, these word content variables were also manipulated in the present experiments, using a word pool that was created by Brainerd et al. (2012). The pool consists of 256 nouns that were sampled from available norms for frequency and concreteness. There are four sub-pools with 64 words apiece: (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 (d) low-frequency/abstract nouns (e.g., *gist*, *rating*). The mean frequencies (per million in printed text) and mean concreteness values (7-point scale) were high frequency/concrete (HF-C) = 73.3/6.6, high frequency/abstract (HF-A) = 78.5/2.2, low frequency/concrete (LF-C) = 1.9/6.5, and low frequency/abstract (LF-A) = 2.1/2.2. The two study lists that were administered to individual subjects were constructed by randomly sampling 24 words from each of the four Frequency $\times$ Concreteness sub-pools. The cue words on the test lists then consisted of the 96 target words, together with 96 distractors. The distractors for each subject's test list were generated by sampling a further 24 unrepresented words from each of the Frequency $\times$ Concreteness sub-pools.

Two lists were presented to each subject. The words on each list appeared in a distinctive visual context that consisted of all of the words being printed in one of several atypical fonts (e.g., Algerian, Broadway, Script) against one of several background colors (e.g., yellow, white, pink). Thus, each presentation context (list) was distinguished by a unique combination of temporal order, font, and background color (e.g., the first list in Broadway font against a pink background, the second in Algerian font against a yellow background). These are features that are commonly used to make source information more memorable in source-monitoring experiments (cf. Arndt, 2012).

At the start of the experiment, the subjects received general instructions, informing them that two different lists of words would be presented, followed by a memory test. During the study phase, two lists of 54 words each (a 3 word opening buffer, followed by 48 targets, followed by a 3 word closing buffer) were presented. The 48 targets consisted of 12 HF-C

nouns, 12 HF-A nouns, 12 LF-C nouns, and 12 LF-A nouns, for a total of 24 targets from each sub-pool over the two lists. The words on each list were presented at a 2-s rate, centered on the computer screen and printed in 72-point bold type, and there was a 15-s pause between lists. During the test phase, subjects responded in a self-paced manner to 192 probes, presented in random order. Each probe consisted of a cue word and an episodic description of the word, which subjects had to accept or reject according to whether they thought that the description was true or false. Half the cue words were targets (the 24 presented words from each of the 4 sub-pools), and half were distractors (24 unpresented words from each of the 4 sub-pools). The three episodic descriptions, which varied factorially over targets and distractors, were: (a) L_1 = presented on List 1? (b) L_2 = presented on List 2? (c) $L_1 \cap L_2$ = presented on List 1 and List 2? Thus, 64 probes of each of the 3 types were administered (32 for targets, with 8 for each Frequency $\times$ Concreteness sub-pool, and 32 for distractors, with 8 for Frequency $\times$ Concreteness sub-pool). The instructions that subjects received prior to the test told them to: accept an L_1 probe if they remembered a word as being presented on List 1, regardless of whether they also remembered it as being presented on List 2; accept an L_2 probe if they remembered a word as being presented on List 2, regardless of whether they also remembered it as being presented on List 1; accept an $L_1 \cap L_2$ probe only if they remembered a word as being presented on both lists. The instructions also stated that half the test words would be targets and half would be distractors.

Results and Discussion

Recall that the design was 3 (probe type: L_1 , L_2 , $L_1 \cap L_2$) $\times$ 2 (list: 1 vs. 2) $\times$ 2 (word type: target vs. distractor) $\times$ 2 (frequency: high vs. low) $\times$ 2 (concreteness: concrete vs. abstract). Summary statistics (mean acceptance probabilities and standard deviations) for the cells of this design appear in Table 1. Summary statistics of greater interest appear in Table 2, where the bias-corrected target acceptance probabilities are reported for the cells of the design. In this table and in the analyses reported below, the two-high-threshold bias correction (e.g., Snodgrass & Corwin, 1988) was used, to preserve comparability with Brainerd et al. (2012). However, the data were also analyzed with three other familiar methods of correcting for response bias (one-high threshold, d' , and A'), and all of the empirical patterns were the same.

The important values in Table 2 are the rows of bias-corrected values for $p(L_1 \cap L_2)$, the overdistribution probabilities for List 1 and List 2. A pair of observed-predicted numbers is reported for each cell of the design. The observed value is the direct measure of the conjunction illusion that was obtained in this experiment, whereas the predicted value is the indirect subadditivity estimate that was reported by Brainerd et al. (2012) for the same design cell.

We asked two questions about the overdistribution results in Table 1. First, were some or all of the observed values of $p(L_1 \cap L_2)$ reliably greater than zero, and second, if so, were they reliably different than the indirect estimates that appear in the predicted cells? With respect to the first question, if subadditivity is not caused by overdistribution, so that subjects are not experiencing conjunction illusions, values of $p(L_1 \cap L_2)$ will not exceed zero. To test

that prediction, we computed a one-sample t test for each of the eight observed values, under the null hypothesis that $p(L_1 \cap L_2) = 0$ in each instance. The results appear in the two rows that are labeled $t_{=0}$. Note that all eight test statistics are highly reliable, as were the mean values of $p(L_1 \cap L_2)$ for List 1 and for List 2. Qualitatively, this is somewhat stronger evidence of overdistribution than the subadditivity estimates reported by Brainerd et al. (2012). In their experiment, the subadditivity estimates for two of the cells in Table 2 (the HF-C and HF-A cells for List 2) did not differ reliably from zero. Another overdistribution result from the earlier experiment that was also present here concerned the effect of temporal order: Subadditivity estimates of $p(L_1 \cap L_2)$ were larger for List 1 than for List 2 and so were direct conjunctive estimates, $t(58) = 3.65$, $SE_M = .02$, $p < .0005$.

It appears, then, that subjects were in fact experiencing the illusion that some words appeared on both lists. However, the earlier line of reasoning about subadditivity—that it is due to disjunctive probes being poorer retrieval cues than nondisjunctive ones—may be still partially correct. More explicitly, even if subjects are experiencing conjunction illusions, the subadditivity effect could be *partially due* to differences in retrieval effectiveness. That hypothesis predicts that although all of the direct $p(L_1 \cap L_2)$ estimates were greater than zero (Table 2), they should be smaller than the indirect subadditivity estimates for the same cells. To test that prediction, we computed a one-sample t test for each of the eight pairs of estimates, under the null hypothesis that the two estimates did not differ reliably. The results appear in the two rows that are labeled $t_{p=0}$. Note that, in general, the paired direct-indirect estimates differed reliably: Seven of the eight individual tests were reliable, and the mean absolute value of the $t(58)$ statistic was 4.44. However, the pattern of differences was not such as to confirm the predicted difference in retrieval effectiveness. That is because differences between the paired estimates were not consistently in the predicted direction: Although the subadditivity estimate was reliably larger than the conjunctive estimate in four of the comparisons, the opposite was true in three of the comparisons. Further, pooling across the eight comparisons, the mean subadditivity estimate of $p(L_1 \cap L_2)$ was virtually the same as the mean conjunctive estimate (.16 versus .13), and those means did not differ reliably. Thus, in addition to providing direct support for the overdistribution hypothesis, the data failed to provide any support for the retrieval effectiveness hypothesis.

