TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Sciences (2013) 1–27

ISSN:1756-8757 print / 1756-8765 online
DOI: 10.1111/tops.12039

Superposition of Episodic Memories: Overdistribution and Quantum Models

Charles J. Brainerd,[a] Zheng Wang,[b] Valerie. F. Reyna[a]

[a]*Department of Human Development, Cornell University*
[b]*School of Communication, Center for Cognitive and Brain Sciences, The Ohio State University*

Received 1 May 2012; received in revised form 15 August 2012; accepted 17 August 2013

Abstract

Memory exhibits episodic superposition, an analog of the quantum superposition of physical states: Before a cue for a presented or unpresented item is administered on a memory test, the item has the simultaneous potential to occupy all members of a mutually exclusive set of episodic states, though it occupies only one of those states after the cue is administered. This phenomenon can be modeled with a nonadditive probability model called overdistribution (OD), which implements fuzzy-trace theory's distinction between verbatim and gist representations. We show that it can also be modeled based on quantum probability theory. A quantum episodic memory (QEM) model is developed, which is derived from quantum probability theory but also implements the process conceptions of global matching memory models. OD and QEM have different strengths, and the current challenge is to identify contrasting empirical predictions that can be used to pit them against each other.

Keywords: Episodic superposition; Overdistribution; Fuzzy-trace theory; Quantum probability; Recognition; Source monitoring

1. Introduction

In this article, we consider how best to explain and model a memory analog of the superposition principle, which is the physical law that stimulated the development of quantum probability. According to that principle, before a measurement of a physical system is taken, the system has a simultaneous potential to occupy all possible combinations of its physical states or properties, notwithstanding that those combinations are mutually

Correspondence should be sent to Charles J. Brainerd, Department of Human Development, Cornell University, Ithaca, NY 14850. E-mail: cb299@cornell.edu

The PC-based software (Windows and R version) that was used to conduct the analyses of the OD model that are reported in this article is available at http://www.human.cornell.edu/hd/brainerd/research.cfm incompatible, but it occupies only one of them after a measurement is taken. The classical demonstrations are the Stern–Gerlach experiment on vertical and horizontal components of angular momentum in spin-½ particles (Gerlach & Stern, 1922) and the Feynman two-slit experiment on electron landing distributions (Feynman, Leighton, & Sands, 1965). In the Stern–Gerlach experiment, sequences of spin measurements show that prior to measurement, *individual* particles have the simultaneous potential to spin up, down, left, and right. In the two-slit experiment, sequences of landing distributions at a detector after a stream of electrons passes through a single aperture A versus a single aperture B versus both apertures show that prior to the landings, *individual* electrons have the simultaneous potential of passing through A, through B, or through neither.

The phenomenon that we focus on in this article, episodic subadditivity, is produced by a memory analog of the two-slit experiment. We sketch the phenomenon in the first section below and summarize results from prior experiments that provide grist for theoretical explanation. In the second section, we discuss an explanation that is predicated on a memory principle called over-distribution and present new experimental data for a model that implements that principle. In the third section, we discuss a quantum superposition explanation that grows out of recent work in quantum cognition by Busemeyer and colleagues (Busemeyer & Bruza, 2012; Busemeyer, Wang, & Lambert-Mogiliansky, 2009; Busemeyer, Wang, & Townsend, 2006) and present results for a model that implements this alternative explanation. In the final section, we briefly compare the results for the two models.

2. Memory slits

The memory paradigm that we will be considering, conjoint recognition, bears considerable resemblance to the Feynman two-slit experiment at a procedural level and also with respect to the findings that it produces. The procedural resemblance lies in the fact that subjects respond to recognition tests that open different "memory slits" through which they are told to allow only items that occupy a stipulated episodic state to pass. There are usually two memory slits, denoted T and R for reasons that will become apparent presently, and the items that are supposed to pass through T occupy episodic states that exclude those that are occupied by the items that are supposed to pass through R. There are two versions of the conjoint recognition paradigm, an item-memory version (Brainerd, Reyna, & Mojardin, 1999) and a source-memory version (Brainerd, Reyna, Holliday, & Nakamura, 2012), which involve different varieties of mutual exclusivity of episodic states. The exclusivity is logical in the item-memory version (an item cannot be both presented and not presented) and empirical in the source-memory version (it can be arranged that an item only originates from one source).

In the item version, subjects first encode a set of memory targets, typically a word, picture, or sentence list. They then respond to a recognition test on which three distinct types of test cues are administered: (a) targets (e.g., *chair*, *robin, steak, ...*); (b) distractors that preserve salient features of targets, usually their meaning (e.g., *sofa*, *oriole,* *roast*, …); and (c) distractors that are semantically and physically unrelated to targets (e.g., *lake*, *car*, *mayor,* …). These three types of test cues are then factorially crossed with three types of recognition instructions: (d) accept targets but reject both related and unrelated distractors (abbreviated as T); (e) accept related distractors but reject both targets and unrelated distractors (abbreviated as R); (f) accept targets and related distractors but reject unrelated distractors (abbreviated as TR). Note the analogy between the T, R, and TR measurements and the A, B, and AB measurements in the Feynman two-slit experiment. Note, too, that the episodic states specified in the instructions about which cues to accept are mutually exclusive, like the notion of a discrete particle having the simultaneous potential to occupy different spatial positions. To be clear about that, a test cue cannot simultaneously be an item that was presented on the study list and an item that was not presented on the study list.

The surprising finding is that many experiments with this paradigm point to the conclusion that before a memory test for a cue is administered, cues for presented items (*chair*, *robin, steak*) or for unpresented related items (*sofa*, *oriole, roast*) have the simultaneous potential to occupy mutually incompatible episodic states (for a review, see Brainerd & Reyna, 2008). Furthermore, an item's remembered episodic state seems to be an emergent property of our memory tests, rather than a stable property of memory itself. How all of this works and how evidence of this episodic superposition is secured can be explained with the aid of Table 1.

In Table 1, the operational distinction between superposition, true memory, false memory, and forgetting is illustrated for the related distractor *Coke* when subjects must accept/reject three statements about it after studying a list on which the targets *Pepsi*, *7-Up*, and *Sprite* appeared: (a) It is a target. (b) It is new but related to a target. (c) It is either a target or new but related to a target. True memory for *Coke* means that subjects remember it as being absent from the study list but related to items that were on the list. If so, the second and third statements will be accepted, and the first will be rejected, which is the pattern in the first column of Table 1. 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 second column of Table 1. Forgetting means that subjects remember *Coke* as not having been on the study list and as not being related to

Table 1
Operational definition of superposition in semantic false memory

	Memory State			
	True Memory	False Memory	Superposition	Forgotten
Coke: target?	No	Yes	Yes	No
Coke: new but related to target?	Yes	No	Yes	No
Coke: target or new but related to target?	Yes	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.

anything on the study list—as being an unrelated distractor rather than a target or a related distractor. If so, all three statements will be rejected, which is the pattern shown in the last column of Table 1. Superposition refers to the logically contradictory situation in which *Coke* is remembered as being a target *and* a related distractor. If so, all three statements will be accepted, which is the pattern in the third column of Table 1. Finally, note that if the test probe were a target (e.g., *Pepsi*) rather than a related distractor, the third and fourth columns of Table 1 would still be the superposition and forgetting columns, but false memory would correspond to the first column and true memory to the second.

Table 1 also illustrates the key point that superposition cannot be separated from true or false memory in a standard experiment in when subjects only make old/new judgments because, assuming appropriate corrections for response bias: (a) Acceptance of "It is a target" for a target does not guarantee rejection of "It is new but related to a target" (true memory), and (b) acceptance "It is a target" for a related distractor does not guarantee rejection of "It is new but related to a target" (false memory) Thus, when *Pepsi* is presented as a cue for an old/new judgment in a standard design, an "old" response may mean true memory or superposition, and likewise, when *Coke* is presented as a cue for an old/new judgment in a standard design, an "old" response may mean false memory or superposition. However, conjoint recognition designs are able to disentangle these possibilities by factorially crossing the three types of test cues (targets, related distractors, unrelated distractors) with the three types of probes in Table 1, which are denoted T (for "target"), R (for "related distractor"), and T∪R (for "target or related distractor"), respectively. Only one probe is administered per cue to individual subjects, naturally, to avoid repeated testing effects, but across subjects, all three probes are administered for each cue.

We now show how conjoint recognition separates superposition from true memory, false memory, and forgetting, and to do that, we need some simple notation for measurements of these states. For any arbitrary test cue, let $P(\mathrm{T}\cap{\sim}\mathrm{R})$, $P({\sim}\mathrm{T}\cap\mathrm{R})$, $P(\mathrm{T}\cap\mathrm{R})$, and $P({\sim}\mathrm{T}\cap{\sim}\mathrm{R})$ be the probabilities that the cue occupies the target, related distractor, both target and related distractor, and unrelated distractor states, respectively. It is easy to see how one would normally obtain empirical estimates of the probabilities of the cue occupying the target, related distractor, and unrelated distractor states. Respectively, those values are just the probability of accepting a cue under T instructions, which will be denoted $P(\mathrm{T})$, the probability of accepting the cue under R instructions, which will be denoted $P(\mathrm{R})$, and the probability of *rejecting* a cue under TR instructions, which will be denoted $1-P(\mathrm{T}\cup\mathrm{R})$. Estimating the "both" (superposition) state is quite another matter, owing to subjects' metacognitive knowledge of logical exclusivity. We cannot administer memory probes such as "*Coke* is both a target and a related distractor" or "*Pepsi* is both a target and a distractor." Thanks to meta-cognitive knowledge, subjects would simply respond "no" to such probes, and if they responded "yes," they would probably be dropped from the experiment for inattentiveness or sheer perversity. This obstacle to direct measurement is overcome by taking advantage of a mathematical constraint that flows from target-distractor exclusivity.

