American Psychological Association

Journal of Experimental Psychology: Learning, Memory, and Cognition

0278-7393/19/$12.00

2019, Vol. 45, No. 6, 1042–1065
http://dx.doi.org/10.1037/xlm0000639

A Hybrid Model of Source Monitoring in Paired-Associates Learning

Francis S. Bellezza and Jennifer K. Elek
Ohio University, Athens

Three source-monitoring models were tested using the data of Bellezza, Elek, and Zhang (2016), who presented word pairs with each word in 1 of 4 locations. Given 1 word as a cue, participants had to remember the other word as well as the 2 corresponding locations. Results included (a) locations of the cue and target words were identified equally well; (b) source identification of unrecalled words was above chance; (c) the correct identification of the cue word was positively correlated with that of the target word; and (d) the location of the cue word was frequently confused with the location of the target. Three multinomial processing-tree models were tested to explain these results: a word-code model, an event-code model, and a hybrid model. The hybrid model was able to fit the data from the 5 experimental conditions of Bellezza et al. data. The model also fit data both from an experiment using four background colors as the source attributes and from a validation experiment manipulating visual-imagery instructions. The parameter values of the hybrid model suggested that source performance using locations was based predominantly on a memory code labeled an *event code* that included a source axis describing the locations of the 2 words, whereas performance using color relied more on *word codes* that associated each word with its color. It appears that different source attributes draw upon different combinations of cognitive processes, but each process occurs within the framework of the hybrid model.

Keywords: source monitoring, multinomial processing trees, paired-associate learning, mediation learning

How source information is remembered is an important question in the study of human memory and has been the focus of intensive research. Researchers have generally assumed that source information is stored in memory on the item level. Tests of source identification have typically been accompanied by recognition tests with source identification assumed to be contingent on successful recognition, as evidenced in reviews of source-monitoring research such as that of Mitchell and Johnson (2009). If an item is not recognized, then its contextual information cannot be remembered. Bellezza et al. (2016) suggested that the study of source monitoring in paired-associates learning can provide information about source memory that cannot be obtained by studying only recognition learning. Episodic memory is sometimes comprised of events consisting of combinations of components, each of which may be relatively unfamiliar and included in an unfamiliar pattern. Or, these components may be similar from day to day but occur in different arrangements. See Tversky and Zacks (2013) for a discussion of how events are perceived. Thus, some memory events are assembled from contiguous but unrelated features of an experience, and source information may have to be later retrieved by an unanticipated test query. For example, the sparse question "How was your vacation?" can result in a variety of types of information being remembered, including information regarding times and places. Consequently, for the successful identification of source information the components of each event must be bound together in memory (Chalfonte & Johnson, 1996).

Bellezza et al. (2016) suggested that the study of source information in paired associates can provide information about how source information is remembered in multicomponent events. To explore this notion, they presented pairs of words in a 2×2 array. Later, participants were presented one of the words and had to recall the other. In addition, the location of each of the two words had to be identified. Three studies were performed comprising a total of five experimental conditions. These conditions were in Study 1 (a) a list of pairs of unrelated words; in Study 2 a list made up of both (b) unrelated words and (c) related words; and in Study 3 a list of (d) pairs of unrelated words and (e) pairs of words and letters.

Source-Monitoring Results Using a Paired-Associates Procedure

In the paired-associates learning paradigm the primary measure of remembering is recall of the target word. The term *identification* is used here to represent the event of source identification for either the cue word or target word. In their five test conditions Bellezza et al. (2016) found four unexpected results. These were (a) the equality of cue and target source identification, (b) the source identification of unrecalled targets, (c) the positive correlation between cue and target source identification, and (d) the

This article was published Online First July 19, 2018.

Francis S. Bellezza and Jennifer K. Elek, Department of Psychology, Ohio University, Athens.

Jennifer K. Elek is now at the National Center for State Courts, Williamsburg, Virginia.

We thank Emily Brandenburg for her assistance in collecting and scoring the data and Ru Zhang for comments on an early version of this article.

Correspondence concerning this article should be addressed to Francis S. Bellezza, Department of Psychology, Ohio University, Athens, OH 45701. E-mail: bellezza@ohio.edu

reversal of cue and target locations. These results are further described below. The statistics summarizing these results are displayed in Table 1.

Equality of Source Identification of Cues and Targets

An unexpected result, occurring in all conditions, was that the proportion of target-word locations identified was approximately the same as the proportion of cue-word locations. This was true even though every cue word in the list was available during test but only about half the target words were available, because only about half were recalled. Across the five conditions shown in Table 1 the median proportion of identified sources was .54 for cue words and .53 for target words. When target words were recalled, the median proportions of cue and target locations identified were .71 and .72, averaged across the five conditions. When the target words were not recalled the median proportions of cue and target locations identified were .38 and .35, respectively.

Source Identification of Unrecalled Target Words

The proportion of target locations identified following no target recall were significantly above chance level, indicating that target word location was often identified even though the target word itself could not be recalled and therefore was not present. Source information describing one item in the pair may become associated to the other item. This possibility has been discussed using a variety of paradigms (Ball, DeWitt, Knight, & Hicks, 2014; Bell, Mieth, & Buchner, 2017; Cook, Marsh, & Hicks, 2006; Malejka & Bröder, 2016; Kurilla & Westerman, 2010; Starns & Hicks, 2008; Starns, Hicks, Brown, & Martin, 2008).

Correlation of Source Identification Performance for Cues and Targets

When target recall occurred, the phi correlation between the cue and target identification performance had a median value of .77 across the five conditions. When recall did not occur, the median phi correlation was approximately .37, which was significantly above the chance level of .11. Hence, performance for cue identification and for target identification were positively correlated even when the target word was not available. This result suggested that information stored in memory to identify target location was somehow associated with the cue word.

Reversal of Locations in Source Identification

The errors made in identifying cue and target locations were related. The cue location was often identified as the target location with the target location identified as the cue location. This frequency of location reversals in the five conditions from Bellezza et al. (2016) are displayed in Table 1. Given that both locations were identified incorrectly, the probability was $1/7 = .14$ that a reversal error would occur by randomly choosing one of these seven possible combinations of the two errors. However, the median proportion of reversed locations across the three studies when words were located incorrectly was approximately .50, given that the target was recalled. When recall did not take place, the mean proportion of reversed locations over the five conditions of the three studies was .19. This value was significantly above the chance value of .14, $t(4) = 4.70$, $p < .005$.

Bellezza et al. (2016) described a number of interesting results. The problem remains to create a model that accounts for data collected from a paired-associates learning procedure combined with a source-monitoring task.

The Present Research

In this research, we tested three models described below as potential explanations for the findings of Bellezza et al., 2016. This was done by testing all five data sets reported by Bellezza et al. The best model was the hybrid model. An additional experiment was performed using color instead of location as a source. Again, it was shown that the hybrid model could explain these data. The results suggested that the success of the hybrid model could be generalized to another source feature. With color it was found that word codes were used more extensively to identify color than event codes, which was not the case when using location. This difference in the two sources is described by performing an

Table 1
Summary of Statistics From the Five Experimental Conditions Reported by Bellezza et al. (2016)

		Study 1	Study 2		Study 3		
Statistic	Item type	Unrelated words	Unrelated words	Related words	Letter cues	Word cues	Median value
Proportion of items recalled		.44	.25	.50	.39	.66	.45
Source identification-overall	Cue	.54	.48	.53	.54	.54	.54
	Target	.51	.45	.53	.53.	.55	.53
Source identification with recall	Cue	.71	.80	.69	.80	.62	.71
	Target	.72	.78	.69	.79	.63	.72
Source identification with no recall	Cue	.39	.38	.34	.40	.38	.38
	Target	.35	.35	.31	.37	.36	.35
Phi correlation-recall		.76	.85	.85	.77	.76	.77
Phi correlation no recall		.24	.30	.44	.37	.37	.37
Proportion reversals—recall		.50	.50	.54	.51	.48	.50
Proportion reversals—no recall		.19	.18	.20	.17	.22	.19

Note. Source identification refers to the proportion of locations correctly identified. Recall = recall of the target word. Phi correlations and reversals = the relation between cue and target source identifications.

identification-component analysis based on the hybrid model. Also, when using color rather than location as an attribute, source-identification performance was greater for cue words than for target words when target recall did not occur.

In addition, we conducted a validation experiment, Experiment 2, to determine if the values of the parameters of the hybrid model behaved in a manner that matched each parameter's interpretation in the model. The hypothesized memory structures of source axes and word codes were also validated as functioning as mediating memory structures. Finally, a procedure is recommended to determine if the source of unrecalled target words are remembered above chance level when the hybrid model is not used to analyze the data.

The Three Source-Monitoring Models

Because source-monitoring is usually studied in conjunction with recognition learning, most models of source-monitoring are an extension of some recognition-learning model (Banks, 2000; Batchelder & Riefer, 1990; Glanzer, Hilford, & Kim, 2004; Rotello, Macmillan, & Reeder, 2004). The procedure used by Bellezza et al. (2016) was a paired-associates source-monitoring paradigm with the two words in each pair presented in different locations, so any model must include a way to associate the two items in the pair. Our goal was to describe, evaluate, and compare three models that try to account for the data of Table 1 to determine which model is the best.

The first model, a word-code model, assumes that source information about an item may become associated to the memory representation of the other item in the pair. The second model is an event-code model based on the concept of an event code as described by Bellezza et al. (2016). This model assumes that a visual image associates the two words in a pair, and, in addition, a visual axis containing source information about the two words becomes associated to this image.

The third model is a hybrid model which is a combination of the event-code model and the word-code model. This model allows for source information to be associated with either an event code or a word code. We conclude that the hybrid model is superior to the other two models and give reasons for this conclusion.

The Word-Code Model

In the word-code model the cue word can retrieve location information about both the cue and the target word. Similarly, if the target word is recalled, it may also aid in identifying the location of the cue word. That is, the two types of location information may be accessible in memory by both words in the pair but probably to different degrees depending on whether recall of the target word takes place.

Description of the model. This model is diagrammed in Figure 1 with associations represented by the solid and the dotted lines. The solid lines represent the associative links between the two words and between each word and its source information. Each dotted line represents an association between the cue word and source information of the target word and an association between the target word and source information of the cue word. The possibility that one word in a pair may retrieve source information from memory about the other word has been discussed by

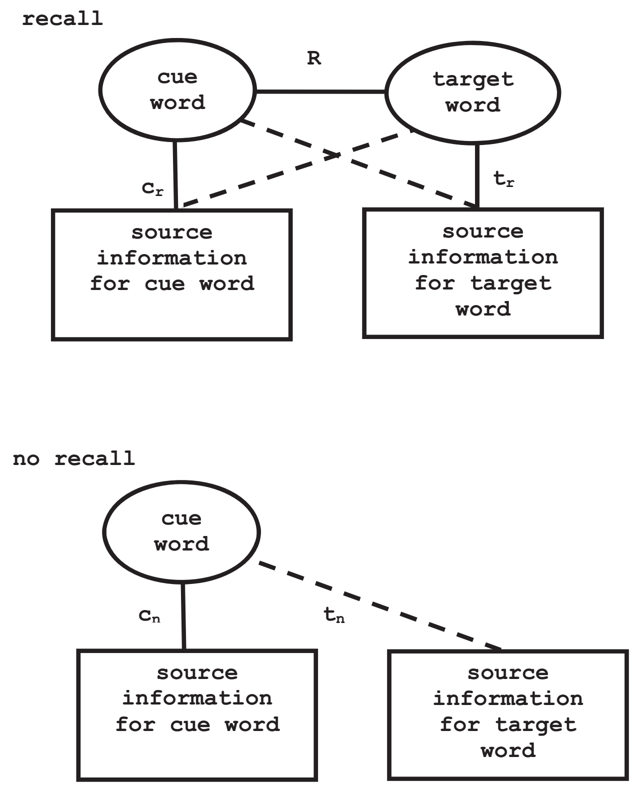

Figure 1. Representations of source location information associated with individual word representations in paired-associates learning. The solid lines represent the association between the two words in the pair and the association of each word with its source. The dotted lines represent the links by which each word in the pair may be associated with source information of the other word in the pair. R = proportion of target words recalled; c_r = proportion of cue word locations retrieved when target word recalled; c_n = proportion of cue word locations retrieved when target word not recalled; t_r = proportion of target word locations retrieved when target word recalled; t_n = proportion of target word locations retrieved when target word not recalled.

a number of researchers (Ball et al., 2014; Bell et al., 2017; Cook et al., 2006; Malejka & Bröder, 2016; Starns et al., 2008). See also Jones (1976) and Ross and Bower (1981).

Parameter R represents the probability that the cue word will retrieve the target word from memory so that it can be recalled. When recall occurs both the cue word and the target word are present, so parameter c_r represents probability that the location of the cue word will be identified by the presence of the cue and target words, and t_r represents the probability that the location of the target word will be identified by that same combination. If the target word is not recalled, the c_n represents the probability that the location of the cue word will be identified by the presence of only the cue word, and parameter t_n represents the probability that the location of the target word will be identified by only the cue word. Therefore, when no target word is recalled, the information for identifying the location of either word is provided only by the cue. The presence of the cue and target words, rather than the cue alone, provides context that is close to the encoding context experienced when the pair was presented for study. Therefore, identification performance for both words should be better when recall takes place (Tulving & Thomson, 1973).

The processing tree for the word-code model is shown in Figure 2. The word-code model and the other models presented here are multinomial processing-tree (MPT) models (Batchelder & Riefer, 1999; Erdfelder et al., 2009; Riefer & Batchelder, 1988). The fractional values shown in the tree represent guessing probabilities. The processes following recall failure of the target word are the same as when recall occurred except that the corresponding parameters involved are c_n and t_n.