Although the foregoing results are of chief interest, we naturally conducted a full 3 (probe type) $\times 2$ (list) $\times 2$ (frequency: high vs. low) $\times 2$ (concreteness: concrete vs. abstract) analysis of variance (ANOVA) of the bias-corrected response probabilities for targets. That ANOVA produced the results that would be expected from an inspection of Table 2. Specifically, there was a Probe Type $\times$ List interaction, $F(2, 116) = 12.34$, $MSE = .05$, $\eta^2 = .11$, and a List $\times$ Frequency interaction, $F(1, 58) = 39.61$, $MSE = .03$, $\eta^2 = .41$. The Probe Type $\times$ List interaction simply demonstrates that source memory was reliable in this experiment: As can be seen in Table 2, L_1 acceptance rates were higher for List 1 targets than for List 2 targets and L_2 acceptance rates were higher for List 2 targets than for List 1 targets, with both differences being reliable by post hoc tests (Tukey). The List $\times$ Frequency simply means that the effect of word frequency depended on list: As can be seen in Table 2, acceptance rates were higher for LF targets than for HF targets on List 1 but not List 2.

Experiment 2

In the first experiment, subjects displayed reliable conjunction illusions when disjunctive memory probes were replaced by conjunctive ones, a pattern than was predicted by the overdistribution principle. Further, although there were reliable differences between most of the paired conjunctive versus subadditive estimates of $p(L_1 \cap L_2)$, their respective grand means were virtually the same and did not differ reliably. Pooling over conditions, the average observed probability was close to the average predicted probability. In the next experiment, we tested a simple process explanation of how conjunction illusions might arise in this type of design. Although no target was presented more than once, they still differed somewhat in their retrieved levels of memory support, for the obvious reason that some were recognized as old and others were not. Such differences in retrieved memory support might be the basis for conjunction illusions, as follows. Suppose that for targets that retrieved enough support to be perceived as old, subjects tended to accept $L_1 \cap L_2$ probes for those whose retrieved levels of support were especially high, relative to other targets. In other words, targets that retrieve enough support to be perceived as old will also differ among themselves with respect to how much support is retrieved, and subjects may decide that those at the high end of the scale were presented on both lists.

If this scenario is correct, it should be possible to reduce or eliminate conjunction illusions by re-anchoring the scale of retrieved memory support in such a way that the high end can no longer be used to accept $L_1 \cap L_2$ probes because those probes only involve items with lower retrieved support. Within a particular list and a particular combination of the content manipulations in Experiment 1, differences in memory support would have been chiefly subjective because all targets were presented once. It is possible to create larger, objective differences in memory support by merely presenting some targets twice and others once. This effectively re-anchors the high end of the scale, which is now occupied by twice-presented targets. Suppose that those targets are never presented as part of conjunctive probes on memory tests. Under that arrangement, the high end of the memory support scale is not accessed by conjunctive probes; all conjunctive probes access objectively lower values of the scale because the targets that figure in those probes were only presented once. Consequently, conjunction illusions should shrink or disappear, if retrieved differences in memory support were the basis for accepting conjunctive probes in Experiment 1.

Method

Subjects—The subjects were 56 undergraduates who participated in the experiment to fulfill a course requirement.

Materials and Procedure—The materials and procedure were the same as Experiment 1, except for the following modifications. The two lists that were administered to individual subjects contained all of the items that previously appeared on these lists (24 exemplars of each Concreteness $\times$ Frequency combination, plus opening and closing buffers) but also some additional items. Explicitly, each list was lengthened slightly to accommodate the addition of eight filler items that were presented in random positions on both lists. On the test lists, none of these repeated items appeared in any of the test probes.

Results and Discussion

As in Experiment 1, the design was 3 (probe type: L_1 , L_2 , $L_1 \cap L_2$) $\times 2$ (list: 1 vs. 2) $\times 2$ (word type: target vs. distractor) $\times 2$ (frequency: high vs. low) $\times 2$ (concreteness: concrete vs. abstract). Table 2 contains the mean acceptance rates (and *SDs*) for all probes, for all list conditions. The summary statistics of primary interest appear in Table 3, which contains the bias-corrected (two-high-threshold) target acceptance probabilities for the cells of the design. As before, the important entries in Table 3 are the rows of bias-corrected values for $p(L_1 \cap L_2)$, the conjunction illusion probabilities for List 1 and List 2. Because this experiment included a feature (repeated targets) that was intended to reduce conjunction illusions by re-anchoring the high end of the scale of perceived memory support, the main questions are whether such illusions were still present, and if they were, whether they were noticeably smaller than in Experiment 1.

The bias-corrected values for the probability accepting each of the 8 conjunctive probes (2 lists $\times 2$ levels of frequency $\times 2$ levels of concreteness) appear in Table 3. To answer the first question, if the subjects were again experiencing conjunction illusions, some of those values will exceed zero. That prediction was tested by computing a one-sample *t* test for each of the eight values, under the null hypothesis that $p(L_1 \cap L_2) = 0$. The results of those tests appear in the two rows of Table 3 that are labeled $t_{=0}$. As in Experiment 1, all eight test statistics were reliable, with mean List 1 and List 2 $t_{=0}(55) = 3.50$ and 7.62 (both $ps < .0001$), respectively. With respect to the second question, there was no reduction in the strength of conjunction illusions because the mean value of $p(L_1 \cap L_2)$ for List 1 was virtually the same as in Experiment 1 (.15 versus .16), and the mean for List 2 was noticeably larger than before (.09 versus .32). Indeed, when a one-sample *t* test was computed with .09 as the predicted mean for List 2, the test statistic was highly reliable, $t(55) = 9.55$, $SE = .03$, $p < .0001$. (This difference was also reliable by a between-groups *t* test of the mean value of $p(L_1 \cap L_2)$ for the List 1 data of the two experiments.) The straightforward conclusion is that re-anchoring the high end of the scale of perceived memory support in such a way that it cannot be used as a basis for conjunction illusions neither eliminated nor reduced the tendency to experience conjunction illusions.

As in Experiment 1, there was no evidence that conjunction illusions differed for concrete versus abstract words: The mean values of $p(L_1 \cap L_2)$ were .21 and .22 for concrete and abstract items, respectively. With respect to frequency, however, the mean value of $p(L_1 \cap L_2)$ was higher for low-frequency than for high-frequency items (.30 vs. .21). (See the ANOVA, below, for significance tests.) The major new finding was that conjunction illusions rose to the level of conjunction fallacies in some conditions. The distinction between conjunction illusions and fallacies in episodic memory is that the former are experiential inconsistencies, whereas the latter are logical contradictions—analogueous to conjunction fallacies in probability judgment (e.g., Tversky & Kahneman, 1983). As we saw, conjunction illusions violate experience because targets were presented only one list, but there is no logical contraction because it is possible for them to appear on more than one list. In fact, there were some filler items that appeared on both lists. However, again on analogy to conjunction fallacies in probability judgment, certain relations between the acceptance probabilities for conjunctive versus nonconjunctive probes are objectively

impossible. Specifically, even if some of the tested targets had appeared on both lists, it cannot be more likely that a target appeared on both lists than on either individual list: The set of targets that appear on both lists is a proper subset of the targets that appear on the individual lists. Inspection of Table 3 reveals the presence of these logically contradictory relations among the acceptance probabilities.