Specifically, we take advantage of a constraint between $P(\mathrm{T})$, $P(\mathrm{R})$, and $P(\mathrm{T}\cup\mathrm{R})$ that happens to be the same as the additivity constraint that classical physics imposes in the Feynman two-slit experiment when A and B are both open. Because targets and distractors are mutually exclusive, it is objectively true that $P(\mathrm{T}\cap\mathrm{R}) = 0$. From the perspective of episodic memory, however, we can entertain the possibility that some cues are accepted under T instructions because retrieved information is consistent with *both the T and R states*, rather than T only, and that the same circumstance may govern responses under R instructions. If either or both of these scenarios holds, then $P(\mathrm{T}\cap\mathrm{R}) > 0$. To estimate the actual value of $P(\mathrm{T}\cap\mathrm{R})$, we exploit 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, $P(\mathrm{T}) + P(\mathrm{R}) = P(\mathrm{T}\cup\mathrm{R}) + P(\mathrm{T}\cap\mathrm{R})$. Note that if $P(\mathrm{T}\cap\mathrm{R}) = 0$, the relation between $P(\mathrm{T}) + P(\mathrm{R})$ and $P(\mathrm{T}\cup\mathrm{R})$ must be additive; that is, $P(\mathrm{T}) + P(\mathrm{R}) = P(\mathrm{T}\cup\mathrm{R})$. Suppose, however, that some test cues occupy the logically contradictory superposition state of being remembered as both presented and unpresented, so that $P(\mathrm{T}\cap\mathrm{R}) > 0$. It follows that rather than being additive, the relation between $P(\mathrm{T}) + P(\mathrm{R})$ and $P(\mathrm{T}\cup\mathrm{R})$ must be *subadditive*. Specifically, $P(\mathrm{T}) + P(\mathrm{R}) > P(\mathrm{T}\cup\mathrm{R})$ because according to the rule, $P(\mathrm{T}) + P(\mathrm{R}) - P(\mathrm{T}\cap\mathrm{R}) = P(\mathrm{T}\cup\mathrm{R})$, and the larger the difference between $P(\mathrm{T}) + P(\mathrm{R})$ and $P(\mathrm{T}\cup\mathrm{R})$, the more items are being remembered as being both presented and unpresented. It is crucial to bear in mind here that in an actual experiment, no particular relation between $P(\mathrm{T}) + P(\mathrm{R})$ and $P(\mathrm{T}\cup\mathrm{R})$ is forced by the conjoint recognition design; indeed, the observed relation could be *superadditive* as well additive or subadditive.

For a target such as *Pepsi* or a related distractor such as *Coke*, then, the quantity $P(\mathrm{T}\cap\mathrm{R}) = P(\mathrm{T}) + P(\mathrm{R}) - P(\mathrm{T}\cup\mathrm{R})$ indexes the extent to which cues are superposed on multiple episodic states of an experiment before episodic memory is measured—notwithstanding the mutual exclusivity of those states. Positive values demonstrate episodic superposition; negative and zero values rule it out. Now that we understand how episodic superposition is measured, what do the data show?

They provide strong evidence of superposition. The initial evidence was reported by Brainerd and Reyna (2008), who reviewed over 100 previously published sets of conjoint-recognition data. All of the experiments conformed to the general methodology that was described above, with the materials that supplied the targets and related distractors varying from words to pictures to sentences to narratives. A convenient method of summarizing the data for both targets and related distractors is to compute the ratio $[P(\mathrm{T}) + P(\mathrm{R})] \div P(\mathrm{T}\cup\mathrm{R})$ for each data set, where, again, $P(\mathrm{T})$, $P(\mathrm{R})$, and $P(\mathrm{T}\cup\mathrm{R})$ are the observed probabilities of accepting cues (either targets or related distractors) under T, R, and TR instructions, respectively. Note that if $P(\mathrm{T}\cap\mathrm{R}) > 0$, so that the relation between $P(\mathrm{T}) + P(\mathrm{R})$ and $P(\mathrm{T}\cup\mathrm{R})$ is subadditive for a data set, the ratio $[P(\mathrm{T}) + P(\mathrm{R})] \div P(\mathrm{T}\cup\mathrm{R})$ must exceed unity. We computed this ratio for all of the data sets that were reviewed by Brainerd and Reyna. The results appear in Fig. 1, where values of $[P(\mathrm{T}) + P(\mathrm{R})] \div P(\mathrm{T}\cup\mathrm{R})$ have been plotted against the corresponding values of $P(\mathrm{T}) + P(\mathrm{R})$.[1]

The subadditivity picture could not be more apparent, for both targets and related distractors. The target data are plotted in panel A, and the related distractor data are plotted

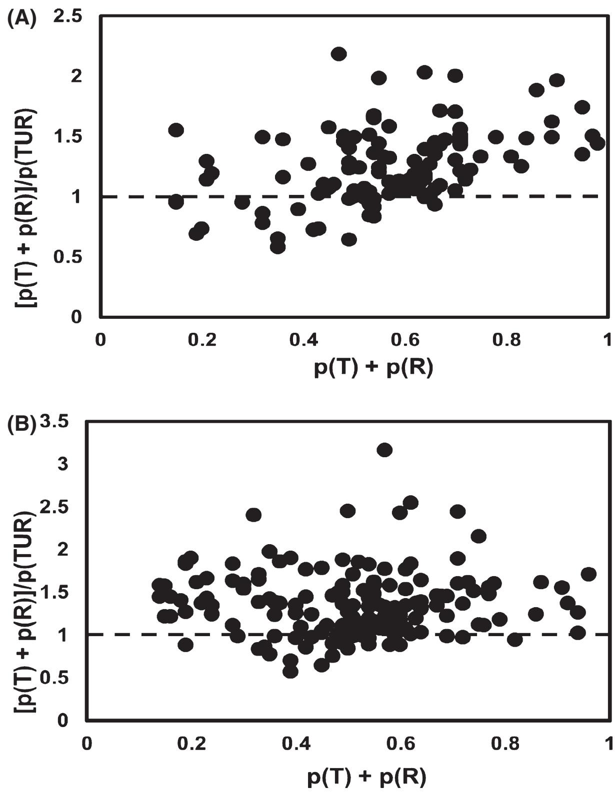

Fig. 1. The superposition effect in the conjoint recognition data sets reviewed by Brainerd and Reyna (2008), which is the statistic $[P(\mathrm{T}) + P(\mathrm{R})] \div P(\mathrm{T}\cup\mathrm{R})$. Data for presented targets are plotted in panel A, and data for unpresented related distractors are plotted in Panel B. In both instances, dots that fall above the line indicate data sets in which a certain proportion of the items were remembered as being both targets and related distractors.

in panel B. The question of interest is whether $[P(\mathrm{T}) + P(\mathrm{R})] \div P(\mathrm{T}\cup\mathrm{R})$ exceeds unity, and it can be seen that it does in the great preponderance of data sets, for both targets and related distractors. The subadditivity patterns in panels A and B can be tested for statistical significance in two ways. First, one can compute χ^2 tests of goodness-of-fit, or second, one can compute one-sample t tests. In the χ^2 tests, additivity/superadditivity is the null hypothesis and subadditivity is the alternative hypothesis, so that one simply counts the number of points that fall above the unity line versus the number that fall on or below it and evaluates the null hypothesis's prediction that all points will fall on or below the line. In the t-tests, one merely computes the mean of the ratio $[P(\mathrm{T}) + P(\mathrm{R})] \div P(\mathrm{T}\cup\mathrm{R})$ for the plotted data and then computes a one-sample test in which the hypothesized mean value of the ratio is ≤ 1. With the data sets in Fig. 1, not surprisingly, these χ^2 and t-tests produce null hypothesis rejections at very high levels of confidence (all ps $< .001$).

More recent findings that revealed the same pattern as in Fig. 1 and tested hypotheses about manipulations that might affect levels of superposition were reported by Brainerd, Reyna, and Aydin (2010). In all of the results in Fig. 1, superposition was measured at the group level (i.e., using data that had been pooled across subjects). A methodological adjustment in the experiments reported by Brainerd et al. allowed superposition to be computed at the level of individual subjects as well. Robust superposition was again observed in the group data, but further, it proved to be the dominant pattern at the individual level. Within each of the various conditions of the first experiment reported by Brainerd et al., for instance, 85% of the subjects, on average, exhibited superposition.

As mentioned, there are two versions of the conjoint-recognition paradigm. Although the above evidence of superposition is extensive, all of it is for the item version. It is natural to ask whether the picture is the same for the source version. This is far from idle speculation because the nature of the target-distractor exclusivity is fundamentally different in the two paradigms. It is logical in the item version but empirical in the source version. The source procedure runs as follows. First, instead of encoding one set of memory targets, subjects encode *two or more* sets. In the simplest case, one set of targets is encoded in context C1, while a second set is encoded in a different context C2. The sets are mutually exclusive because no target that appears in one context can appear in the other. Typically, the two contexts are List 1 and List 2, with targets on List 1 appearing in a font (e.g., Arial) and color (e.g., red) that are different from the font (e.g., Broadway) and color (e.g., blue) in which List 2 targets appear. Second, subjects respond to a recognition test on which three distinct types of test cues are administered: (a) C1 targets (e.g., *factory, shoes, diamond, ...*); (b) C2 targets (e.g., *tower, library, home, …*); and (c) unpresented distractors that are semantically and physically unrelated to targets (e.g., *cell, engine, ocean, …*). The three types of cues are factorially crossed with three types of recognition instructions: (d) T: accept C1 targets but reject C2 targets and distractors; (e) R: accept C2 targets but reject C1 targets and distractors; and (f) TR: accept both C1 and C2 targets but reject distractors. Thus, the incompatibility between targets and distractors is empirical in that no target that was presented in C1 was also presented in C2. Subjects are informed of this arrangement before they encode any of the targets, and they are reminded of it before they respond to the recognition test.