Results. All experiments reported in this article were performed using a protocol approved by an institution review board. Many of the descriptive statistics reported by Bellezza et al. (2016) were computed for individual participants whose scores were then averaged. Because aggregated data were used to test the three models described here, the descriptive statistics referred to from Bellezza et al. will, for the most part, also be based on aggregated data shown in the Appendix. The aggregated data were analyzed by means of the maximum-likelihood methods recommended by Riefer and Batchelder (1988) and Hu and Batchelder (1994). Also, see Riefer and Batchelder (1991) for a discussion of the effects of individual differences among participants on parameter values in the model. The likelihood-ratio statistic, G^2, was used as a goodness-of-fit measure. The Type I error rate, alpha, was set to a value of .05. The program multiTree (Moshagen, 2010) was used to fit the word-code model to the aggregated data from the five experimental conditions reported by Bellezza et al. For each condition, frequencies of the five response categories accompanying recall were used along with frequencies from the five response categories accompanying no recall, as shown in the Appendix. Also, the values of the five parameters of the word-code model, parameters R, c_r, t_r, c_n, and t_n, were estimated. For each experimental condition the model produced a G^2 goodness-of-fit statistic with 4 degrees of freedom whose distribution is asymptotically chi-square (Riefer & Batchelder, 1988). The five $G^2(4)$ values were all greater than 263.0 with $ps < .001$. Clearly, the model did not provide a good explanation for frequencies of identifying the cue and targets locations.

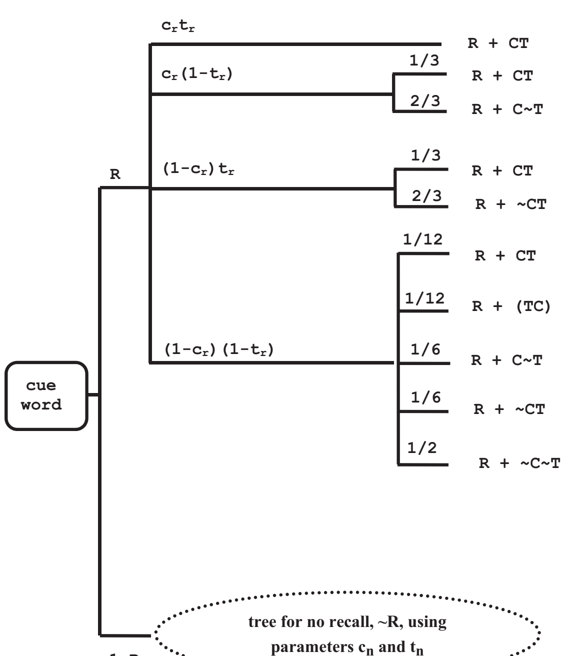

Figure 2. Word code source-monitoring model for paired-associates learning. The symbol R = the event of recall; ~R = the event of no recall; CT represents the correct identification of the location of cue and target words; (TC) = switching the locations of the two words; ~C = an error in location of the cue word; ~T = an error in the location of the target word. R = proportion of target words recalled; c_r = proportion of cue word locations identified when target word recalled; c_n = proportion of cue word locations identified when target word not recalled; t_r = proportion of target word locations identified when target word recalled; t_n = proportion of target word locations identified when target word not recalled.

Some aspects of performance, however, were well accounted for by the word-code model. Recall performance could successfully be accounted for because the value of parameter R measured directly the proportion of target words recalled. Also, when the four findings of Bellezza et al. (2016) were compared with predictions of the word-code model, the word-code model generated the correct proportion of location identifications for both the cue and target words. When target recall occurred, the word-code model identified a median .71 of the cue locations and a median of .75 of the target locations from the five conditions, which were similar to the median values observed and shown in Table 1. With no recall, the two median identification proportions were .35 and .32, the same median values shown in Table 1. Hence, the word-code model could explain the levels of source identification for cues and targets and thereby also show the result that accuracy of identification was the same for cues and targets.

The model did not do as well accounting for the relationship between identifying the location of the cue and identifying the location of the target word. When target recall occurred, the model predicted a median phi value of .07 across the five conditions, and when recall did not occur, the value was .12. The median values shown in Table 1 were .75 and .33, respectively.

Finally, the calculated proportions of errors in which cue and target locations were reversed across the five conditions were .14 with target recall and .14 with recall failure, which are levels that can be accounted for by guessing. The observed median proportions shown in Table 1 were .50 with target recall and .19 without target recall.

Discussion. In the word-code model two source identification parameters were provided for each word in the pair; c_n and c_r for the cue word in the case of no recall and recall, respectively, and similarly t_n and t_r for the target word. Both the parameter values c and t were dependent on whether the target word was recalled. As a result, the word-code model reproduced the proportions of correct location identifications almost perfectly.

Though the model did well, accounting for the level of location identification for cues and targets, it had problems when trying to account for the phi correlations and the proportion of location reversals. In the word-code model the processes shown in Figure 1 the events of successful location identification of cues and targets are stochastically independent, except for the constraint that both items cannot occupy the same location. Identifying the correct location of the cue cannot affect the probability of identifying the correct location of the target. There is no mechanism in the model by which accuracy in identifying the location of the cue of a particular item is related to identifying the location of the target.

The Event-Code Model

Because of the failures of the word-code model, an event-code model was proposed based on the Bellezza et al. (2016) event-code explanation of their results. In this explanation the participant forms an event code for each pair by using a visual image to associate the two words based on the experimenter's instructions (Bellezza, 1981; Paivio, 1969, 1971; Richardson, 1998), and this mnemonic image represents the co-occurrence event of the two words. This model is an alternative to the word-code model in which a word code is created separately for each word in the pair.

This explanation further assumed that this event code was represented in memory as a hierarchical network made up of propositions using labeled associations, even though this information was experienced as a visual image. According to Anderson and Bower (1974), a propositional representation linking the words of the pair provides a node to which context and source information can be associated. Each propositional tree is divided into two subtrees, a fact subtree and a context subtree, with the fact subtree retrieved first. From the fact subtree comes subject and predicate enabling target recall, and from the context subtree comes the nodes and links for place and time of occurrence. Thus, when a pair of words is presented, it and its associated source information are encoded into a memory network as an event. Later, when the target word is to be retrieved from memory by the cue word, the event code is first retrieved by the cue word and an attempt is made to extract the target word from its associated network information. This is followed by retrieval of context information. Thus, the relations among the information in the event code are encoded using propositions that may include features or stimulus attributes but only as explicit components of those propositions (Anderson, 1983; Balota, 1983).

Description of the model. To create a dependency between the two pieces of location information, we further assumed that the locations of the two items in the pair were perceived and represented in memory as two points on a linear axis created in memory that reflected the positions of the two pair items located in the 2 × 2 visual array. This is referred to as the source axis. The six possible axes are shown in Figure 3. This spatial axis was assumed to become associated to the interactive image that was formed from the two words that enabled target recall to take place. Consequently, the event code was made up of the interactive image supporting target recall and the source axis supporting location identification as part of the presentation context. These components are positively related in the event code, but each might be retrieved separately.

The processing tree for the event code model is shown in Figure 4. There is a probability R that the event code is to be retrieved at test and the correct response word is extracted from the image formed for the pair. Next, there is a probability A_r that the axis containing information relevant to spatial location of the two words will also be retrieved from the event code given that target recall has occurred. If the source axis is retrieved, then the orientation of the words on the axis must be retrieved with a probability O_r. Note that if the orientation of the words on the axis is not retrieved (with probability $1 - O_r$) then the orientation is guessed,

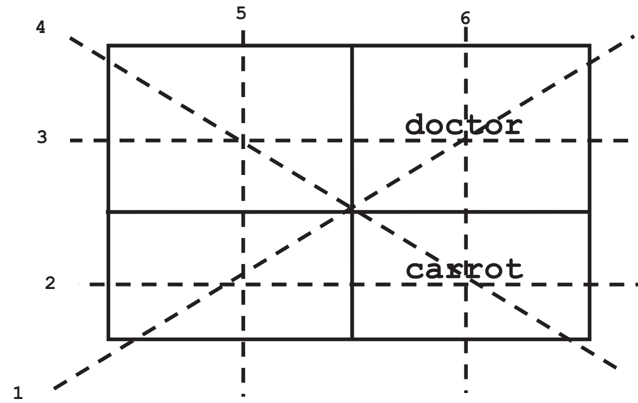

Figure 3. The six axes describing the locations of the two words in the pair.

with the correct locations being identified with probability .50. Thus, it follows that the locations of the two words would be reversed with probability .50 when orientation is guessed.

If recall does not occur, with probability $1 - R$, a source axis of the locations of the two items may, nevertheless, be retrieved from memory. Evidence for this derives from two results: (a) Identification measures for the cue and target items were positively correlated above chance level even when the target words were not recalled, as shown in Table 1; and (b) even without target recall, reversal of the cue and target locations occurred in memory above the expected level. Hence, we assumed that there exists a probability, A_n, that the axis for the pair can be successfully retrieved even when target recall does not occur. This is based on the assumption that the source axis may be retrieved from the event code even if the image of the event code does not support target recall, as suggested in Bellezza et al. (2016, Figure 3). Then, similar to parameter O_r when recall occurs, there is a probability, O_n, that orientation information can then be retrieved and that the locations of the two words will be remembered. The values of A_n and O_n were expected to be smaller than the values of A_r and O_r. The presence of the recalled target word acts as an additional retrieval prompt for the source axis (Tulving & Thomson, 1973). The fractional values shown in the processing tree represent guessing probabilities.

Results and discussion. A model made up of these five parameters, R, A_r, O_r, A_n, and O_n, did better than the word-code model in fitting the data from the five conditions. The $G^2(4)$ values for the five data sets were 31.19, 11.47, 3.31, 17.85, and 4.30. For two of the conditions, the data for the pairs of related words in Study 2 and the data for the letter–word pairs of Study 3, the model fit the data with $p > .05$. For the other three conditions the event-code model did not fit the data. These goodness-of-fit values represent a considerable improvement over the word-code model.

The Hybrid Model

We modified the event-code model in the following way: When the source axis for the pair cannot be retrieved from memory, then an attempt is made to retrieve the location information from the two representations of the individual words. Hence, word codes

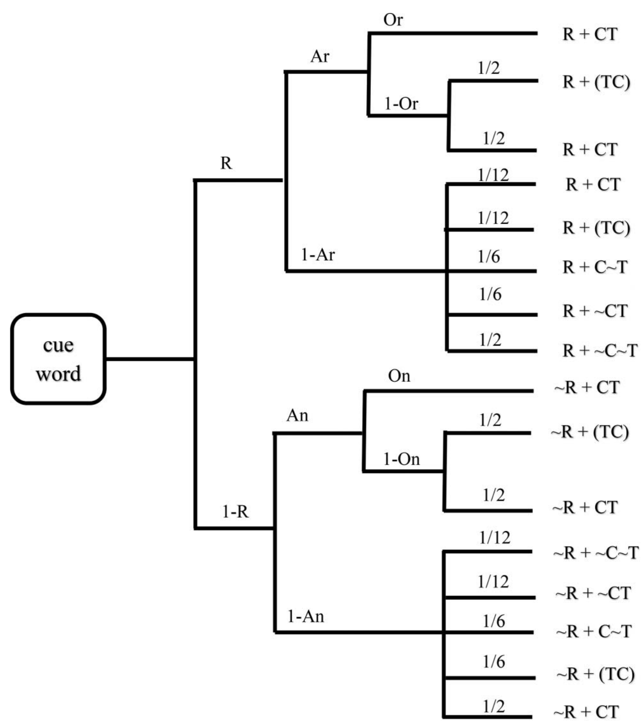

Figure 4. Processing tree for the event-code model. The symbol R = recall; CT = identifying location of cue and target words; (TC) = switching the locations of the two words; ~C = an error in location of cue word; ~T = an error in the location of the target word. The frequency is (TC) has been removed from the frequency for ~T~C. R = proportion of target words recalled; A_r = the proportion of source axes retrieved with target recall; A_n = the proportion of source axes retrieved with no target recall; O_r = the proportion of retrieved source axes correctly oriented when target recalled; O_n = the proportion of retrieved source axes correctly oriented when target not recalled.

can exist in the hybrid model as well as an event code for a pair but are used only if the source axis for a pair cannot be retrieved.

The rationale for this was as follows: The words used in the Bellezza et al. (2016) studies were typically words representing concrete objects. This facilitated the formation of composite visual images combining images from the individual words. Nevertheless, there may be pairs for which a participant might have been unable to create an interactive image easily. In that case the locations of individual words may have been separately attended to during study time of the pair. Thus, location information could become associated only to representations of the individual words, as suggested in the word-code model.

Description of the model. The processing tree for the hybrid model is shown in Figure 5. As in the event-code model, when the target word is recalled, the axis is retrieved with probability A_r. Then there is a probability O_r that the orientation of the words on the axis will be remembered. If the orientation cannot be remembered, then the participant guesses the orientation. If the source axis for the pair cannot be retrieved from memory, which occurs with probability $1 - A_r$, then the location information for the cue word and target word are retrieved from memory with probabilities c_r and t_r, respectively, based on a retrieval process using the two words as cues for the word codes.