Specifically, if each value of $p(L_1 \cap L_2)$ is compared to the values of $p(L_1)$ and $p(L_2)$ for that same condition, it can be seen that for List 2, the value of $p(L_1 \cap L_2)$ is larger than the value of both $p(L_1)$ and $p(L_2)$ for three of the four Frequency $\times$ Concreteness combinations. We computed a significance test for this conjunction fallacy using the mean values of $p(L_1 \cap L_2)$, $p(L_1)$, and $p(L_2)$ for List 1 and List 2. For List 1, the means were .15, .30, and .24, respectively, and a univariate F test with follow-up post hoc tests revealed that $p(L_1 \cap L_2)$ was reliably smaller than either nonconjunctive acceptance probability. That, of course, is consistent with the set-subset relation between targets presented on individual lists versus both lists. For List 2, on the other hand, the means were .36, .28, and .27, respectively. Now, a univariate F test with follow-up post hoc tests revealed that $p(L_1 \cap L_2)$ was reliably larger than either nonconjunctive acceptance probability. Thus, for targets presented on List 2, subjects were more likely to remember them as being presented on both lists and on either individual list, which is impossible.¹

Although the foregoing results are of chief interest, we again computed a full 3 (probe type) $\times$ 2 (list) $\times$ 2 (frequency: high vs. low) $\times$ 2 (concreteness: concrete vs. abstract) ANOVA of the bias-corrected response probabilities for targets. As is obvious from a visual inspection of Table 3, there were main effects for list, $F(1, 55) = 14.16$, $MSE = .11$, $\eta^2 = .21$, $p < .0001$, and for frequency, $F(1, 55) = 29.76$, $MSE = .10$, $\eta^2 = .34$, $p < .0001$. Mean acceptance probabilities for all three probe types were higher for List 2 targets than for List 1 targets and higher for low-frequency targets than for high-frequency targets. The most important result was a List $\times$ Probe Type interaction, $F(2, 110) = 23.96$, $MSE = .07$, $\eta^2 = .40$, $p < .0001$, which occurred for two reasons. First, as in Experiment 1, it was due to the fact that source memory was reliable. It can be seen in Table 3 that L_1 acceptance rates were higher for List 1 targets with L_1 probes than with L_2 probes (post hoc tests showed that the difference was reliable), though acceptance rates for List 2 targets did not differ reliably for the two types of probes. Second and more important, the List $\times$ Probe Type interaction was caused by the aforementioned fact that conjunction illusions became conjunction fallacies on List 2 (i.e., acceptance rates for $L_1 \cap L_2$ probes were much higher for List 2 targets).

Another result of some interest is a Frequency $\times$ Concreteness interaction, $F(1, 55) = 20.80$, $MSE = .08$, $\eta^2 = .28$, $p < .0001$. Post hoc comparisons revealed what is visually apparent in Table 3: (a) Neither the size nor the direction of word frequency effects was influenced by concreteness, but (b) acceptance rates were higher for concrete than abstract words when words were high-frequency. In short, there was some tendency for acceptance rates to be higher for concrete words, but it was modulated by frequency.

¹As a further illustration of the illogical consequences of $p(L_1 \cap L_2)$ being larger than $p(L_1)$ or $p(L_2)$, consider the fact that for some combinations of these probabilities, impossible values of $p(L_1 \cup L_2)$ will result from the $p(L_1 \cup L_2) = p(L_1) + p(L_2) - p(L_1 \cap L_2)$ rule. For instance, $p(L_1 \cup L_2)$ will be *smaller* than either $p(L_1)$ or $p(L_2)$ if $p(L_1) = .2$, $p(L_2) = .2$, and $p(L_1 \cap L_2) = .3$, and $p(L_1 \cup L_2)$ will be *negative* if $p(L_1) = .2$, $p(L_2) = .2$, and $p(L_1 \cap L_2) = .5$.

Experiment 3

In Experiment 2, conjunction illusions remained when a repetition manipulation was introduced that re-anchored the high end of the scale of the retrieved memory support by creating objective differences in such support. The high end of the scale was then made inaccessible on memory tests because repeated targets did not appear in any of the probes. Nevertheless, conjunction illusions neither disappeared nor declined, and in fact, they increased to the level of conjunction fallacies in some cells of the design. In Experiment 3, we used a stronger version of this manipulation, in order to determine whether conjunction illusions would be reduced or eliminated when the high end of the scale of retrieved memory support was re-anchored at an even higher value by strengthening the repetition manipulation.

That was accomplished by dividing the two lists that were presented in the first two experiments into three lists of equal length; that is, the number of presentation contexts was increased from two to three but the total number of targets in each Frequency $\times$ Concreteness cell (24) was held constant. This allowed us to present each of the filler items three times (because there were now three lists), while holding the encoding and testing loads for target items and distractors constant. Therefore, any resulting differences in the conjunction illusion could not be due to differences in the memory loads that were imposed by the three-list versus two-list designs but, instead, would be due to increased repetition of filler targets.

Subjects

The subjects were 52 undergraduates who participated in the experiment to fulfill a course requirement.

Materials and Procedure

The materials and procedure were the same as Experiment 2, except for the following two modifications. The first change involved the study lists. Three lists were presented rather than two by pooling the list structures that were previously used and dividing them into three lists with equal numbers of targets from each of the Frequency $\times$ Concreteness cells. Hence, whereas there were 12 targets from each of the four cells on each of the two previously administered lists, there were now 8 targets per cell on each list. Because there were three lists, subjects were exposed to three rather than two unique combinations of background color, letter font, and presentation order. The second change was that the number of times that the filler items were presented was increased from two to three, with each filler item being presented once per list.

Results and Discussion

The design was 4 (probe type: $L_1, L_2, L_3, L_1 \cap L_2 \cap L_3$) $\times$ 2 (list: 1 vs. 2) $\times$ 2 (word type: target vs. distractor) $\times$ 2 (frequency: high vs. low) $\times$ 2 (concreteness: concrete vs. abstract). The mean acceptance rates (and *SDs*) for all probes for all list conditions appear at the bottom of Table 1. The summary statistics of primary interest appear at the bottom of Table 3, which contains the bias-corrected target acceptance probabilities for the cells of the

design. The important entries in Table 3 are the rows that contain the bias-corrected values for $p(L_1 \cap L_2 \cap L_3)$, the conjunction illusion probabilities for Lists 1, 2, and 3. The main question is whether the conjunction illusion decreased or disappeared in this experiment. It might, owing to more extreme re-anchoring of the scale of retrieved memory support: The objective difference between the high end of that scale and the levels of support that could be retrieved by the targets in memory probes (all of which had been present once) was greater than in Experiment 1.