Nevertheless, Brainerd et al. (2012) found, in a series of experiments, that this paradigm produced robust evidence of superposition in source memory. Across the experiments, at the group level, the average value of the ratio $[P(\mathrm{T}) + P(\mathrm{R})] \div P(\mathrm{T} \cup \mathrm{R})$ was far above unity. At the individual level, across experiments, nearly three-quarters of the subjects displayed superposition. Thus, regardless of whether different episodic states were mutually exclusive for logical or empirical reasons, data from conjoint-recognition experiments point to the conclusion that before recognition tests are administered, episodic memory is in a superposition of incompatible episodic states.

3. Modeling episodic superposition: Model and data

Superposition was not a serendipitous finding. It was predicted by a theoretical conception of how retrieval operates, which is called overdistribution (OD). That conception is implemented in a mathematical model of the paradigm, whose parameters have specific process interpretations. When this OD model is analyzed, subadditive relations between $P(\mathrm{T}) + P(\mathrm{R})$ and $P(\mathrm{T} \cup \mathrm{R})$ fall out as predictions; that is, either these relations are subadditive or the model will not fit conjoint-recognition data, so that statistical tests of fit are also tests of superposition.

This forms the substance of the present section, which begins by presenting the model, the process interpretation of its parameters, and the derivation of the superposition effect. The derivation specifies that superposition is caused by certain types of memory representations whose retrieval distributes test cues to too many episodic states, even when those states are mutually exclusive. Next, we show how this model is applied in research, by reporting a new experiment with the source version of conjoint recognition, fitting the model to the data, estimating its parameters, and showing how the parameters predict exact amounts of superposition.

3.1. Overdistribution model

The model (Brainerd et al., 1999; Brainerd, Wright, Reyna, & Mojardin, 2001) implements fuzzy-trace theory's dual-trace analysis of representation (Reyna & Brainerd, 1995). When items are presented during the study phase, subjects are assumed to store two types of episodic traces in parallel: verbatim and gist. Verbatim traces are representations of items' exact surface content—their orthography and phonology in the case of words—along with contextual features (criterial and noncriterial) that accompany their presentation. (Storing contextual features is what makes memory representations "episodic," of course. Criterial features distinguished one context from another, whereas noncriterial features do not.) Gist traces are representations of items' semantic content and other relational information, along with contextual features (because gist traces, too, are episodic). Here, a distinction that has been significant in explaining counterintuitive findings about false memory is that some of the semantic content that is stored in gist traces is particular to individual items while other semantic content connects multiple items (Brainerd & Reyna, 2012). For instance, suppose, as in our earlier example, that *Pepsi*, *7-Up,* and *Sprite* all appear on a study list. "Cola" is a semantic feature that applies to one of these words, while "soft drink" applies to all. On conjoint recognition tests, test cues provoke retrieval of verbatim and gist traces, both of the cued items and other items, in parallel, with the two types of traces leading to the same response under some types of instructions but different responses under other instructions. Because verbatim traces store realistic surface features, their retrieval produces vivid mental reinstatement of their prior presentation, a phenomenology that is usually called recollection in the memory literature. Because gist traces do not store realistic surface features, their retrieval produces global feelings of confidence that items with this meaning were recently studied without vivid recollection of their actual presentation, a phenomenology that is usually called familiarity (although vivid gist memories can also occur for semantically related lists; Brainerd & Reyna, 2005).

The OD model that implements these distinctions runs as follows.[2] As before, for an arbitrary test cue, let $P(\mathrm{T})$, $P(\mathrm{R})$, and $P(\mathrm{T} \cup \mathrm{R})$ be the probabilities of accepting a T probe (agreeing that it is a target), an R probe (agreeing that it is a related distractor), or a TR probe (agreeing that it is not an unrelated distractor). When the test cue is a target, such as *Pepsi*, the model's expressions for these empirical probabilities are

$$P_T(\mathrm{T}) = V_T + (1 - V_T)(1 - E_T)G_T, \tag{1}$$

$$P_T(\mathrm{R}) = (1 - V_T)E_T + (1 - V_T)(1 - E_T)G_T, \tag{2}$$

and

$$P_T(\mathrm{T} \cup \mathrm{R}) = V_T + (1 - V_T)E_T + (1 - V_T)(1 - E_T)G_T. \tag{3}$$

V_T is the probability that subjects retrieve verbatim traces of *Pepsi*, E_T is the probability that they retrieve verbatim traces of other related targets (e.g., *Sprite*; that is, a cue for one target has the ability to retrieve the verbatim trace of another target, causing subjects to conclude that the target is a related distractor), and G_T is the probability they retrieve gist traces of *Pepsi.* (The T-subscripts indicate that the retrieval cue is a target rather than a related distractor.) These expressions implement simple, uncontroversial psychological ideas. For instance, if verbatim traces of *Pepsi* are retrieved, subjects ought to accept T probes, reject R probes, and accept TR probes because they recollect this word's prior presentation. Also, if neither verbatim traces of *Pepsi* nor verbatim traces of closely related targets are retrieved but gist traces of *Pepsi* are retrieved, subjects ought to accept all three types of probes because the semantic information in such traces is consistent with *Pepsi* being a target or a related distractor.

Although the psychological ideas in Eqs. 2–4 are uncontroversial, the expressions lead to the surprising $[P_T(\mathrm{T}) + P_T(\mathrm{R})] > P_T(\mathrm{T}\cup\mathrm{R})$ result. To see that, first add the right sides of Eqs. 2 and 3, to produce the sum $V_T + (1-V_T)E_T + 2(1-V_T)(1-E_T)G_T$, and then subtract the right side of Eq. 4, the difference being $(1-V_T)(1-E_T)G_T$. Thus, the model predicts superposition as a qualitative pattern, and quantitatively, it predicts *exact* amounts of superposition once the three parameters are estimated. Psychologically, the superposition prediction falls directly out of the notion that a gist trace can be used to accept any cue (e.g., *Pepsi*) whose semantic content (e.g., "soft drink") matches that of the trace, regardless of whether it happens to be a target or a distractor. This means that superposition is due to a gist-driven *overdistribution* process. Logically, a target cue such as *Pepsi* cannot be both a target and related distractor, but its gist trace is consistent with both episodic states. Hence, processing gist traces on tests for targets will tend to distribute them to the related distractor state as well as to the target state, notwithstanding the logical incompatibility, and processing gist traces on tests for related distractors will distribute them to the target state as well as the related distractor state.

To see that this is so, consider the model's expression for acceptance of related distractor cues such as Coke under T, R, and TR instructions, where the R-subscripts denote that the test cue is a related distractor rather than a target:

$$P_{\mathrm{R}}(\mathrm{T}) = (1 - V_{\mathrm{R}})P_{\mathrm{R}} + (1 - V_{\mathrm{R}})(1 - P_{\mathrm{R}})G_{\mathrm{R}}, \tag{4}$$

$$P_{\mathrm{R}}(\mathrm{R}) = V_{\mathrm{R}} + (1 - V_{\mathrm{R}})(1 - P_{\mathrm{R}})G_{\mathrm{R}}, \tag{5}$$

and

$$P_{\mathrm{R}}(\mathrm{T} \cup \mathrm{R}) = V_{\mathrm{R}} + (1 - V_{\mathrm{R}})P_{\mathrm{R}} + (1 - V_{\mathrm{R}})(1 - P_{\mathrm{R}})G_{\mathrm{R}}. \tag{6}$$

V_{R} and G_{R} are the probabilities of retrieving verbatim and gist traces, respectively, of related distractors' corresponding targets (e.g., *Pepsi* for the distractor *Coke*). P_{R} is the probability that subjects retrieve ersatz verbatim memories of *distractors*—traces that produce illusory vivid recollections of distractors' prior "presentations." The presence of this process in the model is consistent with a well-established finding in the literature: when subjects falsely accept related distractors as targets, it is often accompanied by phantom recollective experiences (e.g., Frost, 2000; Heaps & Nash, 2001). The key point is that when the right sides of Eqs. 5 and 6 are added and the right side of Eq. 7 is subtracted from the sum, superposition shows up again. Specifically, $[P_{\mathrm{R}}(\mathrm{T}) + P_{\mathrm{R}}(\mathrm{R})] > P_{\mathrm{R}}(\mathrm{T} \cup \mathrm{R})$ because $[P_{\mathrm{R}}(\mathrm{T}) + P_{\mathrm{R}}(\mathrm{R})] - P_{\mathrm{R}}(\mathrm{T} \cup \mathrm{R}) = (1 - V_{\mathrm{R}})(1 - P_{\mathrm{R}})G_{\mathrm{R}}$. Another point of interest is that the psychological basis for superposition is the same OD principle: Although *Coke* is a related distractor and not a target, the fact that gist retrieval is consistent with both episodic states means that it will distribute *Coke* to the target state as well as the related distractor state.

3.2. Illustrative experiment

Next, we briefly apply OD to the data of an experiment that used the source version of conjoint recognition. Mathematically, OD is identical for the source and item paradigms, the only difference lying in some of the memory processes that the parameters refer to. We relegate the discussion of those differences to the Appendix.