It may be that recall of the target word does not occur. An event-code source axis may nevertheless exist in memory as part of the event code. If the axis is retrieved with probability A_n, then there is a probability O_n that the orientation of the words on the axis will be remembered. As in the case of target recall, if the axis formed for the pair cannot be retrieved from memory, then the location information for the cue word and target word are retrieved

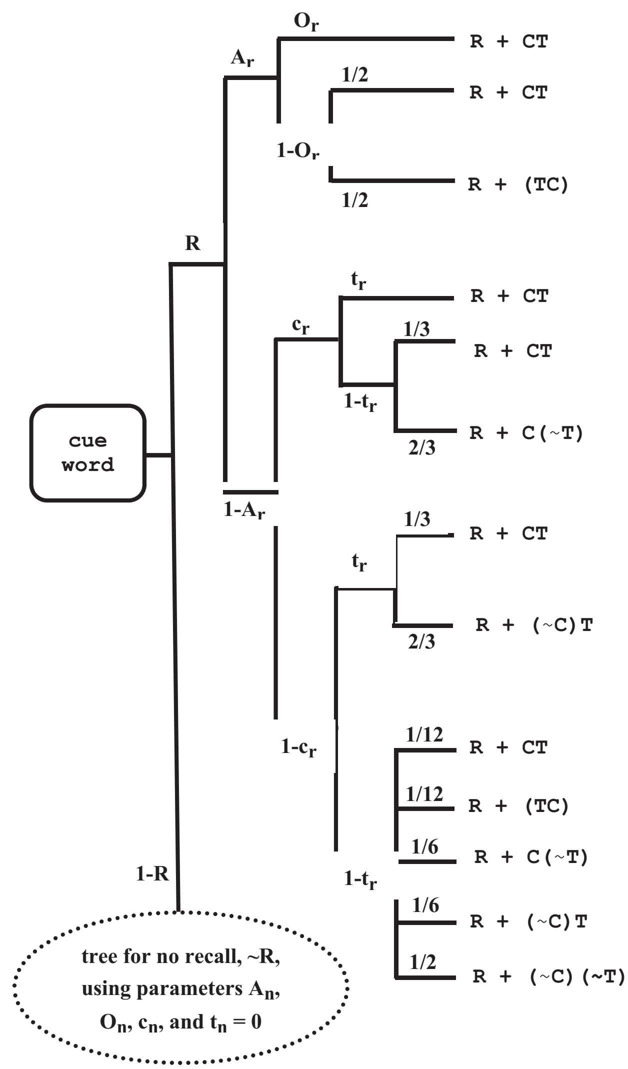

Figure 5. Processing tree for the hybrid source-monitoring model. The symbol R = recall; CT = identifying the location of the cue and target words; (TC) = switching the locations of the two words; (~C) = an error in location of cue word; (~T) = an error in the location of the target word. The frequency is (TC) has been removed from the frequency for ~T~C. R = proportion of target words recalled; A_r = the proportion of source axes retrieved with target recall; A_n = the proportion of source axes retrieved with no target recall; O_r = the proportion of retrieved source axes correctly oriented when target recalled; O_n = the proportion of retrieved source axes correctly oriented when target not recalled; c_r = proportion of cue word locations identified when target word recalled; c_n = proportion of cue word locations identified when target word not recalled; t_r = proportion of target word locations identified when target word recalled.

from memory with probabilities c_n and t_n, respectively. The fractions shown in the tree in Figure 5 represent the probabilities of guessing the correct locations when all types of retrieval fail. The lower part of the processing tree, which represents processes following recall failure of the target word, is not shown in Figure 5, but the processes are the same except that the corresponding parameters involved are A_n, O_n, c_n, and t_n.

Results. The model has nine parameters, but there were only 9 *df* available in each data set from the five conditions of Bellezza et al. (2016). Therefore, the parameters could be estimated but with no degrees of freedom remaining for a goodness-of-fit test (Riefer & Batchelder, 1988). When the parameters of this saturated model were estimated using multiTree (Moshagen, 2010), however, each of the five values of parameter t_n and a number of other parameters were close to a value of zero.

Bootstrapping. Because many parameter values were at the boundaries of the parameter space of (0,1), the assumption of an asymptotic χ^2 distribution for the goodness-of-fit statistic G^2 when the null hypothesis of a model fit is tested may not be met (Bell et al., 2017; Efron & Tibshirani, 1993; Klauer & Oberauer, 1995). Therefore, a bootstrap analysis (Moshagen, 2010) was performed. This analysis gave much the same results as the maximum likelihood tests for the overall fit of the hybrid model. In both the maximum likelihood and in the bootstrap analyses, the value of the parameter t_n in each of the five sets of data was not significantly different from zero.

Setting t_n to zero. The value of $t_n = 0$ indicates that if the source axis was not retrieved following recall failure, then the probability of identifying the location of the target word above chance level when using only the word code of the cue word was zero. This is shown in Figure 1. With the result of $t_n = 0$ there is no line between the cue word alone and source information for the target. Any memory of the location of the target word had to be mediated by the source axis. Therefore, in subsequent analyses of the hybrid model parameter t_n was assumed to be zero. This constraint allowed a goodness-of-fit test to be performed for the hybrid model using the one available degree of freedom within each data set. The mean values for all parameters and their standard deviations based on a parametric bootstrap procedure (Klauer & Oberauer, 1995; Moshagen, 2010) are shown in Table 2. The values of the $G^2(1)$ goodness-of-fit statistics for the five conditions and their *p* values based on the bootstrap analysis are also displayed. The hybrid model fit each of the five sets of data. The observed frequencies as well as the corresponding frequencies generated by the hybrid model are presented in the Appendix. Of the 15 values of the parameters c_r, t_r, and c_n from the five conditions describing the effects of the word codes, nine of them were not significantly different from zero. Nevertheless, the word-code component of the hybrid model was necessary for the model to fit the data, as suggested by the failure of the event-code model described above. The amount of variance accounted for by the parameters c_r, t_r, and c_n in these studies was relatively small.

Discussion. As shown in Table 2, the source axis of the pair was more often successfully retrieved when the target word was recalled, with parameter A_r ranging from .68 to .80, than when the target word was not recalled, where A_n ranged from .10 to .17.

Table 2
Goodness-of-Fit Statistics and Parameter Values Obtained for the Hybrid Model Using Data From the Three Studies of Bellezza et al. (2016)

	Study 1	Study 2		Study 3	
Parameters	Unrelated pairs	Unrelated pairs	Related pairs	Letter cues	Word cues
R	.44 (011)	.25 (011)	.50 (013)	.39 (011)	.66 (011)
Source parameters for event-code representations					
Ar	.75 (018)	.79 (026)	.80 (016)	.79 (020)	.68 (015)
An	.10 (017)	.16 (017)	.11 (019)	.15 (016)	.17 (023)
Or	.78 (027)	.84 (039)	.75 (029)	.86 (023)	.68 (027)
On	.57 (166)	.71 (114)	$.35_0$ (174)	.75 (107)	.46 (128)
Source parameters for word-code representations					
cr	$.08_0$ (052)	.26 (090)	$.03_0$ (036)	.22 (070)	$.02_0$ (023)
cn	.11 (022)	$.03_0$ (020)	$.03_0$ (025)	$.02_0$ (016)	$.03_0$ (025)
tr	.12 (052)	$.06_0$ (070)	$.06_0$ (050)	.18 (071‘)	.07 (034)
tn	.00	.00	.00	.00	.00
Goodness-of-fit statistics					
df	1	1	1	1	1
G^2	2.15	.56	1.35	2.35	.00
p	.15	.51	.46	.16	.98

Note. R = proportion of target words recalled; A_r = the proportion of source axes retrieved with target recall; A_n = the proportion of source axes retrieved with no target recall; O_r = the proportion of retrieved source axes correctly oriented when target recalled; O_n = the proportion of retrieved source axes correctly oriented when target not recalled; c_r = proportion of cue word locations identified when target word recalled; c_n = proportion of cue word locations identified when target word not recalled; t_r = proportion of target word locations identified when target word recalled; t_n = proportion of target word locations identified when target word not recalled. The number in parentheses, when multiplied by .001, gives the value of the standard deviation of the estimate of each parameter. The values with subscripts of zero are not significantly different from zero. All statistics were estimated using bootstrapping procedures with N = 1,000 available in the computer program multiTree (Moshagen, 2010).

Parameter A_n, however, was always significantly greater than zero. The parameter O_r had a median value of .78 and was significantly greater than zero. The median value of the parameter O_n was .57 but the variance of parameter O_n tended to be large. Correctly recalling the target word was positively related to the correct identification of the location of both words in the pair.

Elimination of parameter t_n. Setting t_n to zero is compatible with the finding that source identification is dependent on item recognition and is relevant to the current theorizing about source memory's dependence on item memory (Bell et al., 2017; Malejka & Bröder, 2016). As suggested by Bellezza et al. (2016), when an interacting visual image is formed, a second representation of the cue word is stored in memory. This second representation of the cue word is transformed in meaning compared to the first meaning stored in memory when studied alone. If this transformed cue is not available in memory, but only the original encoding of the cue word is available, then the target word cannot be recalled and identified.

This result also leads to the question of whether unrecalled targets could be recognized if a recognition test were added to our paradigm (Humphreys & Bowyer, 1981). There seems to be a number of ways to test recognition. The target word targets and foils could be presented with or without their corresponding cue words, but, in general, we would expect source identification to be better for the category of recognized targets than for the category of unrecognized targets.

The Complexity of Models

For a model to be accepted as a useful description of a set of data it must fit that data (Klauer & Wegener, 1998; Myung, 2000; Wagenmakers & Farrell, 2004). That is, the model must meet some criterion of goodness of fit. For a MPT model, an adequate fit is defined as a nonsignificant G^2 statistic associated with the maximum likelihood estimate of the parameter values of the model. This test of inference allows for the acceptance of the null hypothesis which implies that there is no significant difference detected between the frequencies comprising the observed data and the frequencies generated from the estimated values of the model's parameters. If the value of G^2 is statistically significant, then interpreting the model parameters would provide an explanation of the frequency data generated by the model when that data has been found to be significantly different from the data observed.

It is necessary for a model to pass the goodness-of-fit test, but it is not sufficient. There is also the problem of complexity. Models may be made to fit by increasing the number of parameters used in the model. Although increasing the number of parameters may help a model fit a particular set of data, the generalizability of the model decreases because a good fit can be the result of capitalizing on error variance (Myung, 2000; Vandekerckhove, Matzke, & Wagenmakers, 2015). Various measures can be used to compare and choose among models differing in complexity, but, importantly, these alternative models must have already been shown to fit the data. So far, we have only one model, the hybrid model, which requires eight parameters but fits the data collected using the paired-associates source-monitoring procedure. When another model is found that fits the data with fewer parameters, a variety of measures of complexity may suggest that the simpler model is to be preferred.

Experiment 1: Color as a Source Attribute

The hybrid model can explain the data from Bellezza et al. (2016), but the question arises of how adequately can the hybrid model fit data when some other source feature is used, such as color. If the hybrid model can successfully account for the data collected using color, then the model is of greater interest and value than if useful only for location. Furthermore, if the model fits color data, then the pattern of parameter values can be compared to the pattern of values obtained using location. Differences in the two sets of parameter values might suggest which memory processes are most important in identifying each attribute.

The source attribute used in the studies discussed so far was the location of each word in a two-dimensional visual field. Color as a source seems to be a more qualitative attribute than location and seems to provide less information that do the horizontal and vertical coordinates of an object in the visual field. The question asked here is whether the hybrid model can fit source data based on remembering the color background of each word.

Method

The 48 pairs made up of concrete nouns were used that were taken from Study 1 of Bellezza et al. (2016). The pairs were recoded such that Locations 1 to 4 of the words in the original 2 × 2 array became the colors red, yellow, green, and blue, respectively. After each word from the pair was printed on its corresponding rectangular color background, the words were randomly assigned the positions left and right, and the two words were presented for study horizontally in a row with their background colors touching, as displayed in Figure 6.

Thirty-two undergraduate students participated for extra course credit. The procedure used was similar to Bellezza et al. (2016). Participants were informed that each pair would be later tested for word recall and color background. The participants were instructed to try to form a mental picture for each presented pair such that the referents of the two words interacted in some manner. They then received a test booklet with a list of cue words, with each cue word selected randomly from the two words in each presented pair and with the 48 cues in random order. The cue word was followed by a blank in which the target word was to be printed. Next to the cue word and next to the blank for the target word the four colors were listed. After recalling the target word, the appropriate background color had to be circled for each word. Every test item had to be completed fully, even if the target word and the word colors were guessed. No participant selected the same color for the cue and target word of any item.

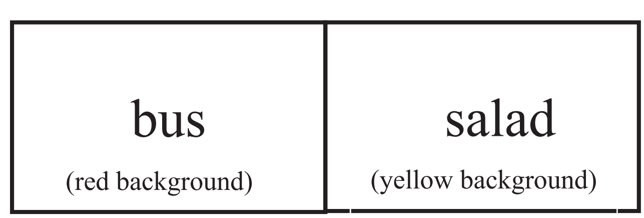

Figure 6. Example of words in a pair presented on their color backgrounds.

Results and Discussion

Before trying to fit the data using the hybrid model, some descriptive statistics using the color attribute were compared to descriptive statistics from Study 1 of Bellezza et al. (2016). These are shown in Table 1 and used location as an attribute. Performance for color identification is displayed in Figure 7. Target recall occurred for .29 of the items using color compared to .44 for recall for when using location as a source. Also, source identification was lower when color was used. Using color as an attribute, we found .46 correct identification of cue words and .33 of target words versus the .54 for cues and .51 found for targets when using location. Surprisingly, the values of .46 and .33 observed for color were significantly different, as shown in Figure 7. That is, when recall occurred, there was no difference in cue and target color identification, but when recall did not occur, identification of the cue word was superior to that of the target word. In all the studies discussed so far using location as a source, identification performance for cues and target words has been only slightly different in value.

Fitting the hybrid model using color as a source attribute. The multiTree program (Moshagen, 2010) was used to fit the hybrid model to the aggregated frequency data shown in Table 3 with color as a source attribute. A bootstrap procedure with N = 1,000 was used. The model fit the color data with $G^2(1) = .75$, $p >$ .50, with the frequencies generated by the model also shown in Table 3. Thus, the hybrid model fit both the location data from Bellezza et al. (2016) and fit the color data reported here. Two nested models (Moshagen, 2010), the word-code model and the event-code model, could be tested separately to determine the extent to which word-codes and event-codes were mediating source-identification performance. Using only the event-code model did not result in an adequate fit, with $\Delta G^2(4) = 121.77$, $p <$ .001. Similarly using only the word-code model did not result in an adequate fit, $\Delta G^2(5) = 53.94$, $p < .001$. Thus, neither of the two models nested in the hybrid model could not fit the data when used by itself.