Inspection of the 12 bias-corrected values for the probability of accepting the conjunctive probes (3 lists $\times$ 2 levels of frequency $\times$ 2 levels of concreteness) reveals that subjects were still experiencing conjunction illusions because all 12 values were well above zero. Further, the average value of $p(L_1 \cap L_2 \cap L_3)$ was virtually the same as the average value of $p(L_1 \cap L_2)$ in Experiment 2 ($M_s = .25$ and $.26$, respectively). The hypothesis that subjects were still experiencing conjunction illusions was tested for statistical significance by computing a one-sample t test for each of the 12 values, under the null hypothesis that $p(L_1 \cap L_2 \cap L_3) = 0$. The results of those tests appear in the three rows of Table 3 that are labeled $t_{=0}$. As in prior experiments, all test statistics were reliable, with mean List 1, List 2, and List 3 $t_{=0}$ ($df = 51$) = 4.76, 3.48, and 6.13 (all $p_s < .008$), respectively. Because the average value of $p(L_1 \cap L_2 \cap L_3)$ was virtually the same as the average value of $p(L_1 \cap L_2)$ in Experiment 2, we did not compute a significance test for that small difference. As in Experiment 2, the straightforward conclusion about retrieved memory support is that re-anchoring the high end of the scale and making it inaccessible by memory probes neither eliminated nor reduced the tendency to experience conjunction illusions.

In the first two experiments, there was no evidence that conjunction illusions differed for concrete versus abstract words, but in both experiments, conjunction illusions were more marked for low- than for high-frequency targets. In this experiment, conjunction illusions did not vary reliably as a function of either concreteness or frequency, although the null finding for frequency is due to an interaction with concreteness. (See the ANOVA below for significance tests.) However, two other results that emerged in Experiment 2 were also present in this experiment. First, recall that conjunction illusions were more marked for the last presented list in Experiment 2. The same was true here. The mean bias corrected acceptance probabilities for Lists 1, 2, and 3 were .25, .19, and .32, respectively. Second, recall that conjunction illusions rose to the level of conjunction fallacies in some of the design cells of Experiment 2, and in particular, the mean value of $p(L_1 \cap L_2)$ was larger than the mean values of $p(L_1)$ and $p(L_2)$ for the last presented list. That pattern was present in this experiment, too, as the mean value of $p(L_1 \cap L_2 \cap L_3)$ in Table 3 for List 3 was .32, which was notably larger than the mean values of $p(L_1)$, $p(L_2)$, and $p(L_3)$, which violates the set-subset relation between the latter versus the former. A univariate F test showed that the values of $p(L_1 \cap L_2 \cap L_3)$, $p(L_1)$, $p(L_2)$, and $p(L_3)$ differed reliably, $F(3, 153) = 21.77$, $MSE = .03$, $\eta^2 = .21$, $p < .0001$, and follow-up post hoc tests revealed that $p(L_1 \cap L_2 \cap L_3)$ was reliably larger than each of the other values. In addition, there was now some limited evidence of memory conjunction fallacies for earlier lists. Specifically, it can be seen in Table 3 that the mean value of $p(L_1 \cap L_2 \cap L_3)$ was noticeably larger than the mean value of $p(L_2)$ for List 1 targets and List 2 targets. For both lists, univariate F tests with follow-up

post hoc tests showed that the acceptance probabilities for the four types of probes differed reliably and that $p(L_1 \cap L_2 \cap L_3) > p(L_2)$.

In Experiment 2, it will be remembered that for List 1, although there were robust conjunction illusions in all four Frequency $\times$ Concreteness combinations, the mean values $p(L_1)$ and $p(L_2)$ were reliably larger than the mean value of $p(L_1 \cap L_2)$, which squares with the set-subset relation between nonconjunctive and conjunctive probabilities. In this experiment, however, nonconjunctive probabilities were generally not larger than conjunctive probabilities. Indeed, there was only instance in which the mean value of a nonconjunctive probability was reliably larger than the mean value of the corresponding conjunctive probability—namely, $p(L_3)$ versus $p(L_1 \cap L_2 \cap L_3)$ for List 2.

For the overall significance testing machinery, we computed a 4 (probe type) $\times$ 3 (list) $\times$ 2 (frequency: high vs. low) $\times$ 2 (concreteness: concrete vs. abstract) ANOVA of the bias-corrected response probabilities for targets. Although there were main effects for both frequency and concreteness, the key result for the two word content manipulations was a Probe Type $\times$ Frequency $\times$ Concreteness interaction, $F(3, 153) = 8.65$, $MSE = .17$, $\eta^2 = .15$, $p < .0001$. Post hoc comparisons revealed that the effects of frequency and concreteness were different for conjunctive versus nonconjunctive probes. For nonconjunctive probes, these tests showed that (a) acceptance probabilities were higher for low- than for high-frequency words (.26 vs. .16), regardless of concreteness, and (b) acceptance probabilities were not affected by concreteness. For conjunctive probes, on the other hand, these tests showed that (c) acceptance probabilities were higher for low-frequency words when they were concrete but not when they were abstract, and (d) acceptance probabilities were higher for abstract words when they were low-frequency but not when they were high-frequency. The tendency for conjunction illusions to be more pronounced for low-frequency targets was previously observed in Experiment 2, but this is the first evidence that they may also be influenced by concreteness.

Further results of note are the significance tests for the aforementioned effects of list on conjunction illusions. There was a main effect for probe type, $F(3, 153) = 5.91$, $MSE = .21$, $\eta^2 = .11$, $p < .001$ because, overall, the mean value of $p(L_1 \cap L_2 \cap L_3)$ was larger than the means of the three nonconjunctive probes; that is, the net picture was one of conjunction illusions rising to the level of conjunction fallacies. As mentioned, however, conjunction illusions became statistically reliable conjunction fallacies for all nonconjunctive probes on the last presented list but not on the lists that preceded it. Thus, either the values of $p(L_1 \cap L_2 \cap L_3)$ increased on the last list, or the values of the nonconjunctive probabilities decreased, or both. Inspection reveals that it was the first alternative, which was also the pattern in Experiment 2.

Finally, as in the first two experiments, there was a reliable Probe Type $\times$ List interaction, $F(6, 305) = 2.95$, $MSE = .11$, $\eta^2 = .05$, $p < .008$, and as in Experiment 2, it occurred for two reasons. First, post hoc analysis revealed that source memory was reliable for two of the three lists (List 1 and List 3). Second and more important, post hoc analysis also revealed that $p(L_1 \cap L_2 \cap L_3)$ was reliably larger for List 3 targets than for List 1 or List 2 targets, but it did not differ for List 1 versus List 2 targets. As in Experiment 2, then, conjunction

illusions were stronger for the most recently presented items and became conjunction fallacies.

General Discussion

The overdistribution principle stipulates that as a matter of course, episodic memory assigns events to too many states, including states that are logically or experientially incompatible with one another. We investigated such illusions in order to resolve a fundamental uncertainty about this hypothesis. The uncertainty is that it is based chiefly on memory disjunction fallacies, which are subadditive relations between acceptance probabilities for disjunctive versus nondisjunctive memory probes: When probes are administered for two or more mutually exclusive episodic states (Was *couch* on List 1? Was *couch* on List 2?), the sum of their respective acceptance probabilities is subadditive with respect to the acceptance probability for an equivalent disjunctive probe (Was *couch* on List 1 or List 2?). By the axioms of classical probability theory, this means that the probability that an event is remembered as belonging to two (or more) mutually exclusive states is greater than zero.