Seventy subjects participated in a design that involved two steps. First, they were exposed to three word lists (List 1, List 2, and List 3), each consisting of 36 items (2-word starting and ending buffers, plus 32 focal words). To ensure that each list was distinctive, the background color against which the words were presented differed between the lists, as did the fonts in which the words were printed. Second, the subjects responded to a 192-item test list composed of four types of test cues: List 1 targets, List 2 targets, List 3 targets, and distractors. The four types of test cues were factorially crossed with four types of recognition probes, to which the subjects responded "accept" or "reject": (a) presented on List 1; (b) presented on List 2; (c) presented on List 3; and (d) presented on List 1 or List 2 or List 3. Of the 192 test items, there were 12 exemplars of each of the four types of cues, resulting 48 probes in total. Each cue was tested with only one of the four types of recognition probes, within an individual subject. However, the cues were rotated through the four types of recognition probes across subjects—so that overall, each cue was tested with each type of probe one-quarter of the time. Thus, the methodology extended the simple two-slit design to three slits (List 1, List 2, List 3), which can be handled with a relatively minor adjustment of the model (see Appendix).

The experiment also included two content manipulations, word concreteness and word frequency, that were factorially crossed with each other on both the study and test lists. These content manipulations were included for theoretical reasons—that there were reasons to predict that superposition would be more marked for abstract words and low-frequency words (Brainerd et al., 2012). The exact reasons are that experiments in the false memory literature suggest that abstract words create weaker verbatim traces that concrete words and that low-frequency words create weaker verbatim traces than high-frequency words (Brainerd & Reyna, 2005) so that abstract words and low-frequency words would lower the values of the verbatim parameters of the OD model, V_T and V_R, which would then increase superposition by increasing the values of the $(1-V_T)(1-P_T)G_T$ and $(1-V_R)(1-P_R)G_R$ terms. However, in this article, we do not consider this feature of the experiment further. Instead, we concentrate on the fundamental question of whether OD, which predicts superposition as a qualitative pattern, fits the data and is able to make quantitative predictions that fall within traditional statistical boundaries.

The fit of the model has already been established by Brainerd et al. (2012)—using the usual indexes of fit, such as G^2 difference. The G^2 difference is $G^2_{OD}-G^2_{saturated} = -2\cdot\ln$ (Likelihood(data|OD)) when $p_i(yes|m)$ is based on the OD model, and $G^2_{saturated} = -2\cdot\ln$ (Likelihood(data|saturated)) when $p_i(yes|m)$ is based on the saturated model. In the saturated model, $p_i(yes|m) = n_i(yes)$ is the observed proportions in the experiments. That is, the saturated model has the best fit possible since it perfectly describes the observed data; thus, it is used as a comparison against the OD model. For the present experiment, a G^2 critical value (based on $df = 24$) of 36.42 was required to reject the null hypothesis of model fit at the .05 level for the experiment as a whole. The model was fit using 72 parameters to 96 data points, which leaves 24 degrees freedom for the experiment as a whole: There are 12 list-item conditions with 8 degrees of freedom (free response probabilities) per condition. As shown in Eqs. A7–A14 in the Appendix, the model estimates a total of six parameters for each of those conditions, four memory parameters and two response-bias parameters. Thus, there are two remaining degrees of freedom for each condition, and $12 \times 2 = 24$ degrees of freedom for the experiment as a whole. The observed G^2 difference value was 29.43, so that fit was acceptable. (At the level of individual conditions, the null hypothesis of fit could not be rejected in any of the 12 conditions.) With fit established, we compared observed levels superposition for the various conditions of this experiment to the levels that were predicted by the model. The factorial structure was 3 lists $\times$ 2 levels of concreteness (concrete vs. abstract) $\times$ 2 levels of frequency (high vs. low), for a total of 12 conditions. Observed values of $P(L_1)$, $P(L_2)$, $P(L_3)$, and $P(L_1 \cup L_2 \cup L_3)$, for each type of target cue (List 1, List 2, List 3) in each of the 12 conditions, were calculated and corrected for response bias, using the distractor data (see footnote 1). Those values are reported in Table 2. Next, the observed level of superposition was calculated for each condition, which is the statistic $[P(L_1) + P(L_2) + P(L_3)]-P(L_1 \cup L_2 \cup L_3)$, and those values are also reported in Table 2. It can be seen that this experiment produced robust superposition effects: Across the 12 conditions, the average

Table 2
Bias-corrected acceptance probabilities and OD statistics

	Word Content			
List-Context/Statistic	High-frequency Concrete	High-frequency Abstract	Low-frequency Concrete	Low-frequency Abstract
List 1 targets				
p(L1?)	0.37	0.27	0.48	0.39
p(L2?)	0.16	0.12	0.35	0.40
p(L3?)	0.17	0.13	0.28	0.36
p(L1 or L2 or L3?)	0.34	0.28	0.51	0.46
Subadditivity[a]	0.36	0.24	0.60	0.69
List 2 targets				
p(L1?)	0.16	0.07	0.20	0.25
p(L2?)	0.18	0.30	0.35	0.32
p(L3?)	0.14	0.14	0.21	0.21
p(L1 or L2 or L3?)	0.32	0.38	0.36	0.41
Subadditivity[b]	0.16	0.13	0.40	0.37
List 3 targets				
p(L1?)	0.15	0.15	0.23	0.24
p(L2?)	0.26	0.20	0.24	0.37
p(L3?)	0.31	0.24	0.36	0.35
p(L1 or L2 or L3?)	0.38	0.27	0.43	0.40
Subadditivity[c]	0.34	0.38	0.40	0.56

Notes. All tabled values were corrected for response bias, using the two-high threshold correction of signal detection theory.
[a]Referring to Eqs. A7–A10, these subadditivity values were computed from $(p_{\mathrm{M,L1}} + p_{\mathrm{E2,L1}} + p_{\mathrm{E3,L1}}) - p_{\mathrm{I,L1}}$, after each of these probabilities had been corrected for the influence of response bias.
[b]These subadditivity values were computed from $(p_{\mathrm{M,L2}} + p_{\mathrm{E1,L2}} + p_{\mathrm{E3,L2}}) - p_{\mathrm{I,L2}}$, after each of these probabilities had been corrected for the influence of response bias.
[c]These subadditivity values were computed from $(p_{\mathrm{M,L3}} + p_{\mathrm{E2,L3}} + p_{\mathrm{E1,L3}}) - p_{\mathrm{I,L3}}$, after each of these probabilities had been corrected for the influence of response bias.

difference between the expected value of $[P(\mathrm{L}_1) + P(\mathrm{L}_2) + P(\mathrm{L}_3)] - P(\mathrm{L}_1 \cup \mathrm{L}_2 \cup \mathrm{L}_3)$, which is zero, and its actual value is .38. This is a highly reliable difference. At the level of individual subjects, averaging across the 12 conditions, more than 70% exhibited superposition.

The next question is whether OD predicts observed levels of superposition with high accuracy. We know from the previous section that in the simple two-slit design, the predicted amount for targets in a condition is just $(1-V_{\mathrm{T}})(1-E_{\mathrm{T}})G_{\mathrm{T}}$ (The expression is slightly different for the three-slit design; see Appendix.). To predict exact levels of superposition for each condition, one simply estimates the parameters for each condition and computes predicted values. That was done, and the predicted values were compared to the corresponding observed values. The results are shown in Fig. 2, where it can be seen that the predicted-observed correspondence was quite good, with 95% of the variance accounted for. Beyond fit, another question is the degree to which OD's predictions exactly calibrate reality that there is correspondence between the absolute values of each

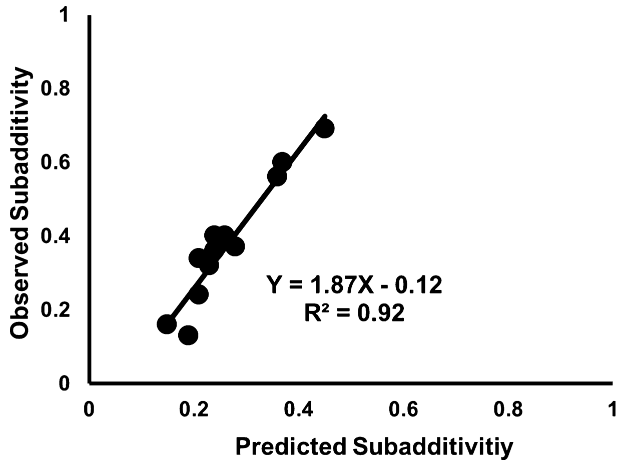

Fig. 2. Observed levels of superposition across the List × Concreteness × Frequency combinations of the experiment versus levels of superposition predicted by the OD model. The measure of the superposition is the subadditivity statistic [P(C1) + P(C2) + P(C3)] ÷ P(C1∪C2∪C3).

condition's observed and predicted superposition statistics. If calibration is exact, the best-fitting regression equation will have an intercept of 0 and a slope of 1. It can be seen in Fig. 2 that OD overpredicted the data of this experiment. The intercept was slightly below zero, but the slope was 1.87.

To investigate whether overprediction is the rule with OD, we reanalyzed the data of two previous experiments that used the same procedure and the same list conditions as the present experiment—specifically, Experiments 3 and 4 of Brainerd et al. (2012). The results are displayed in Fig. 3, with data from Experiment 3 appearing in Panel A and data from Experiment 4 appearing in Panel B. Comparing the best-fitting regression equations across the data sets in Figs. 2 and 3, there is a slight tendency for the model to over-predict superposition: Across data sets, the mean value of the intercept is −.02 and the mean value of the slope is 1.25.