Comparing the parameter values for location and color. The pattern of parameter values found for the hybrid model were different when analyzing the two different types of source data. Because identical word pairs were used in each experiment, any difference in a parameter value when comparing the attributes of location versus color represents the difference in the frequency

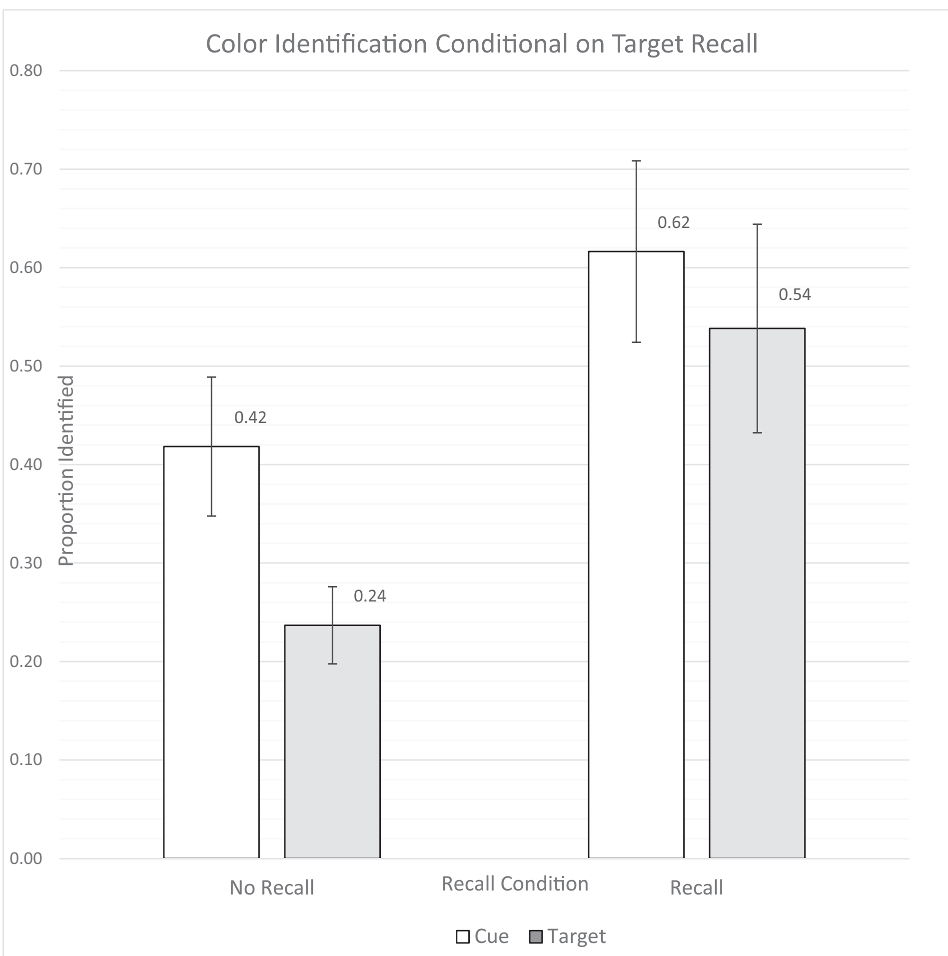

Figure 7. Proportion of cue and target colors identified for the unrecalled and recalled pairs in Experiment 1 averaged across participants. Error bars represent 95% confidence intervals.

Table 3

Observed and Model-Generated Frequencies of Each Response Category for Pairs Using Color as the Source Attribute in Experiment 1 and Using Location in the One-Image and Two-Image Conditions of Experiment 2

Condition	Response category: Cue and response color correct	Cue and response color reversed	Only cue color correct	Only response color correct	Neither cue nor response color correct	Total
Experiment 1: Color attribute						
Correct recall	162 (162.0)	34 (34.0)	71 (71.0)	55 (55.0)	116 (116.0)	—
Incorrect recall	165 (165.0)	73 (73.0)	301 (301.0)	131 (139.8)	428 (419.2)	1536
Experiment 2: Single image condition						
Correct recall	164 (164.0)	32 (32.0)	35 (35.0)	50 (50.0)	71 (71.0)	—
Incorrect recall	209 (209.0)	72 (72.0)	216 (216.0)	147 (147.0)	444 (444.0)	1536
Experiment 2: Two images condition						
Correct recall	489 (489.0)	93 (93.0)	68 (68.0)	116 (116.0)	194 (194.0)	—
Incorrect recall	65 (65.0)	37 (37.0)	82 (82.0)	68 (74.0)	228 (222.0)	1536

Note. The numbers in parentheses are the frequencies produced by the hybrid model. The two conditions of Experiment 1 were fit simultaneously. The category "Neither cue nor response color correct" does not include the category "Cue and response color reversed."

which the cognitive process corresponding to that parameter was being used for the two different sources. Therefore, both sets of data, the source as location data from Study 1 of Bellezza et al. (2016) and the color data reported here, were analyzed simultaneously assigning a separate set of parameter variables to each type of data. As expected (Batchelder & Riefer, 1990; Erdfelder et al., 2009; Riefer & Batchelder, 1988), the resulting goodness-of-fit statistic was the sum of the statistics found when analyzing each data set separately, $G^2(2) = 2.90$, $p > .25$.

Shown in Table 4 are the separate parameter values of the hybrid model for location and color. Also shown are probability levels of z tests used (a) to determine if the value of each parameter was significantly different from zero and (b) to determine if corresponding parameter values in the two data sets were significantly different from one another. The z tests used means and standard errors estimated using a parametric bootstrapping procedure where $N = 1{,}000$ (Moshagen, 2010).

One caveat: The two sets of data compared were collected a number of years apart. Any statistically significant differences found between the two sets of parameter values must be interpreted with some caution. We cannot assume that every difference found in the two sets of parameter values was caused by the differences in the source attributes used. Differences may result from other variables such as the change in the population of college participants used at different times, the change of experimenters, and the fact that in Study 1 of Bellezza et al. (2016) each participant was presented the pairs on an individual computer screen, whereas in the study reported here pairs were projected to the group on one large screen. In addition, there may have been other, more subtle, differences in the circumstances of the two

Table 4

Parameter Values of the Hybrid Model Comparing Location as Source Attribute Using Data of Bellezza et al. (2016, Study 1) and Color as Source Attribute From Experiment 1

Location as source				Color as source				Compare the two strategies
		Test against zero				Test against zero		Test of difference
Parameter label	Parameter value	*SD*	*z*-test *p* value	Parameter label	Parameter value	*SD*	*z*-test *p* value	*z*-test *p* value
R_L	.44	011	***	R_C	.29	011	***	***
Ar_L	.75	018	***	Ar_C	.29	035	***	***
An_L	.10	017	***	An_C	.02	013	n.s.	***
Or_L	.78	027	***	Or_C	.77	089	***	n.s.
On_L	.57	161	***	On_C	.58	413	n.s.	n.s.
cr_L	.08	051	n.s.	cr_C	.17	.047	***	n.s.
tr_L	.12	055	*	tr_C	.10	048	*	n.s.
cn_L	.11	024	***	cn_C	.22	022	***	***

Note. R = proportion of target words recalled; A_r = the proportion of source axes retrieved with target recall; A_n = the proportion of source axes retrieved with no target recall; O_r = the proportion of retrieved source axes correctly oriented when target recalled; O_n = the proportion of retrieved source axes correctly oriented when target not recalled; c_r = proportion of cue word locations identified when target word recalled; c_n = proportion of cue word locations identified when target word not recalled; t_r = proportion of target word locations identified when target word recalled.
[*] z-test significant at $p < .05$. [***] $p < .001$. Standard deviation values should be multiplied by .001. All z tests used means and standard deviations estimated using a parametric bootstrapping procedure where $N = 1{,}000$ (Moshagen, 2010).

experiments (Keppel, 1991). Also, as shown in Table 1, there was variation in the size of the standard errors for the different parameters. The comparisons of parameter values in the two conditions that we report, however, test a number of hypotheses suggesting how cognitive processes might differ when location versus color were the source attributes used. Some tentative conclusions are reported.

Event-code parameters of the hybrid model. Inspection of the event-code parameter values in Table 4 shows that the parameter R_L = .44 was significantly greater than R_C = .29, indicating that recall was better when using location compared to color.

Following recall. Given that target recall occurred, parameter Ar_L represents the probability of retrieving the axis formed for a location pair. Its value, .75, was significantly greater than zero and also significantly greater than Ar_C = .29, which was the probability of retrieving the axis formed for a color pair given that recall occurred. Nevertheless, the value of Ar_C was significantly greater than zero.

The probabilities of remembering the orientation of the two words on the axes were Or_L = .78 and Or_C = .77, both of which were significantly different from zero but not significantly different from one another. We conclude that the source axis played a greater role in source identification when location was used than when color was used. The probability of retrieving the source axis for a pair was larger when location was the source attribute.

Following no recall. When using location as a source attribute, an axis can be retrieved even when the target word was not recalled. This was evidenced in the location data used here by the result that both An_L and On_L were significantly greater than zero. We can see from Table 4 that when using color, neither An_C nor On_C were significantly greater than zero. This means that when the target word was not recalled, there was no possibility of color being identified using the source axis of the event code. And because t_n was assumed to be equal to zero, the color of the target word could not be identified using the cue word when the target word was not recalled.

When target recall occurs using color as the source attribute, there is some way to store the color information for the two words within a memory framework. The two colors are structured not using what we have labeled a source axis, but, perhaps, some nonlinear structure. This means color encoding involves more than simply associating each color to its corresponding word in the pair. The pair of colors must somehow provide an organizing structure or relationship that can be combined with the visual image associating the two words. That is, when using color as an attribute, an event code may contain a color arrangement as indicated by the statistically significant parameter values, Ar_C and Or_C, in addition to the interacting image associating the words in the pair. We discuss below that the term *axis* may not be appropriate for describing the organization of two colors indexed by the model parameter types *A* and *O*.

Word-code parameters for location and color. In the word-code component of the hybrid model, source information about each word in the pair is assumed to be associated with each word's representation in memory. As shown in Table 4, when location was the source attribute and the target word was recalled, cr_L = .08 and tr_L = .12 with only the latter significantly greater than zero. With no target recall, cn_L = .11 was significantly greater than zero. With color as a source attribute the word parameters tended to be larger in value than with location; cr_C = .17, tr_C = .12 and cn_C = .22, with all significantly greater than zero. However, only cn_C was significantly greater than cn_L. The word codes associated with color as a source tended to be larger in value than the word codes formed using location as a source.

In summary, the parameters related to the event code, including *R*, were generally larger when location was the source attribute, whereas the word-code parameters tended to be larger when color was the source attribute. This inverse relation between the values of event code parameters and word code parameters is not surprising because the hybrid model identifies the sources of a pair by first trying to retrieve a source axis from an event code. If the source axis fails to be retrieved, as was often the case with color, an attempt is made to retrieve the word codes.

Computing the Identification Components of the Hybrid Model

As shown in Table 1, Bellezza et al. (2016) found that cue and target identification were similar in value, following both recall and no recall, when using location as a source attribute. When color was used as the source attribute, the results were different. Identification performance was similar in value when recall occurred, but cue identification was greater than target identification when recall failed, as displayed in Figure 7. It would be useful to know precisely which cognitive processes were responsible for the interaction graphed in Figure 7.

There are four components of source identifications for cues and targets that play a role in the hybrid model. The equations for the component proportions based on the model parameters are shown in Table 5. The four components are (a) the cue and target source correctly identified after retrieving the source axis from memory and also remembering the correct orientation, abbreviated in Table 5 as SAM; (b) the cue and target source correctly identified after retrieving the source axis without remembering the correct orientation but correctly guessing it, abbreviated SAG; (c) the cue or target source correctly identified by retrieving the relevant word code from memory, denoted WCM; and (d) the cue or target source correctly guessed after failing to retrieve the relevant word code from memory, denoted WCG. The parametric expressions for the 16 proportions generated by the components shown in Table 5 were derived from the processing tree for the hybrid model shown in Figure 5.

Identifying color following target recall and no recall. The values of the proportions for the four types of identification components for color are shown in the fourth column of Table 5, labeled P(Crt). Each represents the proportion of the 1536 items tested in Experiment 1 that occurred in each of the 16 possible conditions listed. The first four values indicate the proportions of cues and targets that were identified using components SAM and SAG when recall occurred. The second set of four values represents the proportions that were identified using WCM and WCG. For example, when recall occurs the proportions of cues correctly identified was .06 + .01 + .03 + .04 = .14. The corresponding value for the target is also .14. When recall does not occur then the values for cue and target are .31 and .20, respectively.

Table 5

Identification Components of the Hybrid Model Expressing the Proportions of Correct Cue and Target Identifications Mediated by Source Axes and Word Codes With Values From Data Using Color as a Source Attribute

Identification component	Performance	Model equation	P(Crt)	P(Crt|R)
		Source identification based on source axis with target recall		
SAM	Cue remember	R × Ar × Or	.06	.22
	Target remember	R × Ar × Or	.06	.22
SAG	Cue guess	R × Ar × (1-Or) × (1/2)	.01	.03
	Target guess	R × Ar × (1-Or) × (1/2)	.01	.03
		Source identification based on word code with target recall		
WCM	Cue remember	R × (1-Ar) × cr	.03	.12
	Target remember	R × (1-Ar) × tr	.02	.07
WCG	Cue guess	R × (1-Ar) × (1-cr) × [(1/12) × tr + 1/4]	.04	.15
	Target guess	R × (1-Ar) × (1-tr) × [(1/12) × cr + 1/4]	.05	.17
		Source identification based on source axis with no target recall		
SAM	Cue remember	(1-R) × An × On	.01	.01
	Target remember	(1-R) × An × On	.01	.01
SAG	Cue guess	(1-R) × An × (1-On) × (1/2)	.00	.00
	Target guess	(1-R) × An × (1-On) × (1/2)	.00	.00
		Source identification based on word code with no target recall		
WCM	Cue remember	(1-R) × (1-An) × cn	.16	.22
	Target remember	− - - - - - - - - - - -	.00	.00
WCG	Cue guess	(1-R) × (1-An) × (1-cn) × (1/4)	.14	.19
	Target guess	(1-R) × (1-An) × [(1/12) × cn + 1/4]	.19	.26

Note. R = proportion of target words recalled; A_r = the proportion of source axes retrieved with target recall; A_n = the proportion of source axes retrieved with no target recall; O_r = the proportion of retrieved source axes correctly oriented when target recalled; O_n = the proportion of retrieved source axes correctly oriented when target not recalled; c_r = proportion of cue word locations identified when target word recalled; c_n = proportion of cue word locations identified when target word not recalled; t_r = proportion of target word locations identified when target word recalled. SAM (source-axis memory) indicates cue and target correctly remembered after retrieving the source axis and remembering correct orientation; SAG (source-axis guess) represents cue and target correctly identified after retrieving the source axis with correct guess for orientation; WCM (word-code memory) represents cue (target) correctly remembered after retrieving its word code; WCG (word-code guess) represents cue (target) correctly guessed after failing to retrieve its word code; P(Crt) represents the proportion of correct identifications; P(Crt|R) represents the proportion of correct identifications conditional on target word recall or no target recall.