However, episodic memory may not obey the axioms of classical probability theory, and if not, something other than overdistribution may be the cause of subadditivity. In particular, it could be due to differences in the relative effectiveness of disjunctive versus nondisjunctive probes as retrieval cues. Here, theoretical and empirical arguments can be marshaled for the superiority of nondisjunctive probes over disjunctive ones. Although this retrieval effectiveness hypothesis can explain subadditivity, the overdistribution principle also predicts conjunction illusions; that when the probability of remembering an event as belonging to mutually exclusive episodic states is directly estimated by administering conjunctive memory probes, estimates will exceed zero. Therefore, we focused squarely on that prediction in our experiments, and it was confirmed in all conditions. Across experiments, there were 28 conditions in which false acceptance probabilities for conjunctive probes were measured. The bias-corrected value was reliably greater than zero in each of them, and over conditions, the mean value was .22. In the initial experiment, the average false acceptance probability over conditions did not differ reliably from the value that was predicted on the basis of an earlier subadditivity experiment, so that the rule $p(L_1) + p(L_2) = p(L_1 \cup L_2) - p(L_1 \cap L_2)$ was satisfied empirically. In the other experiments, conjunction illusions became conjunction fallacies in some conditions because false acceptance probabilities were higher for conjunctive than for nonconjunctive probes. In line with the overdistribution principle, then, the subjects in our experiments experienced conjunction illusions.

As mentioned earlier, a key psychological implication of the overdistribution hypothesis is that for some items, episodic memory is an emergent property of testing because prior to the administration of a probe, events that occurred in a single episodic state have the potential to be remembered as occurring in other, incompatible states as well. Obviously, the circumstance of an item being distributed over a series of mutually incompatible states requires a theoretical explanation. As a stimulus for future hypothesis-driven experimentation, we briefly discuss a pair of straightforward explanations of how memory for an event's episodic status could be an emergent property of testing: (a) familiarity

without recollection and (b) dual episodic traces. In connection with each, we briefly note the types of experimental manipulations that could be used to test the explanation.

Familiarity without Recollection

The first explanation was discussed in the initial studies of subadditivity (Brainerd & Reyna, 2008). It relies on the conventional distinction between recollection and familiarity (e.g., Mandler, 1980), according to which some items are recognized as old because they provoke vivid awareness of their prior presentation while others are recognized as old because they seem very familiar even though their presentation cannot be specifically remembered. According to the first explanation, conjunction illusions and subadditive relations between disjunctive and nondisjunctive acceptance probabilities are in-the-moment attributional distortions, which occur on memory tests for items that provoke familiarity without recollection.

When an item elicits recollection of its prior presentation, subjects know precisely which list it was presented on and, conversely, which list it was not presented on. Thus, attributions for that item on memory tests are internally consistent because subjects will accept correct nonconjunctive probes, and they will simultaneously reject incorrect conjunctive and nonconjunctive probes. Rejections of the latter are accomplished via a process that has been extensively studied in the false memory literature and is variously called recollection rejection (e.g., Brainerd, Reyna, Wright, & Mojardin, 2003), recall-to-reject (e.g., Lampinen, Odegard, & Neuschatz, 2004), or diagnostic monitoring (Gallo, 2004). In contrast, when an item provokes familiarity without recollection, subjects are confident that it belongs to one of the episodic states of an experiment, but they do not know which one. Owing to this combination of confidence and uncertainty, their responses to memory probes are *description dependent*: As long as the episodic state that is described in a probe is a possible one for that experiment, subjects tend to accept it. In the recognition literature, this principle is implemented in models such as process dissociation (Jacoby, 1991) and, in the false memory literature, in models such as conjoint recognition (Brainerd, Reyna, & Mojardin, 1999). Both types of models include parameters that quantify the phenomenon of familiarity-based acceptance of multiple inconsistent episodic states.

Naturally, this would produce conjunction illusions in source designs because $L_1 \cap L_2$ is a possible state of such an experiment, unless it is explicitly ruled out in the instructions that subjects receive. The same mechanism will produce subadditivity in occurrence designs, as well as source designs, as follows. Suppose that *couch* is a target in an occurrence design in which the false memory items are synonyms of targets (e.g., *sofa*). Assume, as in the conjoint recognition model, that recognition of *couch* is recollection-based with some probability R and familiarity-based with some probability $(1-R)F$. Ignoring the influence of response bias, probes that ask if *couch* is a target will be accepted with probability $R + (1-R)F$, probes that ask if it is a synonym of a target will be accepted with probability $(1-R)F$, and probes that ask if it is one or the other will be accepted with probability $R + (1-R)F$. Thus, the sum of two nondisjunctive probabilities exceeds the disjunctive probability by $(1-R)F$, the probability that *couch* is classified as a target on target probes and as a related distractor on related distractor probes (Brainerd & Reyna, 2008). Likewise, in a two-list

source design, assume that *couch* is presented on List 1 and that recognition of *couch* is recollection-based or familiarity-based with the same probabilities. Here, according to the process dissociation model, probes that ask if it was presented on List 1 will be accepted with probability $R + (1-R)F$, probes that ask if it was presented on List 2 will be accepted with probability $(1-R)F$, and probes that ask if it was presented on one or the other list will be accepted with probability $R + (1-R)F$. Again, the sum of the nondisjunctive probabilities will exceed the disjunctive probability by $(1-R)F$.

Although conjunction illusions and subadditivity could thus result from familiarity-driven responding, this explanation has two potential limitations, one conceptual and the other empirical. The conceptual one concerns the plausibility of assuming familiarity-based acceptance of conjunctive probes, considering the differences in perceived memory support that are provoked by familiarity versus recollection. Our commonsense expectation is that memory support will be greater for targets that have been presented on multiple lists than for targets that have been presented once on a single list, so that subjects ought to be more inclined to accept conjunctive probes for items with greater perceived memory support. A well-known finding from the false memory literature is that perceived memory support is *lower* for probes that provoke familiarity than for probes that provoke recollection (e.g., Brainerd et al., 2003; Lampinen, Neuschatz, & Payne, 1998). Consequently, to predict conjunction illusions, it must be assumed that subjects are inclined to accept conjunctive probes for items whose perceived level of memory support is low, relative to other items. The second limitation of the familiarity explanation is that it appears to rule out memory conjunction fallacies—that conjunctive probes could be accepted at higher rates than correct nonconjunctive probes. In the above example for source designs, for instance, those probabilities would be $(1-R)F$ and $R + (1-R)F$, respectively. However, in Experiment 2 and 3, acceptance probabilities were sometimes higher for conjunctive probes than for correct nonconjunctive probes.

To conduct systematic tests of the familiarity explanation, an obvious approach would be to rely on the vast archive of experiments on the recollection-familiarity distinction. Using methodologies such as remember/know judgment and process dissociation, there are several variables that have been found to increase the incidence of recollection-based recognition, relative to familiarity-based recognition. Some well-replicated examples are replacing word lists with picture lists (e.g., Wagner, Gabrieli, & Verfaellie, 1997), decreasing list length (e.g., Rajaram, 1993), testing younger rather than older adults (e.g., Duarte, Graham, & Henson, 2010), generating rather than reading words on study lists (e.g., Hicks et al., 2002), studying lists under longer versus shorter exposure times (e.g., Hirshman, Fisher, Henthorn, Arndt, & Passannante, 2002), studying emotionally-valenced lists rather than neutral ones (e.g., Ochsner, 2000), and studying lists under full rather than divided attention (e.g., Mangels, Picton, & Craik, 2001). If conjunction illusions and subadditivity are due to familiarity-based responding, the prediction with respect to each of these variables is that it will decrease both forms of distortion. More generally, the first explanation expects that in the main, variables that are already known to affect the incidence of familiarity-based responding will have parallel effects on subadditivity and conjunction illusions.