4. Modeling episodic superposition: Quantum model

Built upon basic quantum probability principles, an alternative model called quantum episodic memory (QEM) model was formulated to account for the episodic OD effect. As discussed in this special issue (Wang, Busemeyer, Atmanspacher, & Pothos, 2013), quantum probability theory was developed during the process of inventing a new theory of physics, quantum mechanics, to explain physical phenomena that seemed paradoxical from a classical physics (and hence classical probability) perspective in the 1920s. Likewise, the subadditivity memory result is surprising and paradoxical only from the classical probability perspective. Then, why not drop the classical perspective and see whether quantum probability theory can explain this result? The QEM model is motivated by this

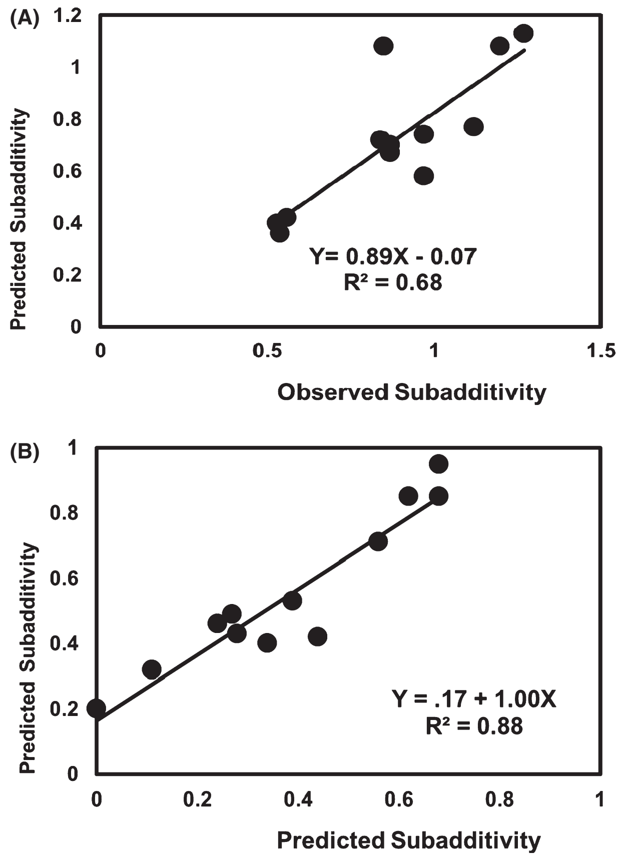

Fig. 3. Observed levels of superposition versus levels of superposition predicted by the OD model for the list conditions of two experiments reported by Brainerd et al. (2012). Panel A contains data from their Experiment 3, while Panel B contains data from their Experiment 4. In both instances, the measure of superposition is the subadditivity statistic $[P(\mathrm{C1}) + P(\mathrm{C2}) + P(\mathrm{C3})] \div P(\mathrm{C1} \cup \mathrm{C2} \cup \mathrm{C3})$.

idea. It is derived from the general principles of quantum probability theory described next.

4.1. Some principles of quantum probability theory

Superposition is a basic principle of quantum probability theory. Classical probability theory assumes that at any moment, a person is in a *definite* state with respect to possible cognitive states. For example, in the 3-list conjoint recognition experimental paradigm, it assumes that during the retrieval process, the memory-cue matching is in any one of the following exclusive states: either verbatim List 1, or verbatim List 2, or verbatim List 3, or the gist, or the distractors. This definite state can change (stochastically) across time; but at each moment, the state is still definite, and the retrieval process produces a definite sample path. In contrast, quantum probability theory assumes that at any moment in the retrieval process, a person is in an *indefinite* (i.e., superposition) state of episodic awareness until a response is made. A superposition state is defined by the fact that all five possibilities (verbatim List 1, verbatim List 2, verbatim List 3, the gist, and the distractors) have the *potential* for being expressed, but none of the five can be assumed at any moment. The superposition conception resonates with the fuzzy, ambiguous, uncertain feelings associated with memory.

Classical probability theory represents an event, such as E, as a subset of the sample space Ω, which contains all the events. A state is represented by a function P: $2^{\Omega} \rightarrow [0,1]$, which assigns probabilities to events. In other words, P(E) is the probability assigned to event E∈Ω. In contrast, quantum probability theory represents an event, such as E, as a subspace of a vector space (a Hilbert space), V, and the vector space V contains all the events. This is illustrated in Fig. 4. It is important to note that the vector space can be arbitrarily high-dimensional although for the simplicity of illustration, a simple three-dimensional vector space is used in this example. In Fig. 4, the event E subspace is a plane spanned by the basis vectors X and Y, that is, the horizontal plane defined by the *X*, *Y* axes in the figure. A different event can be a 1-dimensional ray along the *X*, *Y*, or *Z* axe, or a two-dimensional plane spanned by the *X*, *Z* axes or the *Y*, *Z* axes, or even a three-dimensional subspace spanned by *X*, *Y*, and *Z* (which then becomes the identity). A state, such as a memory state, is represented by a unit length vector S ∈ V, which assigns probabilities to events (the red line in Fig. 4). Corresponding the subspace E is a projector, M_E, which projects points in V (including the state S) onto E. The probability

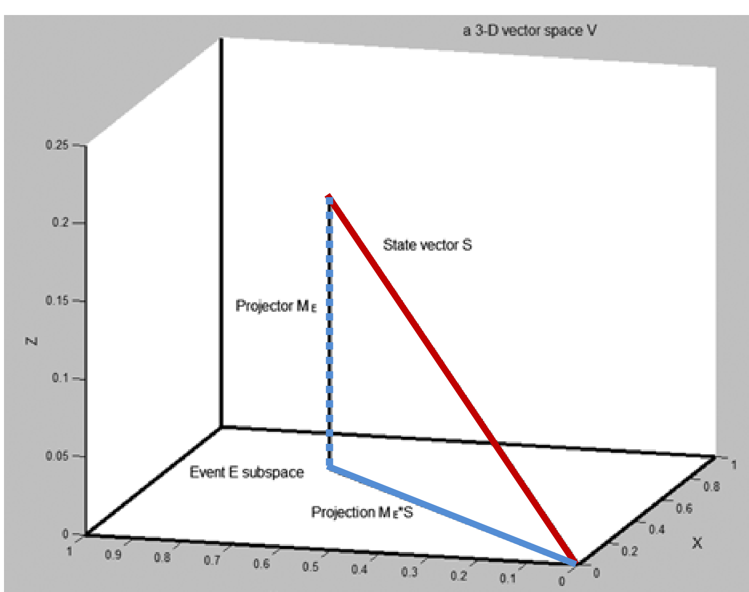

Fig. 4. A simple three-dimensional illustration of basic quantum probability principles used in QEM.

of an event is the squared length of the projection projecting the state vector S to the subspace E: $P(\mathrm{E}) = \|\mathrm{M_E} \cdot \mathrm{S}\|^2$. In Fig. 4, the dotted blue line represents the projector $\mathrm{M_E}$, which projects the state S onto the event subspace E. The solid blue line is the resulting projection $\mathrm{M_E} \cdot \mathrm{S}$, whose squared length is the probability of the event E for the state S.

Observing the figure, it becomes clear that the geometric relations between the state vector S and the subspaces of events determine the probability of the events. Also, the state vector S in fact is a superpositional state with *potential* to be projected onto all possible event subspaces. The QEM model presented next is a five-dimensional model, which is difficult to be illustrated visually and thus is only presented in a precise and abstract manner, but it follows the exact conceptual ideas as shown in the simple example in Fig. 4. In addition, there are many other differences between the quantum versus classical probability theory (see the introduction article to this special issue), but the current QEM model is simple and only needs to use a few mathematical principles of quantum probability theory, as described above, to account for the subadditivity memory result.

4.2. Quantum episodic memory model

First, the QEM model uses a *superposition* of feature vectors to represent the ambiguous memory state. Using feature vectors to represent memory is commonly seen in cognitive models of memory based on classical probability theory, such as MINERVA 2 (Hintzman, 1986, 1988), Search of Associative Memory (SAM; Raaijmakers & Shiffrin, 1981), Retrieving Efficiently from Memory (REM; Shiffrin & Steyvers, 1997), and the Noisy Exemplar model (NEMO; Kahana & Sekuler, 2002). Specifically, in QEM, the memory state is represented using a five-dimensional vector space spanned by five orthonormal basis vectors: V_1, a vector representing verbatim List 1 features; V_2, a vector representing verbatim List 2 features; V_3, a vector representing verbatim List 3 features; G, a vector representing gist features; and U, a vector representing distractor features. The memory state is theorized to be an ambiguous state that superposes these five types of features. To capture this ambiguity, which cannot be done in the aforementioned classical feature vector models of memory, QEM represents the memory state using a unit length vector S formed by a *superposition* of the five basis vectors:

$$\begin{aligned} \mathrm{S} &= (v_1 \cdot \mathrm{V}_1 + v_2 \cdot \mathrm{V}_2 + v_3 \cdot \mathrm{V}_3 + g \cdot G + u \cdot \mathrm{U})/\mathrm{c}, \\ \text{where} \quad c &= (|v_1|^2 + |v_2|^2 + |v_3|^2 + |g|^2 + |u|^2). \end{aligned} \tag{7}$$

In Eq. 7, v_1, v_2, v_3, g, and u are the amplitudes of each basis vector, and c normalizes the sum of the five basis states to maintain the unit length of S. (This normalization procedure is standard in probability models, either quantum or classical, to keep the probabilities for all possible responses happening to the state S sum up to one.) As detailed later, the memory state S is constructed depending on the cue presented on each test (i.e., List 1, List 2, List 3, or distractor).