With no recall very little of the identification originated from the source axis. Most of the identification performance for the cue word was the result of word code retrieval (WCM), .16, and for the target word this value was .00. This was because in the hybrid model $t_n = 0$. If the word codes could not be retrieved, then the cue word was correctly guessed .14 of the time and the target word .19 of the time (WCG). These guessing values are large when color is a source because so few of the source axes or word codes could be retrieved when the target word was not recalled. Correct guessing can sometimes play a large role in source identification when no other relevant information is available in memory (Bellezza, 2009).

Analyzing the item type by recall interaction. The first column of values shown in Table 5 represents the proportions of sources identified with the proportions representing all the items tested. However, only .29 of the pairs tested resulted in correct target recall, whereas the proportion of pairs tested accompanied by no target recall was .71 of the items. The interaction shown in Figure 7 shows the proportion of correct source identifications conditional on recall and on no recall. Therefore, the first column of proportions must be transformed to be conditional on recall and no recall. The second column of values in Table 5, labeled P(Crt|R), represents the proportions of correct color identifications conditional on the recall state. With recall, the four conditional values total .52 for the cue and .49 for the target word. With no target recall the two values are .42 and .27, respectively. These conditional proportions make up the inaction shown in Figure 7.

These values are somewhat different than those shown in Figure 7, because the numbers generated by the hybrid model are derived from aggregate data, whereas the data shown in Figure 7 are the means of conditional proportions computed separately for each participant. However, the pattern is the same: Color identification is almost the same for cue and target words when recall occurs but target identification is poorer than cue identification when target recall does not take place.

Graph of pattern of identification components. Table 5 lists the 16 identification components that comprise the item type by recall interaction when using color as a source. It is often easier to compare identification performance among different experimental conditions by graphing these components. In Figure 8 we compare

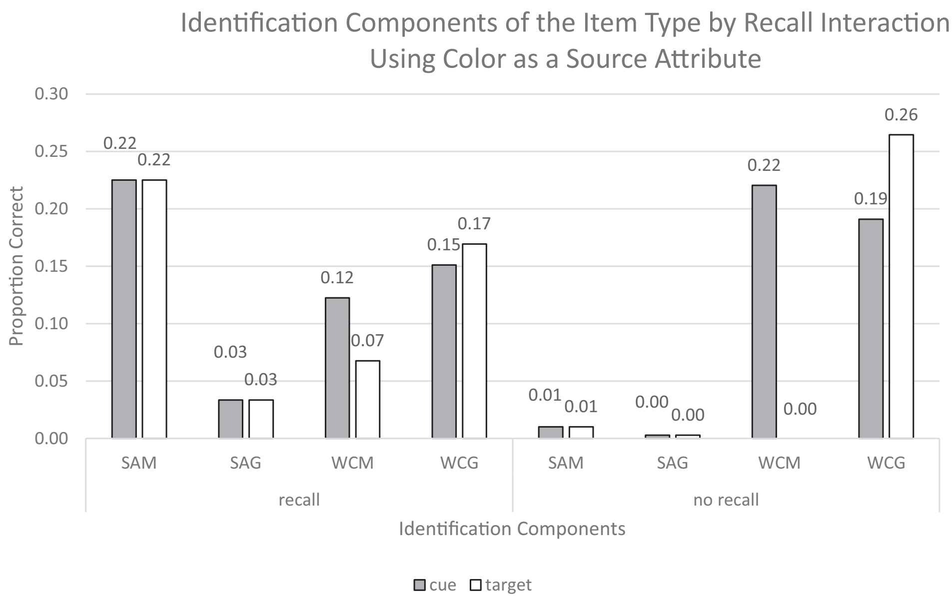

Figure 8. Proportions of correctly identified cue and target color (top panel) or correctly identified cue and target location (bottom panel) given target recall or no target recall generated by the hybrid model for the data of Experiment 1. SAM (source-axis memory) indicates cue and target correctly remembered after retrieving the source axis and remembering correct orientation; SAG (source-axis guess) represents cue and target correctly identified after retrieving the source axis with correct guess for orientation; WCM (word-code memory) represents cue (target) correctly remembered after retrieving its word code; and WCG (word-code guess) represents cue (target) correctly guessed after failing to retrieve its word code.

identification performance for color and location as source attributes using a graphical representation of the identification components.

Using color as a source. The pattern of source identification of color for the four components is graphed in the top panel of Figure 8. On the left side of the histogram are plotted identification performance mediated by the four components of the model when target recall occurred. Identification performance for the cue and target are similar except that more WCM identification appears to have occurred for cues, .12, than for targets, .07, when identification was based on word codes.

On the right side of Figure 8, representing color identification when target recall did not occur, the story is different. Virtually no source identification occurred as a result of identification components representing source axes (SAM, SAG) because neither A_n or O_n was significantly greater than zero. When color was identified using word codes (WCM), the proportion of cues identified was .22, whereas target identification was zero, because $t_n = 0$. Finally, correct guessing (WCG) of the target color, .26, was somewhat better than correct guessing the color of the cue, .19. However, Figure 8 makes clear that the interaction shown in Figure 7 is largely the result of no target word code being available when no recall occurred. This had a large effect because when using color as a source attribute, the participant must rely primarily on word codes because the source axis/arrangement for two colors is generally not available. With no recall the color of the target therefore must be guessed.

Using location as a source. The situation is different for location as a source. The second panel of Figure 8 graphs the identification components of the interaction between item type and recall, where this interaction measuring location was not statistically significant (Bellezza et al., 2016, Table 4). When recall occurred, most of the identification of cues and targets resulted from the source axes, especially SAM, with less from SAG, WCM, and WCG. When target recall did not occur, little identification was based on source axes. For the word codes the level of identification for the cue word was .10 for WCM compared to .00 for the target word. This difference, .10, was less than the corresponding difference of .22 found when color was used as source. With no target recall the WCG difference between cue words and target words when using location was not large.

The hybrid model fit data using both location and color as source attributes. Even though these two types of attributes seemed to be processed very differently, the difference was one of degree. The processes outlined in the hybrid model were used in each case, but event codes seemed to be more easily formed for location than for color. An interesting question is whether the hybrid model will be useful using other source attributes.

Validating the Hybrid Model

Although the hybrid model fit the data of Bellezza et al. (2016), the question remains of whether the parameters of the model actually represent memory retrieval of event codes, interactive imagery, source axes, and word codes in the manner we have proposed. This is the problem of validating the model (Bayen, Murnane, & Erdfelder, 1996; Buchner & Erdfelder, 1996; Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995; Erdfelder & Buchner, 1998; Klauer & Wegener, 1998). Recognition-based MPT source-monitoring models, for example, have been able to fit data collected in a variety of experiments. However, a concern has been that the parameters of these models do not correspond to the cognitive processes proposed as operating in the model. Bayen et al. tested three MPT recognition models. They found that for only the two-high-threshold source-monitoring model did the parameter values change in the manner expected under various experimental manipulations.

Bayen et al. (1996, p. 211) also suggested that MPT models are theories of decision processes not memory processes. We disagree with this characterization because the hybrid model introduced here, as well as other models, are primarily models of memory. In the hybrid model images, source axes, and word codes are considered memory constructs, that is, associative structures, not processes. These structures along with processes representing memory retrieval play important theoretical roles. The structures are identified as the visual images derived from combining the two words of the pair, the source axes contain source information regarding the pairs, and the word codes representing episodes comprised of individual words and context. The hybrid model has some similarity to the model of Batchelder and Riefer (1980) in which pairs of words separated in a list were presented followed by free recall of the individual words. Each pair was drawn from a different semantic category. As in the hybrid model, recall of each pair of words was mediated by an associative structure in memory. In the case of the Batchelder and Riefer model the mediator for recall was the memory representation of the category from which the two words were chosen. Furthermore, their model could separate the processes related to storage and retrieval.

The main process in the hybrid model is memory retrieval. Bias may play a role in choosing a location, but the word pairs have been evenly distributed across the 12 possible location combinations. Decisions were scored as correct or incorrect without specific locations becoming part of the data. Also, the model was fit to the data with the guessing probabilities assigned a priori.

The hybrid model is not what has been labeled a measurement model (Batchelder & Riefer, 1999; Bayen et al., 1996; Buchner & Erdfelder, 1996; Buchner et al., 1995; Erdfelder & Buchner, 1998; Klauer & Wegener, 1998) but is rather a descriptive model. The problem addressed by a measurement model is to assess a set of commonly acknowledged cognitive processes that need to be measured without confounding, without bias, and with minimal error. Rather than having this goal, the hybrid model attempts to explain the data of Bellezza et al. (2016) via associative mediating structures in memory that cannot be directly observed (Bellezza et al., 2016).

But as in a measurement model, one criterion of success for the hybrid model is that each experimental manipulation used should affect parameter values in a manner that depends on what each parameter represents, even though every parameter represents a type of memory retrieval. When a set of experimental conditions are varied, the set of parameter values should change in the directions expected based on assumptions of how the model operates. This represents *convergent validity* because a particular set of experimental manipulations are expected to converge on the manner in which they affect the value of a parameter.

There is another criterion for each model parameter of a measurement model, that of discriminant validity. This criterion suggests that there should be at least one manipulation that influences the value of its corresponding parameter in the manner expected from the model but does not affect the values of the other parameters. Finding such a manipulation for each model parameter establishes the model's discriminant validity. Each such manipulation is said to exhibit process purity with regard to the parameter. These two criteria are important in establishing the validity of a measurement model (Bayen et al., 1996; Buchner & Erdfelder, 1996; Buchner et al., 1995; Erdfelder & Buchner, 1998; Klauer & Wegener, 1998). However, we discuss below whether discriminant validity is relevant to the hybrid model.

The hybrid model has eight parameters all involving either the retrieval of a proposed mnemonic structure itself from episodic memory or the retrieval of item information from a mnemonic structure that can result in word recall or source identification. The use of eight parameters makes the hybrid model more complex and therefore less desirable than a model that has fewer parameters but can fit the data. This issue of complexity and parsimony has been discussed above. Eight parameters also makes the hybrid model more difficult to validate than a less complex model. Therefore, only a start can be made here in validating the hybrid model. Furthermore, only convergent validity will be tested, but with some discussion of convergent validity. Specifically, the validity experiment reported here, Experiment 2, will focus on the three parameters, R, A_r, and A_n related to the event code, with the other model parameters given less attention.

Experiment 2: A Validation Study of the Hybrid Model

Concern has been expressed regarding the role of the interactive imagery instructions in the formation of event codes in the hybrid model.[1] It may be that the importance of imagery instructions has been overestimated. Participants were assumed to form mental pictures based on interactive-imagery instructions, but interactive images may have occurred simply because the words in each pair were presented simultaneously. To alleviate with this concern, our first hypothesis was that the parameter R representing recall should be larger in value following interactive imagery instructions in the one-image condition than following instructions to form a separate image for each word in the two-image condition (Bower, 1972).

Recall of the target word from the cue word of a pair means that an additional prompt is available for retrieving the source axis with source information. That is, the conditions of creating the event code at the time of learning have been more completely recreated at test than if the target word had not been recalled. Hence, by the principle of encoding specificity (Tulving & Thomson, 1973) retrieval from memory of the target word results in better identification of source locations than if no target recall occurred. Recall adds the target word as an additional prompt for source information. Therefore, our second and third hypotheses are that parameter A_r should be larger than A_n in both the one-image and two-image conditions.

The event code is made up of two components, the interactive image and the source axis. The interactive image and the source axis are positively but not perfectly associated in memory and can be retrieved together or separately (Bellezza et al., 2016). Two additional hypotheses follow from this assumption. In the one-image condition there should be a greater number of interactive images than in the two-image condition, especially when made conditional on target recall, which indicates the retrieval and use of the interactive image. That is, recall performance should be positively related to source identification. Hence, our fourth hypothesis is that A_r should be greater in the one-image condition than in the two-image condition. The fifth hypothesis is that parameter A_n will be greater in the one-image condition than in the two-image condition, because there are fewer interactive images in the two-image condition whether target recall occurs or not.

In the hybrid model, it is assumed that memory structures play an important role in mediating source-identification performance (Bellezza et al., 2016). A mediating structure provides distinctive context for the encoding and retrieval of item information but is not manifested by an overt response. That is, the learner is aware of these structures during learning, and by being aware of these structures at the time of test, optimal remembering can occur. This occurs when the visual image is retrieved from memory to facilitate target recall. Also, memory structures establish relationships among the various retrievals that can occur in the hybrid model, as when the source axis allows the identification of location of both the cue and target words. Our Hypothesis 6 states that more source axes will retrieved in the one-image condition than in the two-image condition. As with validating the model parameters, thoroughly validating the proposed associative structures in the model will require additional research.