Dual Episodic Traces

The fundamental difference between the second explanation and the first is that the second assumes that distortions are consequences of stable, enduring properties of episodic traces rather than in-the-moment attributions on memory tests. More particularly, the second explanation assumes that details are routinely *stored in episodic traces* that are consistent with items belonging to multiple states, when they do not belong to multiple states. Therefore, retrieval of such traces could cause subjects to accept conjunctive probes and could generate subadditivity by supporting acceptance of nondisjunctive probes for mutually incompatible states. According to this explanation, representation is promiscuous; some episodic traces contain details that are consistent with too many episodic states. Such representational distortion is a shared principle of contemporary dual-trace models, such as fuzzy-trace theory (FTT; Brainerd et al., 2003), retrieving effectively from memory (REM; Shiffrin, 2003), and storing and retrieving knowledge and events (SARKAE; Nelson & Shiffrin, 2013). In these models, target presentations lead to the storage of two distinct types of episodic traces—called verbatim and gist in FTT, episodic and lexical-semantic in REM, and item and knowledge in SARKAE. The first type of trace captures the specifics of *individual* target presentations, whereas the second captures a variety of additional information that goes beyond individual presentations, such as their semantic content, contextual details from past encounters with them, inferences about them, and so forth.

It is retrieval of gist (lexical-semantic/knowledge) traces that directly foments conjunction illusions and subadditivity in such models. The reason is that traces of this second type usually contain details that are consistent with an item belonging to multiple episodic states—multiple states of the current experiment but also of past encounters (see Nelson & Shiffrin, 2013). In source designs, then, a List 1 target can produce acceptance of both L1 probes and L2 probes and so can a List 2 target, resulting in subadditivity, because subjects retrieve traces containing details that are consistent with both states. For the same reason, retrieval of such traces supports acceptance of conjunctive probes for List 1 and List 2 targets. In occurrence designs, targets can produce acceptance of both target and related distractor probes and so can related distractors because, again, subjects retrieve traces that contain details that are consistent with both states.

A practical advantage of the dual-trace explanation is that it is not subject to either of the limitations of the familiarity explanation. The first limitation was that the familiarity account of conjunction illusions implies that the items for which conjunctive probes are accepted are ones whose perceived levels of memory support are low, relative to other items, and the second was that such an account does not seem to allow conjunction illusions to become conjunction fallacies. With respect to first limitation, the dual-trace explanation does not assume that perceived memory support is necessarily low when gist (lexical-semantic/knowledge) traces are retrieved. On the contrary, experimentation on the phenomenon of phantom recollection suggests that it can be quite high (for reviews, see Arndt, 2012; Brainerd & Reyna, 2005; Lampinen et al., 1998). In this line of research, perceived memory support is measured (e.g., via confidence ratings or remember/know judgments) for related distractors, which can access their gist traces but cannot access their verbatim traces (because they were never presented). For some distractors, subjects report the same strong,

vivid memory support that they report for targets, and in certain designs, such support is reported as often for distractors as for targets (e.g., Lampinen et al., 2005). Turning to the other limitation, the second explanation does not rule out the possibility of false acceptance probabilities for conjunctive probes being higher than correct acceptance probabilities for nonconjunctive probes. Because dual traces are involved, the respective acceptance rates for the two types of probes will depend on: (a) the relative tendencies of conjunctive and nonconjunctive probes to access verbatim versus gist traces and (b) the relative strengths of the two types of traces. Here, common assumptions of dual-trace models are that probes for actual presentation events (e.g., nonconjunctive probes) are better retrieval cues for verbatim traces, that probes for related-but-unpresented events (e.g., conjunctive probes) are better retrieval cues for gist traces, and that as time passes, verbatim traces weaken more rapidly than gist traces (e.g., Brainerd et al., 2003; Nelson & Shiffrin, 2013). Under such assumptions, there are multiple scenarios in which acceptance rates could be higher for conjunctive probes.

A promising strategy for testing the dual-trace explanation is to rely on the extensive literature on semantic false memory. There, using methodologies such as conjoint recognition and phantom ROC (e.g., Lampinen et al., 2004), manipulations have been identified that increase subjects' tendency to base performance on the gist of target events because they increase false memory for semantically-related distractors (e.g., *sofa*), relative to true memory for targets (e.g., *couch*). Examples include administering pre-study cues that focus attention on the semantic content of upcoming lists, increasing intra-list semantic relatedness of targets, performing deep rather than shallow encoding tasks, interpolating post-study buffer activities that require subjects to retrieve semantic content, and administering delayed rather than immediate memory tests (for reviews, see Brainerd & Reyna, 2005; Gallo, 2006). If disjunction fallacies and conjunction illusions are indeed due to processing gist traces, these manipulations ought to increase both forms of distortion. More generally, the literature provides examples of manipulations that increase false memory via two basic routes: by enhancing subjects' ability to store and retrieve the semantic content of events or by interfering with their ability to store and retrieve the surface form of events. According to the dual-trace explanation, manipulations of the first type should be effective at increasing subadditivity and conjunction illusions.

Concluding Comment

Although this paper has been focused on illusions of episodic memory, it should perhaps be noted that there seem to be some remarkable parallels between well-known biases in the sphere of human judgment and memory for occurrence and for source. In the judgment domain, an often-stated principle (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997) is that the probabilities of disjunctive events are underestimated whereas the probabilities of conjunctive events are overestimated. That is because (a) disjunctions of mutually exclusive events are rated as less probable than the sum of the probability ratings of their component events, and (b) conjunctions of events are rated as more probable than some of their components (e.g., "Linda is a feminist bank teller" is rated as more probable than "Linda is a bank teller"). When the present experiments are combined with earlier studies of memory disjunction fallacies, a parallel rule, in which disjunctive events are under-

remembered while conjunctive events are over-remembered, seems to hold for episodic memory.

Acknowledgments

Preparation of this article was supported by National Institutes of Health Grant 1RC1AG036915 to the first and fourth authors.