The five amplitudes (v_1, v_2, v_3, g, u) used to form this superposition are parameters in the QEM model. Only real valued amplitudes are needed to fit the model. The parameter u is always set to be $u = 1$ because the normalization makes this parameter arbitrary. In addition, in the experimental task, two of the recognition probes do not match the test cue, and thus, the amplitudes of the verbatim features for these two probes are equated. For example, if the cue is from List 1, then Lists 2 and 3 probes do not match the cue. Thus, we set $v_2 = v_3$ in this case. Therefore, only three parameters are needed to fit to the observed probabilities from four probe questions, which leave one degree of freedom to test the model for each type of cue in the 3-list experiment described earlier.

Second, based on quantum principles, QEM uses a *projector* to represent each of the four types of recognition probe questions, which have "yes" ("accept") or "no" ("rejection") answers. In the current QEM implementation, the projector is a 5×5 diagonal matrix M with appropriately set zeros or ones on the diagonal (detailed in next paragraph). To answer a probe question is to project the memory state S to the probe M. This projection process is similar to the "memory-cue matching" process theorized in the above mentioned feature vectors in global memory models, but response probabilities are computed differently. As introduced earlier, in quantum probability theory, the probability of an event is the squared length of the projection projecting the state vector to the subspace representing the event. In QEM, the probability of a "yes" answer to a recognition probe is the squared length of the projection $\|M \cdot S\|^2$. The probability of a "yes" answer to the OR probe is computed from 1−p["no" to all the other three probes].

For example, consider when a List 1 test cue is presented. Define S_1 as the superpositional memory state. We fit three parameters (v_1, $v_2 = v_3$, and g) to the four types of probe questions: (a) M_1 is a projector used to respond to the List 1 probe question (i.e., "Presented in List 1"?). Similar to the assumption of OD, both verbatim List 1 features and gist features should lead to the answer "yes." Under this assumption, $M_1 = \text{diag}(1,0,0,1,0)$. The probability of answering "yes" to the List 1 probe is: $P(\text{List 1 response} = \text{yes} \mid \text{List 1 cue}) =$
$\|M_1 \cdot S_1\|^2$. Similarly, (b) M_2 is a projector used to respond to the List 2 probe question. We allow both verbatim List 2 features and gist features to answer "yes." Thus, $M_2 = \text{diag}(0,1,0,1,0)$ and $P(\text{List 2 response} = \text{yes} \mid \text{List 1 cue}) = \|M_2 \cdot S_1\|^2$. (c) M_3 is a projector used to respond to the List 3 probe question. Both verbatim list 3 features and gist features lead to the answer "yes." Thus, $M_3 = \text{diag}(0, 0, 1, 1, 0)$ and $P(\text{List 3 response} = \text{yes} \mid \text{List 1 cue}) = \|M_3 \cdot S_1\|^2$. (d) The probability of "yes" to the OR probe is $P(\text{OR response} = \text{yes} \mid \text{List 1 cue}) = 1 - \|(I - M_3)(I - M_2)(I - M_1) \cdot S_1\|^2$, where I denotes the 5×5 identity matrix.

When a List 2 cue, a List 3 cue, or a distractor is presented, the same process occurs with the following exceptions. For a List 2 cue, we set $v_1 = v_3$, and we estimated a new memory state S_2 with three new parameters. For a List 3 cue, we set $v_1 = v_2$, and we estimated a new memory state S_3 with three new parameters. For a distractor cue, we set $v_1 = v_2 = v_3$, and we estimated a new memory state S_4 with two new parameters.

4.3. An illustrative empirical test

QEM was used to estimate the parameters for the 3-list experimental data described earlier. To be consistent and comparable with the OD model, the QEM model also fits separate parameters to different experimental conditions. Maximum likelihood was used to fit three parameters to each experimental condition (four conditions of concreteness × frequency) for each list cue (List 1, List 2, List 3), and two parameters to each experimental condition for each distractor cue. That is, across all the experimental conditions and test cue combinations, QEM estimates 44 parameters in total. Each combination has four "yes" and four "no" probabilities. We computed a G^2 test statistic as follows. The frequencies for each experimental condition and test cue are based on 70 subjects with two observations per person for the List 1–3 cues and with six observations per person for the distractor cues. Define $n_i(\text{yes})$ and $n_i(\text{no})$ as the proportions of "yes" and "no" responses, respectively, to the probe of i, where i = 1, 2, 3 for List 1, 2, 3, and i = or for the OR probe. Similarly, $p_i(\text{yes}|\text{m})$ and $p_i(\text{no}|\text{m})$ are the probabilities of "yes" and "no" answers, respectively, for List i or the OR probe computed from a model m. The log likelihood for each experimental condition and test cue equals

$$\begin{aligned} \text{Ln}(\text{likelihood (data}|\text{m})) = (2)(70) \cdot [&\ln(p_1(\text{yes}|\text{m}) \cdot n_1(\text{yes})) + \ln(p_2(\text{yes}|\text{m}) \cdot n_2(\text{yes})) \\ &+ \ln(p_3(\text{yes}|\text{m}) \cdot n_3(\text{yes})) + \ln(p_{or}(\text{yes}|\text{m}) \cdot n_{or}(\text{yes})) \\ &+ \ln(p_1(\text{no}|\text{m}) \cdot n_1(\text{no})) + \ln(p_2(\text{no}|\text{m}) \cdot n_2(\text{no})) \\ &+ \ln(p_3(\text{no}|\text{m}) \cdot n_3(\text{no})) + \ln(p_{or}(\text{no}|\text{m}) \cdot n_{or}(\text{no}))]. \end{aligned} \tag{8}$$

Let $G^2_{\text{QM}} = -2\cdot \ln(\text{Likelihood(data|QEM)})$ when $p_i(\text{yes}|\text{m})$ is based on the QEM model, and $G^2_{\text{saturated}} = -2\cdot \ln(\text{Likelihood(data|saturated)})$ when $p_i(\text{yes}|\text{m})$ is based upon the saturated model. In the saturated model, $p_i(\text{yes}|\text{m}) = n_i(\text{yes})$ are observed proportions in the experiments. That is, the saturated model has the best fit possible since it perfectly describes the observed data, and thus it is used as a comparison against the QEM model. Then, a G^2 difference test is defined as follows: $G^2_{\text{diff}} = G^2_{\text{QEM}} - G^2_{\text{saturated}}$, with df = the difference of the numbers of parameters in the saturated model versus QEM = 64−44 = 20.[3] Table 3 summarizes estimated parameters of the model and G^2_{diff} statistics for each of the experimental condition × test cue combinations. The G^2_{diff} tests suggest that this simple quantum episodic model can account for the observed data pattern reasonably well: It fits 13 out of the 16 conditions successfully at the α = .05 level. The total G^2_{diff} statistics across all 16 conditions is 26.75, which is not significant at the α = .05 level (the critical value is 31.41, df = 20), suggesting that across all conditions, the model fits the data well.

5. Concluding comments

The superposition principle that originally stimulated the development of quantum probability theory stipulates that before a measurement is taken, a physical system has

Table 3
The quantum episodic memory model: Estimated parameters and G^2_{diff} tests

	Word Content			
Parameters/G^2_{diff}	High-frequency Concrete	High-frequency Abstract	Low-frequency Concrete	Low-frequency Abstract
List 1 targets/cues				
v_t	0.6477	0.5996	0.6932	0.3860
v_n	0.2009	0.1531	0.2425	0.3574
G	0.8706	0.8838	1.0710	1.2549
G^2_{diff}	0.7661	0.4027	5.2442*	6.0463*
List 2 targets/cues				
v_t	0.4883	0.8106	0.5336	0.6527
v_n	0.4473	0.3570	0.1444	0.3334
G	0.7178	0.9129	0.7861	0.9718
G^2_{diff}	0.0154	1.0056	0.3671	0.1811
List 3 targets/cues				
v_t	0.6080	0.4264	0.6472	0.3433
v_n	0.4725	0.2303	0.2849	0.2837
G	0.8271	0.9129	0.8449	1.0654
G^2_{diff}	0.7108	0.2591	0.0156	0.5963
List 4 distractors				
v_t	0.1591	0.0983	0.0752	0.0770
v_n	0.1591	0.0983	0.0752	0.0770
G	0.4478	0.5670	0.3575	0.4779
G^2_{diff}	2.1029	0.9476	6.3242**	1.7665

Notes. v_t and v_n are the amplitudes of the target/cue list and the distractor (i.e., noise) lists, respectively.
*Significant deviation at the $\alpha = .05$ level (the critical value is 3.84, $df = 1$).
**Significant deviation at the $\alpha = .05$ level (the critical value is 5.99, $df = 2$).

the simultaneous potential to occupy all possible combinations of its states even though the combinations are mutually exclusive, but it can only occupy one of them after a measurement is taken. Thus, the physical states that are observed are emergent properties of the measurement. We have considered a memory analog of this principle, episodic superposition: Before a cue for a presented or unpresented item is administered on a memory test, the item has the simultaneous potential to occupy mutually incompatible episodic states. In most experiments, those states have been target (presented at study), related distractor (not presented but related), target and related distractor, and neither target nor related distractor. In some recent experiments, the states have been context 1, context 2, both contexts, and neither context. In either type of design, superposition of episodic states is revealed by the finding that the probabilities of cues being remembered as belonging to mutually exclusive states are subadditive; that is, $P(\text{T}) + P(\text{R}) > P(\text{T} \cup \text{R})$, where $P(\text{T})$ and $P(\text{R})$ are the respective probabilities of remembering a cue as belonging to each of the mutually exclusive states and $P(\text{T} \cup \text{R})$ is the probability of remembering the cue as belonging to either of these states.