We have described a word code as allowing source information to be retrieved from the representation of the word itself in memory. The parameters c_r, c_n, and t_r have been used as equivalent to the probability of identifying the source of a cue word or target word when a source axis cannot be retrieved. But a word code is actually a mediating memory structure that may be associated with an assortment of source information. When a word code is retrieved it may be that the participant becomes aware of a number of source attributes of the presented word in addition to its location. For example, using a recognition paradigm, Meiser and Bröder (2002) and Meiser (2014) presented items with multiple attributes and found that identification of one attribute was positively related to identification of the other. Awareness of the word code when identifying multiple attributes will optimize remembering. This follows from the principle of encoding specificity (Tulving & Thomson, 1973). The relative frequency of word codes is related to the successful retrieval of source axes. Fewer word codes should be retrieved in the one-image condition than in the two-image condition because in the hybrid model word codes are retrieved only if retrieval of the source axis fails. So, our seventh hypothesis is that more word codes will be retrieved in the two-image condition. An experiment was performed to address these six hypotheses.

Method

Materials. Eight hundred forty-five words were sampled from Toglia and Battig (1978) such that all words had mean ratings above 4.50 on seven-point scales of concreteness, imagery, familiarity, and meaningfulness. Ninety-six words were then randomly selected to construct 48 word pairs with two words in each pair not obviously related in meaning. Examples are father–beach, bird–nurse, and heart–window. The words from each pair were randomly placed in different cells of a two-by-two grid of the kind shown in Figure 3. Each of the six possible combinations of using two spaces from the grid was utilized 8 times. This list was made up of the same pairs used in Study 1 of Bellezza et al. (2016).

Participants and procedure. Sixty undergraduate students participated for extra course credit. The participants were tested in small groups of four to eight with pairs projected on a large screen in the front of the testing room. The 30 participants in the one-image condition were instructed to form a mental picture for each presented pair such that the visual representations of the referents of each word interacted in some manner. The participants were informed that they would later be tested for recall and would have to recall one of the words when given the other. Thirty participants

[1] We thank an anonymous reviewer for suggesting the validation procedure used in Experiment 1.

in the two-image condition were instructed to form a separate mental picture representing the meaning of each word in the pair. These participants were told to form each separate image carefully because they would later be asked questions regarding the contents of each image. For participants in both conditions no reference was made to the 2×2 grid on which the pairs were presented or to the fact that the words of different pairs appeared in various locations in the grid.

Each word pair was presented for 6 s with the entire list presented twice in identical order. Following presentation, participants solved anagrams for 3 min to eliminate any possible effects of short-term memory. They then received a test booklet of 48 blank two-by-two grids with the cue word from each of the 48 pairs in a random order printed to the left of each grid. Each cue word was selected randomly from the two words in each presented pair such that there were an equal number of cues from each location. Participants recalled the appropriate target word by printing both the cue word and the target word in their studied locations. Every test item had to be completed fully, even if the target word and the word locations were guessed. The test was described to the participants as a test of incidental memory because participants in the one-image group were not previously informed that they would have to identify locations and participants in the two-image group were not previously informed that they would have to recall a second word from each pair and remember the locations of the two words.

Results

There are five different sets of results. First, some descriptive statistics are presented indicating the different effects of the two imagery strategies on recall and on source monitoring performance. The second analysis shows that the hybrid model can fit the data of each condition, even though different learning strategies were used. Third, we test Hypotheses 1 to 5 that demonstrate construct validity for the parameters R, A_r, and A_n. Fourth, we test Hypotheses 6 and 7 demonstrating the validity of the source axes and word codes. Fifth, the construct validity of other parameters in the model are discussed.

Descriptive statistics. For the one-image condition the proportion of target words recalled was .67 and for the two-image condition was .24. This result suggested that participants were generally trying to follow mnemonic instructions, although not all participants typically do so (Pash & Blick, 1970). Furthermore, our recall findings were comparable to those of Bower (1972) who reported that .71 of the words were recalled in the condition in which instructions were given to form interactive images and .46 of the words in which instructions to form separate images were given.

This experiment was similar to that of Bower (1972) except that Bower did not mention a later memory test, whereas we instructed participants in the one-image condition to remember the two words in each interactive image and instructed participants in the two-image condition only to remember the contents of the image formed for each word. This difference in the specificity of learning goals in the two conditions may have been confounded with the number of images that had to be created. That is, intention to remember the pairs may have helped performance in the one-image condition more than in the two image condition.[2] However, in a related experiment Bower (1972) tested participants in three conditions: In one condition participants were instructed only to memorize the pairs, in a second condition participants were instructed only to form an interactive image with no mention of a later memory test, and in a third condition participants were instructed to memorize the pairs plus form an interactive visual image. The three groups recalled, respectively, .35, .71, and .77 words from the pairs when cued with one of the words. He concluded that intention to learn during encoding was not as important as the mnemonic strategy used. Thus, in Experiment 2 there could be some advantage in recall in the one-image condition because of participants' awareness of nature of the memory test but it was not as important as the type of image used. As discussed below, even with this somewhat problematical experimental design, the hybrid model nevertheless fit the data and the model parameters could be interpreted.

Level of performance on source identification also was better for participants in the one-image condition than for the two-image condition. The identification of the location of the cue and target were .49 and .51, respectively, for the one-image condition and .43 and .40 for the two-image condition. A 2×2 analysis of variance testing type of instructions and item type showed that overall source identification was better for the one-image condition than for the two-image condition, $F(1, 58) = 4.52$, $MSE = .049$, $p < .05$, partial eta-squared $= .07$. Also significant was the Image Condition by Item Type interaction, $F(1, 58) = 5.87$, $MSE = .005$, $p < .05$, partial eta-squared $= .09$. Source identification for target words in the one-image condition was greater (.51) than in the two-image condition (.40), with no two other interaction means being significantly different.

Figure 9 shows mean performance across participants when identifying the source of the items in each pair conditional on recall of the target word. A $2 \times 2 \times 2$ analysis of variance with factors type of image, type of word in pair, and recall versus no recall of target showed that target recall resulted in better source identification than did no target recall, $F(1, 55) = 137.19$, MSe $= .016$, $p < .001$, partial eta-squared $= .71$. The analysis also detected a significant interaction effect that is not clearly indicated by the confidence intervals in Figure 9. This was the Recall $\times$Word Type interaction, $F(1, 55) = 10.03$, MSe $= .011$, $p < .005$, partial eta-squared $= .15$. When recall occurred, the target word was identified somewhat better (.64) than was the cue word (.59). With no recall, the target word was identified somewhat worse (.32) than the cue word (.35). In every condition, however, cue and target identification were comparable in value, which was one of the results reported by Bellezza et al. (2016) and displayed in Table 1. However, the difference between the one-image and two-image conditions did not reach the level of statistical significance when measuring source identification, $F(1, 55) = 2.78$, $p > .10$.

Fitting the hybrid model to the one-image and two-image data. The multiTree program (Moshagen, 2010) was used to fit the hybrid model to the aggregated frequency data and to generate the results using a bootstrap procedure with $N = 1{,}000$ (Moshagen,

[2] We thank David Riefer for pointing out that the difference in the specificity of learning goals in the two conditions was confounded with the number of images that had to be created.

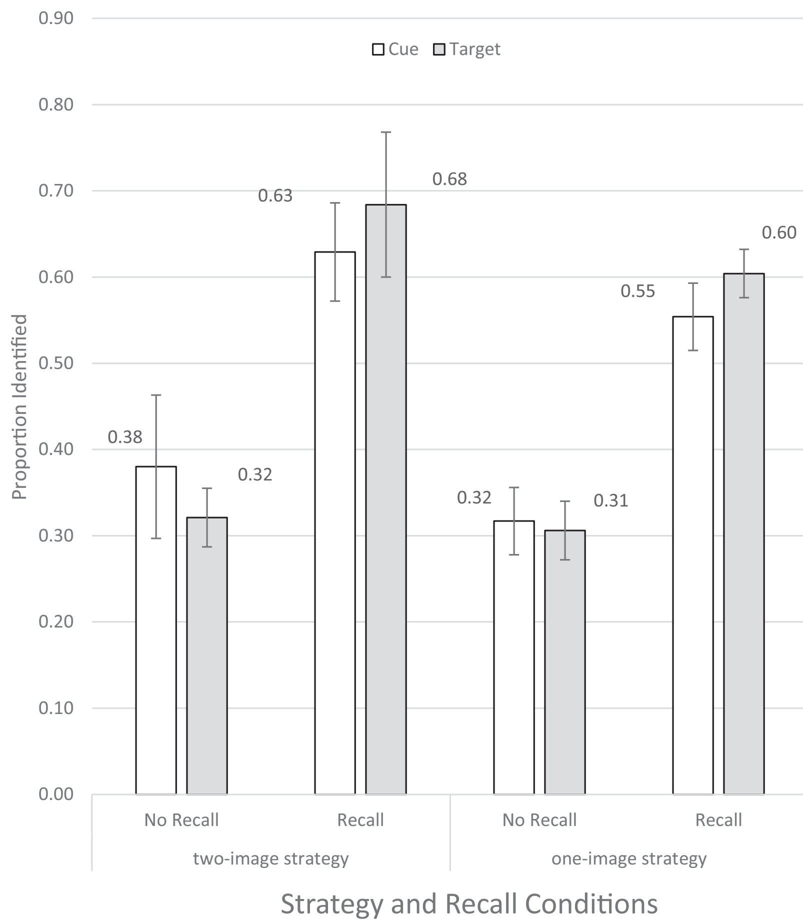

Figure 9. Proportions of cue words and target words identified for the unrecalled and recalled pairs in the one-image and in the two-image learning conditions in Experiment 2. Error bars represent 95% confidence intervals.

2010). To obtain some indication of the extent to which the two sets of parameter values differed, both sets of data were analyzed simultaneously with a separate set of parameters estimated for each type of data (Batchelder & Riefer, 1990; Erdfelder et al., 2009; Riefer & Batchelder, 1988). The probability of a Type I error, alpha, was set to .05. With a small effect size, w = .08, the power, that is, the probability of rejecting the null hypothesis that the model fits the data when it does not, was .98 (Erdfelder, Faul, & Buchner, 1996).

The resulting goodness-of-fit statistic was $G^2(2) = .71$, $p > .50$, where each of the two sets of frequencies, when tested separately, also were fit by the hybrid model. The parameter values for each of the two conditions are shown in Table 6. Also shown in Table 6 are z tests comparing the parameter values in the two conditions and z tests to determine if each parameter value was significantly greater than zero. The observed frequencies and the frequencies generated by the model are displayed in Table 3. It should be noted that all the parameters of the hybrid model had to be used. When the data from each condition was tested separately with only the nested event-code model or with only the nested word-code model, none of the four tests approached a satisfactory fit. If only the event-code portion of the hybrid model was needed for paired-associates data, as in two of the conditions of Bellezza et al. (2016), the hybrid model would be a less complex and more parsimonious model requiring only four parameters.

Construct validity of parameters *R*, A_r, and A_n. Our first hypothesis was confirmed. As shown in Table 6, recall in the one-image condition, R_1 = .67 was significantly larger than recall in the two-image condition, R_2 = .24, replicating the results of Bower (1972).

Confirming the second and third hypotheses, parameter A_r was greater in the one-image condition when target recall occurred, Ar_1 = .51, than when target recall did not occur, An_1 = .05. This was also true in the two-image condition with Ar_2 = .42 and An_2 = .10.

The fourth and fifth hypotheses were that the imagery manipulation would result in better retrieval of source axes, A_r and A_n in the one-image condition than in the two-image condition. The parameter for the one-image condition, Ar_1 = .51, was significantly larger than the parameter for the two-image condition, Ar_2 = .42.

Somewhat surprisingly, the fifth hypothesis was not confirmed. With no target recall there was a higher probability of retrieving the source axis for pairs in the two-image condition than in the

Table 6
Parameter Values of the Hybrid Model When Using the One-Image Encoding Strategy or the Two-Image Encoding Strategy in Experiment 2

One-image encoding strategy				Two-image encoding strategy				Compare the two strategies
Test against zero				Test against zero				Test of difference
Parameter label	Parameter value	*SD*	*z*-test *p* value	Parameter label	Parameter value	*SD*	*z*-test *p* value	*z*-test *p* value
R_1	.67	012	***	R_2	.24	011	***	***
Ar_1	.51	021	***	Ar_2	.42	040	***	**
An_1	.05	020	**	An_2	.10	016	***	**
Or_1	.75	039	***	Or_2	.72	076	***	n.s.
On_1	.81	266	**	On_2	.93	100	***	n.s.
cr_1	.02	026	n.s.	cr_2	.11	059	*	*
tr_1	.17	034	***	tr_2	.22	061	***	n.s.
cn_1	.03	027	n.s.	cn_2	.11	023	***	***

Note. R = proportion of target words recalled; A_r = the proportion of source axes retrieved with target recall; A_n = the proportion of source axes retrieved with no target recall; O_r = the proportion of retrieved source axes correctly oriented when target recalled; O_n = the proportion of retrieved source axes correctly oriented when target not recalled; c_r = proportion of cue word locations identified when target word recalled; c_n = proportion of cue word locations identified when target word not recalled; t_r = proportion of target word locations identified when target word recalled.
* *z*-test significant at $p < .05$. ** $p < .01$. *** $p < .001$. Standard deviation values should be multiplied by .001. All z tests used means and standard deviations estimated using a parametric bootstrapping procedure where N = 1,000 (Moshagen, 2010).

one-image condition, with An_2 = .10 significantly greater than An_1 = .05. This result is contrary to what we expected from the hybrid model. Even without recall of the target word, the source axis of an item would be more available in the one-image condition than in the two-image condition. This was because more source axes should be available in the one-image condition.

It is not clear what the explanation is of this unexpected result. According to the hybrid model, one possibility is that a number of source axes, but not interactive images, were created joining the two locations in the two-image condition. But because of instructions relatively few interactive images were formed compared to the one-image condition. Therefore, some of the source axes formed were not associated with any interactive image but, nevertheless, were available for source identification. Using a similar argument for the one-image condition, the images formed tended to mediate recall efficiently which also resulted in the retrieval of the source axes. Thus, in the one-image condition most source axes formed tended to be retrieved along with the target word.