References

- Arndt J. False recollection: Empirical findings and their theoretical implications. *Psychology of Learning and Motivation*. 2012; 56:81–124.
- Biswas D, Keller LR, Burman B. Making probability judgments of future product failures: The role of mental unpacking. *Journal of Consumer Psychology*. 2012; 22:237–248.
- Brainerd, CJ.; Reyna, VF. *The science of false memory*. New York: Oxford University Press; 2005.
- Brainerd CJ, Reyna VF. Episodic over-distribution: A signature effect of recollection without familiarity. *Journal of Memory and Language*. 2008; 58:765–786.
- Brainerd CJ, Reyna VF, Aydin C. Disjunction fallacies in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2010; 36:711–735.
- Brainerd CJ, Reyna VF, Holliday RE, Nakamura K. Overdistribution in source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2012; 38:413–439.
- Brainerd CJ, Reyna VF, Mojardin AH. Conjoint recognition. *Psychological Review*. 1999; 106:160–179. [PubMed: 10197365]
- Brainerd CJ, Reyna VF, Wright R, Mojardin AH. Recollection rejection: False-memory editing in children and adults. *Psychological Review*. 2003; 110:762–784. [PubMed: 14599242]
- Brainerd C, Wang Z, Reyna V. Superposition of episodic memories: Overdistribution and quantum models. *Topics in Cognitive Science*. 2013; 5:773–799. [PubMed: 24027175]
- Bussemeyer JR, Wang Z, Townsend JT. Quantum dynamics of human decision-making. *Journal of Mathematical Psychology*. 2006; 53:220–241.
- Dennis NA, Hayes SM, Prince SE, Madden DJ, Huettel SA, Cabeza R. Effects of aging on the neural correlates of successful item and source memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2008; 34:791–808.
- Duarte A, Graham KS, Henson RN. Age-related changes in neural activity associated with familiarity, recollection and false recognition. *Neurobiology of Aging*. 2010; 31:1814–1830. [PubMed: 19004526]
- Dubois D, Prade H. Modelling uncertainty and inductive inference: A survey of recent non-additive probability systems. *Acta Psychologica*. 1988; 68:53–78.
- Gallo DA. Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2004; 30:120–128.
- Gallo, DA. *Associative illusions of memory*. New York: Psychology Press; 2006.
- Greene RL, Tussing AA. Similarity and associative recognition. *Journal of Memory and Language*. 2001; 45:573–584.
- Hicks JL, Marsh RL, Ritschel L. The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2002; 28:503–508.
- Hicks JL, Starns JJ. Remembering source evidence from associatively related items: Explanations from a global matching model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2006; 32:1164–1173.
- Hirshman E, Fisher J, Henthorn T, Arndt J, Passannante A. Midazolam amnesia and dual-process models of the word-frequency mirror effect. *Journal of Memory and Language*. 2002; 47:499–516.
- Holliday RE, Hayes BK. Automatic and intentional processes in children's recognition memory: The reversed misinformation effect. *Applied Cognitive Psychology*. 2002; 16:1–16.
- Jacoby LL. A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*. 1991; 30:513–541.

- Kurilla BP, Westerman DL. Source memory for unidentified stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2010; 36:398–410.
- Lampinen JM, Copeland SM, Neuschatz JS. Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2001; 27:1211–1222.
- Lampinen JM, Meier CR, Arnal JD, Leding JK. Compelling untruths: Content borrowing and vivid false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31:954–963.
- Lampinen JM, Neuschatz J, Payne DG. Memory illusions and consciousness: Examining the phenomenology of true and false memories. *Current Psychology: Development, Learning, Personality, Social*. 1998; 16:181–224.
- Lampinen JM, Odegard TN, Neuschatz JS. Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2004; 30:332–342.
- Mandler G. Recognizing: The judgment of previous occurrence. *Psychological Review*. 1980; 87:252–271.
- Mangels JA, Picton TW, Craik FIM. Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research*. 2001; 11:77–95. [PubMed: 11240113]
- Miller MB, Guerin SA, Wolford GL. The strategic nature of false recognition in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2011; 37:1228–1235.
- Miller MB, Wolford GL. Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*. 1999; 106:398–405.
- Multhaup KS, De Leonadis DM, Johnson MK. Source memory and eyewitness suggestibility in older adults. *The Journal of General Psychology*. 1999; 129:74–84. [PubMed: 10216970]
- Nelson AB, Shiffrin RM. The co-evolution of knowledge and event memory. *Psychological Review*. 2013; 120:356–394. [PubMed: 23458086]
- Ochsner KN. Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General*. 2000; 129:242–261. [PubMed: 10868336]
- Pothos EM, Busmeyer JR, Trueblood J. A quantum geometric model of similarity. *Psychological Review*. 2013; 120:679–696. [PubMed: 23915087]
- Rajaram S. Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*. 1993; 21:89–102. [PubMed: 8433652]
- Redden JP, Frederick S. Greater detail can reduce perceived likelihood. *Journal of Experimental Psychology: General*. 2011; 140:159–167. [PubMed: 21261417]
- Rottenstreich Y, Brenner L, Sood S. Similarity between hypotheses and evidence. *Cognitive Psychology*. 1999; 38:110–128. [PubMed: 10090800]
- Rottenstreich Y, Tversky A. Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review*. 1997; 104:406–415. [PubMed: 9127585]
- Shiffrin RM. Modeling memory and perception. *Cognitive Science*. 2003; 27:341–378.
- Snodgrass JG, Corwin J. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*. 1988; 117:34–50. [PubMed: 2966230]
- Tussing AA, Greene RL. Differential effects of repetition on true and false recognition. *Journal of Memory and Language*. 1999; 40:520–533.
- Tversky A, Kahneman D. Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*. 1983; 91:293–315.
- Tversky A, Koehler DJ. Support theory: A nonextensional representation of subjective probability. *Psychological Review*. 1994; 101:547–567.
- Wagner AD, Gabrieli JDE, Verfaellie M. Dissociations between familiarity processes in explicit recognition and implicit perceptual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1997; 23:305–323.
- Zaragoza MS, Mitchell KJ, Payment K, Drivdahl S. False memories for suggestions: The impact of conceptual elaboration. *Journal of Memory and Language*. 2011; 64:18–31. [PubMed: 21103451]

Figure 1.

Episodic overdistribution in false memory for occurrence (Panel A) and false memory for source (Panel B). Panel A data are average subadditivity levels from experiments reported by Brainerd, Reyna, and Aydin (2010), in which delayed tests were administered one week after word lists were studied. Panel B data are average subadditivity levels from experiments reported by Brainerd, Reyna, Holliday, and Nakamura (2012), in which word lists were

sometimes presented in two contexts, sometimes presented in 3 contexts, and were composed of both high and low frequency words.

Table 1

Mean Acceptance Probabilities (Standard Deviations in Parentheses) for Experiments 1–3

List-context/probe type	Word content				Mean
	Hi Frequency	Hi Frequency	Lo Frequency	Lo Frequency	
	Concrete	Abstract	Concrete	Abstract	
Experiment 1					
Targets:					
List 1					
$p(L_1)$	.61 (.26)	.70 (.24)	.49 (.23)	.76 (.23)	.64 (.24)
$p(L_2)$	.53 (.26)	.53 (.26)	.64 (.26)	.59 (.22)	.57 (.25)
$p(L_1 \cap L_2)$	.33 (.26)	.39 (.30)	.25 (.26)	.33 (.26)	.33 (.27)
List 2					
$p(L_1)$	.37 (.29)	.56 (.32)	.18 (.21)	.42 (.28)	.38 (.28)
$p(L_2)$	.51 (.28)	.48 (.28)	.39 (.25)	.47 (.22)	.46 (.26)
$p(L_1 \cap L_2)$	.28 (.26)	.38 (.28)	.18 (.18)	.20 (.21)	.26 (.23)
Distractors:					
$p(L_1)$	.29 (.19)	.40 (.25)	.20 (.18)	.32 (.21)	.30 (.21)
$p(L_2)$	.31 (.21)	.33 (.23)	.23 (.20)	.30 (.19)	.29 (.21)
$p(L_1 \cap L_2)$	.14 (.16)	.43 (.24)	.12 (.17)	.14 (.16)	.21 (.18)
Experiment 2					
Targets:					
List 1					
$p(L_1)$	.55 (.25)	.63 (.27)	.41 (.25)	.56 (.22)	.54 (.25)
$p(L_2)$	.47 (.27)	.50 (.22)	.44 (.21)	.56 (.24)	.49 (.24)
$p(L_1 \cap L_2)$	.35 (.21)	.46 (.28)	.38 (.26)	.40 (.28)	.40 (.26)
List 2					
$p(L_1)$	.48 (.23)	.50 (.30)	.49 (.23)	.45 (.25)	.48 (.25)
$p(L_2)$	.57 (.23)	.52 (.28)	.36 (.25)	.61 (.24)	.52 (.25)
$p(L_1 \cap L_2)$	.44 (.29)	.58 (.26)	.40 (.26)	.36 (.27)	.45 (.27)
Distractors:					