To date, episodic superposition has been explained via a core representational distinction that is used in false memory research—namely, verbatim versus gist traces of experience—and it has been modeled with OD, which implements that distinction with parameters that estimate the respective probabilities that a cue will provoke verbatim and gist retrieval on memory tests. Like support theory (Tversky & Koehler, 1994) and various models of human probability judgment, OD is a neo-classical probability model—"classical" in the sense that its parameter values must fall in the unit interval and "neo" in the sense that probabilities do not obey the additivity axiom of classical probability. Beginning with an initial review of 100 + data sets (Brainerd & Reyna, 2008), OD has been applied to the types of experiments that produce episodic superposition, and it has typically yielded good fits to those data.

Quantum probability theory also appears to provide a natural, intuitive account of the subadditivity phenomenon. A 5-dimensional quantum model, QEM, was derived from the basic principles of quantum probability theory and was found to fit the observed data in 13 of 16 conditions of an illustrative experiment. Compared to OD, it implements alternative process assumptions and uses different mathematical machinery to fit data. On the process side, QEM formalizes the superposition principle as an ambiguous memory state that mixes verbatim information from study lists, gist information, and distractor information. The memory state changes depending on which cue is tested. In other words, the memory state is constructed upon cue presentation. QEM's feature vector representations of memory are consistent with many global matching theories, such as MINERVA 2 (Hintzman, 1986, 1988), SAM (Raaijmakers & Shiffrin, 1981), REM (Shiffrin & Steyvers, 1997), and NEMO (Kahana & Sekuler, 2002). However, these representations are derived from quantum probability theory, which provides a direct formalization of the superposition concept as a fundamental ambiguity in the memory retrieval process. On the mathematical side, global matching models have used various mathematical functions to compute memory-cue matching probabilities, such as memory trace-cue dot product (MINERVA 2) and Bayesian calculation of likelihood (REM). In QEM, as in other quantum models, this probability is computed using the squared length of the projection, which projects the memory state to the probe question.

We applied OD and QEM to a type of experiment that is known to produce consistent episodic superposition results. The global fit test for OD did not produce a null hypothesis rejection, and so OD was statistically acceptable. The model also fit the data of all of the individual conditions. These results were no surprise because OD has delivered acceptable fits to the data of other experiments of this sort (Brainerd et al., 2012). However, the global fit test of QEM also did not produce a null hypothesis rejection. The model fits most of the individual conditions well—13 of 16 in fact—notwithstanding that it is grounded in a different process conception of memory representation. Thus, both OD and QEM performed well and comparably against the usual criterion of a global model fit test. In other words, global fit tests gave comparable results for the two models, though they differed slightly at the level of individual conditions. It is therefore important to consider their relative strengths with respect to other properties of psychological models.

The main strengths of OD are its (a) theoretical specificity, (b) predictive power, and (c) mathematical simplicity. Concerning (a), the process ideas that motivate OD are operationally well defined because they have figured in much prior research on false memory.

Recently, those ideas have also been foci of experimentation in the neuroscience literature (e.g., Dennis, Bowman, & Vandekar, 2012; Reyna, Chapman, Dougherty, & Confrey, 2012). Consequently, when OD posits that episodic superposition is a by-product of a type of remembering in which verbatim retrieval fails but gist retrieval succeeds, we know what that means operationally. Concerning (b), owing to the prior literature on these same ideas, various manipulations have been identified that selectively enhance verbatim or gist retrieval (for a review, see Brainerd & Reyna, 2005). This means that OD can predict particular configurations of manipulations that ought to increase or decrease episodic superposition. Successful tests of such predictions have been reported for both the item and source versions of conjoint recognition (Brainerd et al., 2010, 2012). Concerning (c), the mathematical under-pinning of OD, multinomial modeling, is highly tractable and well understood by researchers (e.g., Batchelder & Riefer, 1999). The statistical machinery for such modeling relies on the theory of maximum likelihood, and off-the-shelf programs are available to handle all steps in experimental application—from initial model coding to data analysis at the level of groups or individuals.

A notable strength of the QEM model is that it is mathematically principled. We derived the model from general principles of quantum probability theory, instead of formulating it specially to account for the episodic subadditivity phenomenon. In addition, QEM is mathematically simple and conceptually intuitive. On the one hand, it only uses a very small part of quantum theory, and the computation involved is simple. On the other hand, in contrast to the common impression that quantum theory is counterintuitive, QEM actually is quite intuitive at a process level. Its formalization of a superposition state of memory captures the fundamental ambiguity and uncertainty in retrieval. Also, as already discussed, the vector representations of verbatim and gist features are consistent with a large group of memory models, which have been shown to explain a large range of recognition and recall data (for a review, see Clark & Gronlund, 1996; Raaijmakers, 2008). In addition, it is worth noting that QEM provides a memory conception for distractor cues, whereas OD uses a bias conception for such cues.

Another strength of QEM is its generalizability. First, its vector representations of memory are more abstract conceptions than the process ideas that motivate OD. As a result, models that rely on such representations can span a broader range of episodic memory data than OD, which was specifically formulated to model conjoint recognition and related designs, such as process dissociation (e.g., Jacoby, 1991) and source monitoring (e.g., Klauer & Kellen, 2010). Thus, although we used QEM to model data from such designs, its potential scope of applications is much broader. Second, quantum probability is a more encompassing variant of probability theory than the variant that is used in OD (see Hughes, 1989). QEM is therefore able to accommodate a more extensive range of empirical violations of classical probability. In fact, similar models derived from quantum probability theory have been used to explain various cognitive phenomena that seem puzzling from a classical probability perspective (e.g., Aerts, 2009; Busemeyer, Pothos, Franco, & Trueblood, 2011; Khrennikov & Haven, 2009; Pothos & Busemeyer, 2009; Atmanspacher & Filk, 2010; Wang & Busemeyer, 2013; also see the introduction to this special issue). This special issue presents many of these applications. In memory research, an example related to conjoint recognition is *super*additivity of response probabilities. In OD, response probabilities are subadditive, approaching additivity as a limit when $V = 1$ or $G = 0$. Superadditivity—that is, $P(\mathrm{T}) + P(\mathrm{R}) < P(\mathrm{T} \cup \mathrm{R})$—is only allowed under certain arrangements of the values of its bias parameters and is not allowed when the influence of bias has been removed from $P(\mathrm{T})$, $P(\mathrm{R})$, and $P(\mathrm{T} \cup \mathrm{R})$. For superadditivity to be allowed when bias has been removed, it would be necessary to relax the restriction that the memory parameters (R, E, and I) must have the same values in all three instructional conditions (see Section 3.1). However, that would require theoretical motivation and would come at the cost of estimating additional memory parameters. That superadditivity might occur in the conjoint recognition parameter is not entirely speculative because there are examples of superadditive response probabilities in the closely related sphere of human probability judgment (Fox, Ratner, & Lieb, 2005). A more complicated implementation of QEM (to include the quantum principle of noncommutative measures) could accommodate various nonadditivity phenomena, including superadditivity.

In addition to the above differences between OD and QEM, it would obviously be desirable if they predicted some contrasting empirical effects that could be used to pit the models against each other in the arena of data. The challenge is to find an assortment of such effects, and further analysis of the models will be necessary to identify them. However, one candidate has been noted: superadditivity. QEM can accommodate superadditivity and OD cannot when response bias is controlled without relaxing an important assumption. The deeper question is whether QEM actually *predicts* superadditivity under certain conditions (e.g., as OD predicts variations in subadditivity under certain conditions). If so, those conditions would provide critical tests of the two models. Evidence of superadditivity would be especially compelling when judged against the backdrop of extant conjoint recognition experiments, which have yet to reveal anything other than subadditive and additive relations.

Acknowledgments

Portions of the research were supported by a grant to the first and third authors from the National Institutes of Health (1RC1AG036915-01) and by grants to the second author from the National Science Foundation (SES 0818277 and SES 1153846). Preparation of this article was supported by fellowships to the first and third authors from the Sage Center for the Study of the Mind. The authors thank Jerome Busemeyer for his valuable discussions through preparing this manuscript. They also thank three reviewers and Wayne Gray for their helpful comments.

Notes

1. The values of $P(\mathrm{T})$, $P(\mathrm{R})$, and $P(\mathrm{T} \cup \mathrm{R})$ for each data set in Fig. 1 were corrected for response bias, using standard signal detection correction methods (Snodgrass &

Corwin, 1988). Brainerd and Reyna (2008) showed that subadditive relations between $P(\mathrm{T}) + P(\mathrm{R})$ and $P(\mathrm{T}\cup\mathrm{R})$ could be caused by differences in levels of response bias for the three testing conditions (b_{T}, b_{R}, and b_{TR}). Specifically, different test instructions (T, R, TR) may produce different levels of bias, and if the arrangement of those differences happens to be $(b_{\mathrm{T}} + b_{\mathrm{R}}) > b_{\mathrm{TR}}$, that could create spurious subadditivity between $P(\mathrm{T}) + P(\mathrm{R})$ and $P(\mathrm{T}\cup\mathrm{R})$ or inflate actual subadditivity.

2. To simplify presentation of the model, the equations do not contain the usual terms for response bias. However, bias terms are always included in data analyses, and the effects of bias are removed from all results. Thus, in considering these equations, simply assume that bias corrections are always present (see Appendix for discussion of response-bias parameters and the expanded version of the equations that contain bias terms).
3. QEM and OD were fit to somewhat different data sets from the experiments. Specifically, OD was fit to a data set with 96 degrees of freedom (free empirical data of probabilities) with separate probability data of the distractor for each target list, while QEM was fit to a slightly simper version of the data set that had 64 degrees of freedom with distractor data averaged across all three target lists.