Other data has been presented here involving the parameters *R*, A_r, and A_n, and these data may provide additional information regarding the validity of these parameters. When comparing color and location as attributes in Experiment 1 recall was greater for location as was A_r. As shown in Table 2, for all five conditions of Bellezza et al. (2016) the value of parameter A_r was markedly greater than the value of A_n. Also, this was true for the color data analyzed in Experiment 1.

For Hypotheses 3 to 5, the predictions are that A_r and A_n will be larger in the condition where recall is greater, Bellezza et al. (2016), in Study 2, tested a list of pairs of unrelated words versus related words using a within-subjects design. Recall, *R*, was greater in value for the related pairs, but the values for A_r were similar in value. However, A_n was greater for the unrelated pairs even though recall was lower. Study 3 tested a list of pairs where the cues were either words or letters and used a within-subjects design as did Study 2. Similarly, Study 3 also produced anomalous results. Recall was larger in value for word cues but A_r was larger for the letter cues. There were no differences in the values of A_n. So, the predictions were generally found in Experiments 1 and 2 using a between-subjects design but not in Bellezza et al. using a within-subjects designs. Different types of pairs appearing in a list may induce different learning strategies for each type of pair (for a discussion of this issue, see Rowland, Littrell-Baez, Sensenig, & DeLosh, 2014). Also, in Study 3 in the 48 pairs presented, one third were letters and one third were words. Bellezza et al. suggested that the letters would comprise a smaller category in memory than would the word cues and thus be more easily retrieved.

Validation of the frequency of source axes. Hypothesis 6 states that more source axes will be retrieved in the one-image condition than in the two-image condition. This is because there is a positive relation between both the visual image and recall of the target word and between the visual image and the source axes. If we let S represent the event of a source axis being retrieved for an item in either the one-image or two-image condition, then P(S) = P(S&R) + P(S&~R), where R represents the event of target recall and ~R represents the event of no target recall. Next, P(S) = P(S|R)P(R) + P(S|~R)P(~R), where P(S|R) = A_r, P(S|~R) = A_n, and P(~R) = [1 − P(R)]. So, P(S) = RA_r + (1-R)A_n. Using the parameter values from Table 6, we find for the one-image Condition P(S) = .36 and for the two-image Condition P(S) = .17. Because there were an aggregated 1440 pairs tested each condition, in the one-image Condition 1440 × .36 suggests approximately 518 source axes were retrieved during testing, whereas in the two-image condition the number 1440 × .17 suggests approximately 245. This result supports Hypothesis 6.

Validation of the frequency of word codes. As explained above in our discussion of Hypothesis 7, P(C), the probability of retrieving the word code for the cue word is not the same as the parameters c_r and c_n, and P(T) is not the same as t_r. Hypothesis 7 states that a greater number of word codes will be retrieved in the two-image condition than in the one-image condition. If we let C represent the event that a cue word-code is retrieved from memory to identify the location of the cue, then P(C) = P(C&S) + P(C&∼S). However, P(C&S) = 0 because retrieval of a source axis means that the word code is not retrieved. Furthermore, P(C&∼S) = P(C&∼S&R) + P(C&∼S&∼R) = P(C|∼S&R) P(∼S|R)P(R) + P(C|∼S&∼R)P(∼S|∼R)P(∼R), where P(∼R) = [1-P(R)]. Substituting the hybrid model parameters into the equation, we get P(C) = $c_r(1 - A_r)R + c_n(1 - A_n)(1 - R)$. When substituting the parameter values from Table 6 for the one-image condition we get .02 and using the parameter values from the two-image condition we get .09. More word codes for cues are retrieved in the two-image condition that in the one-image condition.

If we let T represent the event that a target word is retrieved for an item then, by analogy to P(C) derived above, P(T) = $t_r(1\text{-}A_r)R + t_n(1\text{-}A_n)(1\text{-}R)$. But t_n is assumed to be zero in the hybrid model, so P(T) = $t_r(1\text{-}A_r)R$. This value is .06 in the one-image condition and .03 in the two-image condition. There is a slight advantage in target identification in the one-image condition. In summary, more source axes were retrieved in the one-image condition, but more word codes for cues were retrieved in the two-image condition. This result supports Hypothesis 7.

Construct validity of other parameters. So far, the only parameters mentioned with regard to construct validity are parameters R, A_r, and A_n. The results of Experiment 2 suggest that once that a source axis has been retrieved, the orientation parameters O_r and O_n, in the recall outcomes and no-recall outcomes, respectively, have much the same values. For the one-image Condition Or_1 = .74 and in the two-image Condition Or_2 = .72, which were not significantly different. Similarly, On_1 = .81 and On_2 = .93, which were also not significantly different. Once a source axis was retrieved, the probability of remembering the orientation of the two words on the axis was similar for both conditions.

Based on the experimental manipulation of imagery and learning instructions, the magnitude of the parameters representing memory retrieval from word codes can also be compared conditional on target recall or no recall. For example, when recall occurred, the values for c_r and t_r should be larger in the two-image condition because participants in this condition were instructed to focus on the individual words when creating their visual images. As shown in Table 6, cr_2 = .11 was significantly greater than cr_1 = .02. Also, parameter tr_2 was larger than tr_1 but not significantly so. When target recall did not occur, cn_2 = .11 was significantly larger than cn_1 = .03. These results demonstrate a superiority for the two-image condition when using word codes for source identification.

Study 2 of Bellezza et al. (2016) used word pairs containing either related or unrelated words and found, as expected, that recall was larger for the related pairs compared to the unrelated pairs. Furthermore, this manipulation gave information about parameters Or and On that was not provided by Experiment 2. The value for O_r was greater for the unrelated words, .84, than for the related words, .75. Also, the value for O_n was greater for the unrelated words, .71, than for related words, .35. When tested together these two differences were significant, $G^2(2) = 6.30$, $p < .04$. This outcome resulted from the related words having more similar representations in the propositional network than did the unrelated words. This was true for both the target recall and target unrecalled outcomes. The related words were more confusable and orientation information less reliable.

Discussion

Experiment 2 was primarily a validation study, but it is noteworthy that the hybrid model fit the data of both conditions. The two-image condition was the first time that the fit of the hybrid model was tested when a mnemonic strategy was not used for associating the words in each presented pair. The construct-validation process for the parameters R, A_r, A_n, O_r, O_n, c_r, c_n, and t_r showed these parameter values changing as they should as experimental conditions were varied. The only exception to this was the result that parameter A_n was somewhat greater in the two-image condition than in the one-image condition. Also, some other results from Experiment 1 and from Bellezza (1996) were suggestive of what directions validation studies should proceed in the future. At this point the validation evidence for the hybrid model is tentative and incomplete. More conclusive validation results are necessary.

Establishing discriminant validity for a parameter in a measurement model occurs by finding an experimental manipulation that affects the value of that parameter in a predictable direction but does not affect the values of any of the other parameters in the model (Bayen et al., 1996; Erdfelder & Buchner, 1998; Klauer & Wegener, 1998). The assumption that such a manipulation exists for every parameter in the measurement model is a strong assumption regarding the nature of cognition and its relation to the domain of possible experimental manipulations. It is not clear that discriminant validity is necessary for the hybrid model, which we have labeled a descriptive model, not a measurement model. Experiment 2 tested construct validity using two conditions in which the imagery strategy for remember pair information was varied. However, results from Bellezza et al. (2016) and from Experiment 1 using color as a source attribute were also reviewed for their relevance to the construct validity of the model. In each of these studies varying one experimental condition produced changes in many of the parameter values. Additional efforts are needed to find manipulations that will affect one parameter without affecting the others, if the criterion of discriminant validity is to be satisfied.

Klauer and Wegener (1998) have noted the importance of the power of statistical tests when establishing the discriminant validity of each parameter value in the model. Too little power will fail to find differences among the remaining parameter values that show statistically little change in value but actually may be different. On the other hand, too much power will detect small differences among parameters that should be identical in value. In this latter case, the problem is how to decide whether these differences are substantial enough to not meet the criterion of discriminant validity. The topic of validating and falsifying all types of MPT models, including the hybrid model, requires further research and discussion.

Testing the Source Identification of Unrecalled Words

The hybrid model demonstrates how the source attribute of an unrecalled target word can be identified using a paired-associates learning paradigm. This identification is the result of the use of a memory structure labeled the source axis or source arrangement. It is convenient, however, to be able to study this same phenomenon by analyzing recall performance without having to fit the hybrid model to the data. A procedure for doing this is described.

Bellezza et al. (2016) reported that source location performance for either cue words or target words was above the chance level of .25 when both locations were guessed following no target recall. But participants may not frequently have to guess the locations of both cue and target words. Furthermore, if we identify the source of the cue word, then the probability of guessing the target location is not .25 but .33, because one of the three remaining locations is the location of the target word. This means that if, for whatever reason, the source of the cue word is known to be correct, the target word will be correctly guessed above the .25 level.

What if the proportion of identification errors for the cue words is known, but the location of each error is unknown? This occurs when given the error rate for cue identification. If the cue location for an item is known to be incorrect but its location is unknown, then the probability of the guessing the target word can be computed. Let's say that the cue and target should be in Locations 1 and 2, respectively, that is, Locations (1, 2). Of the 12 possible combinations of placing the cue and target words in the four locations, only 9 can be used for guessing the location of the target word when it is known that the cue location is incorrect. This is because for three choices, Locations (1, 2), (1, 3), and (1, 4), of the 12 choices the cue would be placed in its correct location, which is not the case. Of the nine remaining only two guesses, Locations (4, 2) and (3, 2), result in a successfully placed target word. Hence, if the cue word is not in the correct location and its location is not known, then the probability of guessing the target location is two out of nine or .22.

If ~R signifies the event of no target recall, then let $P(C|{\sim}R)$ be the probability that the cue word was placed in the correct location, for whatever reason, and let and $P(T|{\sim}R)$ be the probability of placing the target word in the correct location by chance. Then expected $P(T|{\sim}R) = (1/3)P(C|{\sim}R) + (2/9)[1 - P(C|{\sim}R)]$, where the value of $P(C|{\sim}R)$, $P(T|{\sim}R)$ and therefore expected $P(T|{\sim}R)$ can be computed for each participant. The difference between observed $P(T|{\sim}R)$ and the expected $P(T|{\sim}R)$ can be tested against zero using a *t* test.

This test was performed on the four sets of data discussed in Experiments 1 and 2. In Experiment 2 a directional *t* test showed that for the one-image condition the target was identified above chance level, $p < .05$. Also, as shown in Table 6, the values of parameters $A_n = .05$ and $O_n = .81$ for the hybrid model were both significantly greater than zero. So, the results of the *t* test corresponded to what could be concluded from testing the significance of the A_n and O_n parameters. In addition, for the two-image condition both procedures demonstrated source identification of unrecalled target words with $p < .005$ using the *t* test with model parameters $A_n = .10$ and $O_n = .93$ both significantly greater than zero.

In Experiment 1 for the color condition the *t* test was nonsignificant. For the hybrid model the parameter values were $A_n = .02$ and $O_n = .58$, both not significantly greater than zero. Finally, in the location condition $p < .001$ for the *t* test with $A_n = .10$ and $O_n = .57$, significantly greater than zero. In the four conditions tested here the *t* tests of the identification performance of unrecalled targets led to the same conclusions as testing the significance of the parameters A_n and O_n.

It is possible that the *t* test may give somewhat different results compared to using the hybrid model. For the *t* test the difference between the observed and expected $P(T|{\sim}R)$ must be computed for each participant, and the mean difference tested against zero. On the other hand, the hybrid model fits data aggregated across participants.

General Discussion

The Hybrid Model and the PAL Procedure

In the paired-associates learning discussed here participants were asked to study pairs of words high in concreteness with each word in a different location or printed on a different color background. Except for the two-image condition of Experiment 2 they were instructed to form a visual image with the referents of the two words interacting in some manner. The hybrid model proposed is an MPT model made up of retrievals of information from memory and from associative structures in memory which act as unobserved mediators of learned information. Information in memory representing any aspect of the experimental presentation connected to the word pair is referred to as an event code. A word from the word pair is later presented to retrieve part of the event code, the interactive image, so that the other word can be recovered from this image. In addition to the image associating the two words, a second structure, the source axis, may be part of the event code. It contains information about the sources of the two words organized in such a manner that retrieval of one source value is accompanied by retrieval of the other. The source axis may be associated with the interactive image in the event code.

Also, there are word codes. While forming an interactive image for the two words in the pair, the participant may also associate the source of each individual word to some representation of that word in memory. This representation, or word code, can be used to identify the word's source if the source axis cannot be retrieved or, in some instances, can be used to identify multiple source attributes. These processes are expressed in the form of the hybrid model diagrammed in Figure 5.

Two characteristics of the hybrid model are problematic. As an MPT model it has eight parameters and so is relatively high on complexity measures. But this is not a fatal flaw (Batchelder & Riefer, 1999), especially because, at present, these is no competing model. The second problem is that construct validation is not complete for all the parameters, especially with regard to the parameter A_n. Many possible experimental manipulations are possible in testing the hybrid model. These include changes in strategy, type of materials, and changes in experimental design. It is possible that future outcomes will suggest that our interpretations of one or more of the more parameters are not well-defined.

Mnemonic Devices and Memory Schemas

Bellezza et al. (2016) used a particular paired-associates procedure that involved four possible source locations for each word in the pair. A mnemonic strategy was provided to the participants to maximize the number of target words available for source testing. As shown in Experiment 2, this mnemonic strategy had a facilitating effect both on word recall and on the identification of the sources of the words. Moreover, it is possible that the manner in which location was used as a source attribute provided a memory schema operating on the visual field that influenced both word recall and the identification of word location. No similar schema seemed to be available with the use of color as a source in Experiment 1.