List-context/probe type	Word content				Mean
	Hi Frequency Concrete	Hi Frequency Abstract	Lo Frequency Concrete	Lo Frequency Abstract	
$p(L_1)$	.23 (.19)	.34 (.18)	.16 (.19)	.21 (.22)	.24 (.20)
$p(L_2)$	.26 (.21)	.39 (.22)	.17 (.19)	.39 (.22)	.30 (.21)
$p(L_1 \cap L_2)$	.27 (.21)	.35 (.23)	.20 (.20)	.19 (.19)	.25 (.21)
Experiment 3					
Targets:					
List 1					
$p(L_1)$	.51 (.34)	.56 (.40)	.34 (.34)	.74 (.34)	.54 (.35)
$p(L_2)$	.39 (.36)	.71 (.36)	.40 (.28)	.37 (.38)	.47 (.35)
$p(L_3)$	.37 (.37)	.56 (.35)	.52 (.34)	.48 (.33)	.48 (.35)
$p(L_1 \cap L_2 \cap L_3)$	.43 (.36)	.39 (.32)	.56 (.38)	.53 (.38)	.48 (.36)
List 2					
$p(L_1)$	.39 (.37)	.43 (.36)	.28 (.27)	.61 (.32)	.45 (.33)
$p(L_2)$	.39 (.33)	.51 (.38)	.48 (.31)	.44 (.37)	.46 (.35)
$p(L_3)$	.55 (.33)	.46 (.37)	.53 (.36)	.42 (.32)	.49 (.35)
$p(L_1 \cap L_2 \cap L_3)$	.34 (.34)	.51 (.38)	.38 (.34)	.41 (.34)	.41 (.35)
List 3					
$p(L_1)$	.43 (.34)	.57 (.41)	.40 (.33)	.36 (.32)	.44 (.35)
$p(L_2)$	.45 (.36)	.46 (.41)	.35 (.27)	.46 (.37)	.43 (.35)
$p(L_3)$	.22 (.30)	.53 (.32)	.34 (.35)	.59 (.35)	.42 (.33)
$p(L_1 \cap L_2 \cap L_3)$	.66 (.32)	.65 (.27)	.38 (.29)	.50 (.33)	.55 (.30)
Distractors:					
$p(L_1)$	.21 (.21)	.33 (.22)	.19 (.16)	.23 (.21)	.24 (.20)
$p(L_2)$	.29 (.23)	.39 (.25)	.19 (.19)	.24 (.25)	.28 (.23)
$p(L_3)$	.32 (.20)	.37 (.22)	.15 (.16)	.26 (.17)	.28 (.19)
$p(L_1 \cap L_2 \cap L_3)$	.30 (.24)	.46 (.26)	.19 (.20)	.18 (.18)	.28 (.22)

Table 2

Bias-Corrected Acceptance Probabilities for Experiment 1

List-context/statistic	Word content				M
	Hi Frequency Concrete	Hi Frequency Abstract	Lo Frequency Concrete	Lo Frequency Abstract	
List 1:					
$p(L1)$	.31	.22	.29	.44	.34
$p(L2)$	.22	.16	.41	.27	.28
$p(L1 \cap L2)$					
Observed	.19	.11	.13	.19	.16
Predicted	.12	.20	.24	.20	.19
$T_{=0}$	5.55***	2.75**	4.06***	6.75***	4.78***
$t_{p=0}$	-2.06*	2.24*	3.36**	0.33	0.89
List 2:					
$p(L1)$	.08	.07	.01	.10	.06
$p(L2)$	.21	.11	.16	.17	.16
$p(L1 \cap L2)$					
Observed	.13	.10	.06	.06	.09
Predicted	-.06	.00	.17	.40	.13
$t_{=0}$	3.76**	3.22**	2.37*	2.51*	2.97**
$t_{p=0}$	-5.48***	-3.22**	4.43***	14.39***	2.53*

Note. Observed = bias-corrected response probabilities from Experiment 1 and Predicted = bias-corrected response probabilities from Experiment 1 of Brainerd, Reyna, Holliday, and Nakamura (2012).

*
p < .02
**
p < .008

p < .0001

Table 3

Bias-Corrected Acceptance Probabilities for Experiments 2 and 3

List-context/statistic	Word content				M
	Hi Frequency Concrete	Hi Frequency Abstract	Lo Frequency Concrete	Lo Frequency Abstract	
Experiment 2					
List 1:					
$p(L_1)$	.32	.29	.25	.35	.30
$p(L_2)$	.21	.11	.28	.36	.24
$p(L_1 \cap L_2)$	.07	.12	.18	.21	.15
$t_{=0}$	2.03*	2.83**	4.64***	4.50***	3.50**
List 2:					
$p(L_1)$	.25	.16	.33	.36	.27
$p(L_2)$	.31	.13	.19	.42	.26
$p(L_1 \cap L_2)$	.39	.23	.48	.32	.36
$t_{=0}$	7.38***	5.36***	10.47***	7.27***	7.62***
Experiment 3: Cornell					
List 1:					
$p(L_1)$	.30	.27	.15	.44	.30
$p(L_2)$	.07	.32	.03	.00	.11
$p(L_3)$	.17	.35	.36	.28	.28
$p(L_1 \cap L_2 \cap L_3)$	.20	.16	.30	.35	.25
$T_{=0}$	3.74***	3.34**	5.43***	6.52***	4.76***
List 2:					
$p(L_1)$	.18	.15	.00	.31	.16
$p(L_2)$	.06	.12	.11	.00	.07
$p(L_3)$	.36	.28	.37	.23	.31
$p(L_1 \cap L_2 \cap L_3)$	.11	.27	.12	.24	.19
$T_{=0}$	2.27*	4.53***	2.49*	4.63***	3.48**

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

List-context/statistic	Word content				M
	Hi Frequency Concrete	Hi Frequency Abstract	Lo Frequency Concrete	Lo Frequency Abstract	
List 3:					
$p(L_1)$	.22	.28	.09	.06	.16
$p(L_2)$	.12	.07	.00	.00	.05
$p(L_3)$	.03	.34	.18	.40	.24
$p(L_1 \cap L_2 \cap L_3)$	.43	.42	.12	.32	.32
$T = 0$	7.74***	7.57***	2.72**	6.52***	6.14***

Note. In the last column, bolded numbers are nonconjunctive response probabilities that differ reliably from there corresponding conjunctive response probability.

* $p < .02$

** $p < .008$

*** $p < .0001$