References

Aerts, D. (2009). Quantum structure in cognition. *Journal of Mathematical Psychology*, *53*, 314–348.

Atmanspacher, H., & Filk, T. (2010). A proposed test of temporal nonlocality in bistable perception. *Journal of Mathematical Psychology*, *54*, 314–321.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*, 57–86.

Brainerd, C. J., Reyna, V. F., Holliday, R. E., & Nakamura, K. (2012). Overdistribution in source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 413–439.

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. doi:10.1016/j.jml.2007.08.006

Brainerd, C. J. & Reyna, V. F., Aydin, C. (2012). Reliability of children's testimony in the era of developmental reversals. *Developmental Review*, *32*, 224–267.

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

Brainerd, C. J., Reyna, V. F., & Aydin, C. (2010). Remembering in contradictory minds: Disjunction fallacies in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36* (3), 711–735.

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

Busemeyer, J. R., & Bruza, P. (2012). *Quantum models of cognition and decision*. Cambridge, UK: Cambridge University Press.

Busemeyer, J. R., Pothos, E. M., Franco, R., & Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological Review*, *118*, 193–218.

Busemeyer, J. R., Wang, Z., & Lambert-Mogiliansky, A. (2009). A quantum probability explanation for decisions that violate the sure thing principle. *Journal of Mathematical Psychology*, *53*, 423–433.

Busemeyer, J. R., Wang, Z., & Townsend, J. T. (2006). Quantum dynamics of human decision-making. *Journal of Mathematical Psychology*, *53*, 220–241.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin and Review*, *3*, 37–60.

Dennis, N. A., Bowman, C. R., & Vandekar, S. N. (2012). True and phantom recollection: An fMRI investigation of similar and distinct neural correlates and connectivity. *Neuroimage*, *59*, 2982–2993.

Feynman, R. P., Leighton, R. B., & Sands, M. (1965). *The Feynman lectures on physics* (Vol. *3*). Reading, MA: Addison-Wesley.

Fox, C. R., Ratner, R. K., & Lieb, D. S. (2005). How subjective grouping of options influences choice and allocation: Diversification bias and the phenomenon of partition dependence. *Journal of Experimental Psychology: General*, *134*, 538–551.

Frost, P. (2000). The quality of false memory over time: Is memory for misinformation "remembered" or "known"? *Psychonomic Bulletin & Review*, *7*, 531–536.

Gerlach, W., & Stern, O. (1922). Das magnetische moment des silberatoms. *Zeitschrift für Physik*, *9*, 353–355.

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.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.

Hughes, R. I. G. (1989). *The structure and interpretation of Quantum mechanics*. Cambridge, MA: Harvard University Press.

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

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, *42*, 2177–2192.

Khrennikov, A. Y., & Haven, E. (2009). Quantum mechanics and violations of the sure thing principle: The use of probability interference and other concepts. *Journal of Mathematical Psychology*, *53*, 378–388.

Klauer, K. C., & Kellen, D. (2010). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review*, *17*, 465–478.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability model explanation for violations of "rational" decision making. *Proceedings of the Royal Society B*, *276* (1665), 2171–2178.

Raaijmakers, J. G. W. (2008). Mathematical models of human memory. In H. L. Roediger III(Ed.), *Cognitive psychology of memory. Vol. 2 of Learning and memory: A comprehensive reference* (pp. 445–466). Oxford, UK: Elsevier.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

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

Reyna, V. F., Chapman, S., Dougherty, M., & Confrey, J. (Eds.) (2012). *The adolescent brain: Learning, reasoning, and decision making*. Washington, DC: American Psychological Association.

Shiffrin, R., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *101*, 547–567.

Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science*, *5*(4).

Wang, Z., & Busemeyer, J. R. (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science*, *5*(4).

Appendix

Consider an experiment in which subjects encode events in two physically distinctive contexts, List 1 and List 2. On memory tests, three types of cues (List 1 targets, List 2 targets, and distractors) are factorially crossed with three types of questions (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 $p_{\mathrm{M,L1}}$ be the probability of accepting a probe that describes such an item as having been presented on List 1, let $p_{\mathrm{E,L1}}$ be the probability of accepting a probe that describes such an item as having been presented on List 2, and let $p_{\mathrm{I,L1}}$ 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 $p_{\mathrm{M},\emptyset}$ be the probability of accepting a probe that describes such an item as having been presented on List 1, let $p_{\mathrm{E},\emptyset}$ be the probability of accepting a probe that describes such an item as having been presented on List 2, and let $p_{\mathrm{I},\emptyset}$ be the probability of accepting a probe that describes such an item as having been presented on List 1 or List 2.

For an L1 target, R_1 is the probability that it provokes recollection of List 1 contextual details, E_2 is the probability that it provokes recollection of List 2 contextual details, and I_1 is the probability that it provokes item memory without recollection of contextual details. For a distractor, b_1 is the probability that it is accepted when a probe asks if it was presented on List 1 and when a probe asks if it was presented on List 2, and $b_{1,1\cup2}$ is the probability that it is accepted when a probe asks if it was presented on either List 1 or List 2. These parameters measure response bias. The expressions for these empirical probabilities are

$$p_{I,\mathrm{L1}} = R_1 + (1-R_1)E_2 + (1-R_1)(1-E_2)I_1 + (1-R_1)(1-E_2)(1-I_1)b_{1,1\cup2}, \quad \text{(A1)}$$

$$p_{\mathrm{E,L1}} = (1-R_1)E_2 + (1-R_1)(1-E_2)I_1 + (1-R_1)(1-E_2)(1-I_1)b_1, \quad \text{(A2)}$$

$$p_{\mathrm{M,L1}} = R_1 + (1-R_1)(1-E_2)I_1 + (1-R_1)(1-E_2)(1-I_1)b_1, \quad \text{(A3)}$$

$$p_{\mathrm{I},\varnothing} = b_{1,1\cup2}, \quad \text{(A4)}$$

$$p_{\mathrm{E},\varnothing} = b_1, \quad \text{(A5)}$$

and

$$p_{\mathrm{M},\varnothing} = b_1. \quad \text{(A6)}$$

In this model, different response-bias parameters are estimated for List 2 than for List 1. Those parameters are denoted b_2 and $b_{2,1\cup2}$.

For application to data, estimates can be obtained for the memory and 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 Eqs. A1–A6 in a multinomial modeling program. For items that are presented on List 2 rather than List 1, a set of expressions that parallel Eqs. A1–A6 can be 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 like the one reported in this article, in which subjects encode events in three contexts, List 1, List 2, and List 3. On a memory test, four types of cues (List 1 targets, List 2 targets, List 3 targets, and distractors) are factorially crossed with four types of questions (Presented on List 1? Presented on List 2? Presented on List 3? Presented on either List 1 or List 2 or List 3?). Thus, there are now 16 distinct probes. For List 1 targets, there are four types of probes with associated empirical probabilities: $p_{\mathrm{I,L1}}$, $p_{\mathrm{E2,L1}}$, $p_{\mathrm{E3,L1}}$, and $p_{\mathrm{M,L1}}$. For distractors, there are also four types of probes with associated empirical probabilities: $p_{\mathrm{I},\varnothing}$, $p_{\mathrm{E2},\varnothing}$, $p_{\mathrm{E3},\varnothing}$, and $p_{\mathrm{M},\varnothing}$. The OD model's expressions for these eight empirical probabilities are as follows:

$$\begin{aligned} p_{I,\mathrm{L1}} =& R_1 + (1-R_1)E_2 + (1-R_1)(1-E_2)E_3 + (1-R_1)(1-E_2)(1-E_3)I_1 + \\ & (1-R_1)(1-E_2)(1-E_3)(1-I_1)b_{1,1\cup2\cup3}, \end{aligned} \tag{A7}$$

$$p_{\mathrm{E2,L1}} = (1-R_1)E_2 + (1-R_1)(1-E_2)(1-E_3)I_1 + (1-R_1)(1-E_2)(1-E_3)(1-I_1)b_1, \tag{A8}$$

$$\begin{aligned} p_{\mathrm{E3,L1}} =& (1-R_1)(1-E_2)E_3 + (1-R_1)(1-E_2)(1-E_3)I_1 + \\ & (1-R_1)(1-E_2)(1-E_3)(1-I_1)b_1, \end{aligned} \tag{A9}$$

$$p_{\mathrm{M,L1}} = R_1 + (1-R_1)(1-E_2)(1-E_3)I_1 + (1-R_1)(1-E_2)(1-E_3)(1-I_1)b_1, \tag{A10}$$

$$p_{\mathrm{I},\varnothing} = b_{1,1\cup2\cup3}, \tag{A11}$$

$$p_{\mathrm{E2},\varnothing} = b_1, \tag{A12}$$

$$p_{\mathrm{E3},\varnothing} = b_1, \tag{A13}$$

and

$$p_{\mathrm{M},\varnothing} = b_1. \tag{A14}$$

In this model, different response-bias parameters are estimated for List 2 than for lists 1 and 3, and different response-bias parameters are estimated for List 3 than for lists 1 and

2. The bias parameters for List 2 are denoted b_2 and $b_{2,1\cup 2\cup 3}$, and the bias parameters for List 3 are denoted b_3 and $b_{3,1\cup 2\cup 3}$.

The only 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 a List 1 target is falsely recollected being presented on List 2. E_3 is the probability that a List 1 target is falsely recollected as being presented on List 3. For application to data, these expressions 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. 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 Eqs. A7–A10 can be written, from which parameter estimates can be obtained and goodness-of-fit tests and parameter significance tests can be conducted.