Both the visual image and the location schema act as mediating structures in memory that aid learning. The visual image is created during the encoding of the pair and stores information that will later help recall. According to encoding specificity (Tulving & Thomson, 1973), the learner must be must have the image as a mediating structure in mind both during encoding and during memory retrieval. The same is true when using the location schema. Although location is part of the visual field during encoding, during retrieval the participant must retrieve the spatial array, that is, one of the six possible axis, before recall. We suggest that like the visual image, the location schema is a mediating structure that helps organize the spatial information. Although these mental structures of images and schemas cannot be directly measured, it is possible to collect verbal reports that can provide validity regarding these structures (Ericsson & Simon, 1993). For example, Reddy and Bellezza (1983) have collected and compared such verbalizations generated during the study and the test phases of a free-recall task.

Mnemonic devices. When a mnemonic strategy is used, as it was here, it must be deliberate and intentional on the part of the learner (Bellezza, 1981; Paivio, 1971). The visual image formed becomes part of the event code for the pair; that is, part of the set of information in memory representing the presentation of a particular pair. When trying to associate unrelated words using visual imagery, semantic memory must be searched (Shiffrin, 1970) in a deliberate manner to try to retrieve some type of information that can relate the two unrelated words in each studied pair. That is, information is needed from semantic memory that serves to link the two words in a manner that might be characterized as using a proto-schema comprising some existing, but weak, connection between the words or as using an ad hoc category (Barsalou, 1983) containing the two words. In either case any weak, but available, relation found will proactively facilitate associating the two words in the newly formed episodic representation (Bellezza, 1996). That is, this transfer will aid in remembering.

Locations as part of a schema. The information regarding the sources of location and color were not learned, as far as we can determine, using a mnemonic device, and no instructions for remembering the source were suggested by the experimenter. In Experiment 2, word location was not even referred to in the participant's instructions but was nevertheless remembered, with source axes more frequently retrieved in the one-image condition than in the two-image condition. We suggest that when using location as a source attribute, it acts as an organizational framework for the source of the words. Specifically, when the two words of a pair were presented within the framework of a familiar two-dimensional visual field, the two objects actually appeared in a mutual relationship. Pairs of locations in a visual field form a familiar framework, which is schematic even if not formally acknowledged as a schema (Tversky, 2001). The two locations can be perceived and imaged independently of the particular objects placed in those locations.

The 2 × 2 grid acts as a schema that can be instantiated in different ways, as shown in Figure 3. The six axes proposed for organizing the locations of the pairs of words enable participants to organize in memory the two locations for each pair. The axes and their accompanying orientations result from the Gestalt principles of similarity and direction. Participants first locate two words, that is, locate the two stimuli in the visual field that are similar to each other as words and then mentally draw a directional line connecting them. The words' position on this line must also be remembered. This is done with reference to the 2 × 2 grid upon which the words are presented.

There may also be an attentional factor explaining why location is more effective than color as an attribute in the storage of source information. Treisman and Gelade (1980; Treisman, 2006) have suggested that the location of a studied item may typically be processed without deliberation because participants must locate each word in the visual field before they can process it further. Hence, location is always attended to, and attention enables the remembering of location. This, however, is not the case when using a color background for a word. As shown in Figure 6, the words from each pair were always located in the same two adjacent location, but printed on different color backgrounds. No visual search was necessary to locate them. Furthermore, Chen and Wyble (2015) have reported that that the location of an object seems to be processed automatically, that is, with less attention than is needed for color.

Locations may organize words for recall. One of the surprises of Experiment 1 was the superiority found both in recall performance and in source identification performance when location rather than color was used as a source attribute. Location as a source attribute seemed to facilitate remembering both the word pair and the source information presented with it. With location as a source attribute target recall occurred for .44 of the pairs but when color was used, this value was .29, even though the mnemonic imagery instructions and pairs were the same in both conditions. Similarly, when location was used, .54 of the cues and .51 of the target were identified. However, when using color these two values were .46 and .33. As shown in Table 4, parameters A_r, O_r, A_n, and O_n were all significantly greater than zero in the location condition, demonstrating the presence of source axes even without target recall, whereas in the color condition only A_r and O_r were greater than zero.

It might be that location also acts a memory mediator for word recall by which word information is stored using the source axes as well as the interacting image. There is evidence indicating that organizing words spatially on a visual field can enhance word recall. In studies involving the long-term retention of words Bellezza (1983, 1986) found that spatial arrangements of words resulted in better word recall when the arrangements were different from each other compared to when the spatial arrangements were the same. Lists of six words were arranged in simple geometric configurations such as a plus sign, a square, the letter K, and so on. The group that studied the lists of words in different configurations recalled more words when given one word from each list as a cue than the group who were presented the words always in the same arrangement. For the data analyzed here, recall may have been enhanced by the distinctive source axes acting as mediators for the words. Thus, the distinctiveness of the six axes describing the pairs of locations may have been a factor in enhancing recall performance.

The organizational process used by each participant to associate the words of a pair was a deliberate strategy of mnemonic learning based on visual imagery. But memory for source location, we suggest, is the result of a type of schema-based learning which occurs unintentionally because of the influence of prior experience. We suggest that the mnemonic procedure and memory schema operate together to form two associated parts of the event code. Information organized by a mnemonic strategy is combined with source information that has been organized in memory using the location schema. It has been suggested that mnemonic devices and memory schemas have many similarities in their structure and operation which enable them to operate together (Bellezza, 1987; Bellezza & Reddy, 1978).

This synthesis of mnemonic-based and schema-based learning would explain why the use of the interactive imagery of the one-image condition in Experiment 2 resulted in better overall recall and source identification than did the imagery used in the two-image condition. If we think of the source-axis as an image, the creation of the interactive image allowed the source-axis to become associated to it. Thus, retrieval of the interactive image could result in both target recall and source identification using the source axis. When target recall failed, the positive correlation of retrieval of the cue-target image with the retrieval of the source-axis means that the source axis was also not likely retrieved, but its retrieval was still possible. This process may be similar to the manner in which disparate components of any experienced event and their combinations are retrieved from memory.

What Is the Memory Structure for Pairs of Colors

The notion of a source axis was used in the sense diagrammed in Figure 3 where the combination of axis and orientation can describe the 12 possible configurations of cue and target words. The conceptual problem is that source axis is a mental construct that is based on source location and represents in memory the two words placed on a linear axis in the two-dimensional visual field. Pairs of colors do not seem to form such an organizational framework in semantic memory that is also available to the cognitive system. Analysis of the data in the color condition in Experiment 1, however, found that A_r and O_r were greater than zero, but not A_n and O_n. That is, source arrangements were playing a role in the color condition when target recall occurred but not when recall did not occur. This was confirmed by showing that the hybrid model could not fit the color data if A_r and O_r were set to zero forcing the model to rely only on word codes for source identification. As with the data from Experiment 2, neither the event-code model by itself nor the word-code model by itself could fit the color data.

These results suggest that there is some way of storing the color information for the two words using a color axis. A better word than *axis*, perhaps, might be the terms *color arrangement* or *color armature*, but the word must convey more than simply associating each color to the appropriate word in the pair, as is the case when using word codes. The source arrangement of colors must provide a structure already existing in semantic memory in which the two colors have a relation separate from the two words in the visual image used for recall. This arrangement can then be included in the event code. In our daily lives we are used to arranging and remembering objects in different locations, but we are not used to arranging and remembering pairs of objects of different colors. It may be that an organization exists for some pairs of colors, such as for red and green (traffic lights representing stop and go) or for red and blue (political orientation of states as conservative and liberal). But pairs of colors, other than those mentioned, do not frequently characterize the events we experience. Thus, previously stored information may provide the basis for forming source arrangements based on some pairs of colors, but it is difficult to conceive of a set of existing memory relations consisting of all pairs of colors.

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning & Verbal Behavior, 22,* 261–295. http://dx.doi.org/10.1016/S0022-5371(83)90201-3

Anderson, J. R., & Bower, G. H. (1974). A propositional theory of recognition memory. *Memory & Cognition, 2,* 406–412. http://dx.doi.org/10.3758/BF03196896

Ball, B. H., DeWitt, M. R., Knight, J. B., & Hicks, J. L. (2014). Encoding and retrieval processes involved in the access of source information in the absence of item memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1271–1286. http://dx.doi.org/10.1037/a0037204

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning & Verbal Behavior, 22,* 88–104. http://dx.doi.org/10.1016/S0022-5371(83)80008-5

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science, 11,* 267–273. http://dx.doi.org/10.1111/1467-9280.00254

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11,* 211–227. http://dx.doi.org/10.3758/BF03196968

Batchelder, W. H., & Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review, 87,* 375–397. http://dx.doi.org/10.1037/0033-295X.87.4.375

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97,* 548–564. http://dx.doi.org/10.1037/0033-295X.97.4.548

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6,* 57–86. http://dx.doi.org/10.3758/BF03210812

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 197–215. http://dx.doi.org/10.1037/0278-7393.22.1.197

Bell, R., Mieth, L., & Buchner, A. (2017). Emotional memory: No source memory without old-new recognition. *Emotion, 17,* 120–130. http://dx.doi.org/10.1037/emo0000211

Bellezza, F. S. (1981). Mnemonic devices: Classification, characteristics, and criteria. *Review of Educational Research, 51,* 247–275. http://dx.doi.org/10.3102/00346543051002247

Bellezza, F. S. (1983). The spatial-arrangement mnemonic. *Journal of Educational Psychology, 75,* 830–837. http://dx.doi.org/10.1037/0022-0663.75.6.830

Bellezza, F. S. (1986). A mnemonic based on arranging words in visual patterns. *Journal of Educational Psychology, 78,* 217–224. http://dx.doi.org/10.1037/0022-0663.78.3.217

Bellezza, F. S. (1987). Mnemonic devices and memory schemas. In M. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes* (pp. 34–55). New York, NY: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4612-4676-3_2

Bellezza, F. S. (1996). Mnemonic methods to enhance storage and retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and cognition: Memory* (2nd ed., pp. 345–380). New York, NY: Academic Press. http://dx.doi.org/10.1016/B978-012102570-0/50012-4

Bellezza, F. S. (2009). Modeling guessing: The influence of multinomial models on the evolution of process-dissociation theory. *Zeitschrift für Psychologie/The Journal of Psychology, 217,* 125–135.

Bellezza, F. S., Elek, J. K., & Zhang, R. (2016). Source monitoring and associative structure. *Journal of Memory and Language, 87,* 144–156. http://dx.doi.org/10.1016/j.jml.2015.09.004

Bellezza, F. S., & Reddy, B. G. (1978). Mnemonic devices and natural memory. *Bulletin of the Psychonomic Society, 11,* 277–280. http://dx.doi.org/10.3758/BF03336829

Bower, G. H. (1972). Mental imagery and associative learning. In L. W. Gregg (Ed.), *Cognition in learning and memory* (pp. 51–88). Oxford, UK: Wiley.

Buchner, A., & Erdfelder, E. (1996). On assumptions of, relations between, and evaluations of some process dissociation measurement models. *Consciousness and Cognition: An International Journal, 5,* 581–594. http://dx.doi.org/10.1006/ccog.1996.0034

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General, 124,* 137–160. http://dx.doi.org/10.1037/0096-3445.124.2.137

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24,* 403–416. http://dx.doi.org/10.3758/BF03200930

Chen, H., & Wyble, B. (2015). The location but not the attributes of visual cues are automatically encoded into working memory. *Vision Research, 107,* 76–85. http://dx.doi.org/10.1016/j.visres.2014.11.010

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2006). Source memory in the absence of successful cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 828–835. http://dx.doi.org/10.1037/0278-7393.32.4.828

Efron, B., & Tibshirani, R. J. (1993). *Introduction to the bookstrap*. New York, NY: Chapman & Hall/CFC. http://dx.doi.org/10.1007/978-1-4899-4541-9

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. A review of the literature. *Zeitschrift für Psychologie/The Journal of Psychology, 217,* 108–124.

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 387–414. http://dx.doi.org/10.1037/0278-7393.24.2.387

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments & Computers, 28,* 1–11. http://dx.doi.org/10.3758/BF03203630

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*. Cambridge, MA: MIT Press.

Glanzer, M., Hilford, A., & Kim, K. (2004). Six regularities of source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1176–1195. http://dx.doi.org/10.1037/0278-7393.30.6.1176

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika, 59,* 21–47. http://dx.doi.org/10.1007/BF02294263

Humphreys, M. S., & Bowyer, P. A. (1981). Cued recall and cue recognition: How similar are the processes? *The American Journal of Psychology, 94,* 65–84. http://dx.doi.org/10.2307/1422343

Jones, G. B. (1976). A fragmentation hypothesis of memory: Cued recall of pictures and of sequential position. *Journal of Experimental Psychology: General, 105,* 277–293. http://dx.doi.org/10.1037/0096-3445.105.3.277

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Klauer, K. C., & Oberauer, K. (1995). Testing the mental model theory of propositional reasoning. *Quarterly Journal of Experimental Psychology, 48,* 671–687. http://dx.doi.org/10.1080/14640749508401410

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology, 75,* 1155–1178. http://dx.doi.org/10.1037/0022-3514.75.5.1155

Kurilla, B. P., & Westerman, D. L. (2010). Source memory for unidentified stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 398–410. http://dx.doi.org/10.1037/a0018279

Malejka, S., & Bröder, A. (2016). No source memory for unrecognized items when implicit feedback is avoided. *Memory & Cognition, 44,* 63–72. http://dx.doi.org/10.3758/s13421-015-0549-8

Meiser, T. (2014). Analyzing stochastic dependence of cognitive processes in multidimensional source recognition. *Experimental Psychology, 61,* 402–415. http://dx.doi.org/10.1027/1618-3169/a000261

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 116–137. http://dx.doi.org/10.1037/0278-7393.28.1.116

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135,* 638–677. http://dx.doi.org/10.1037/a0015849

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42,* 42–54. http://dx.doi.org/10.3758/BRM.42.1.42

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44,* 190–204. http://dx.doi.org/10.1006/jmps.1999.1283

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76,* 241–263. http://dx.doi.org/10.1037/h0027272

Paivio, A. (1971). *Imagery and verbal processes*. New York, NY: Holt, Rinehart, and Winston.

Pash, J. R., & Blick, K. A. (1970). The effect of a mnemonic device on retention of verbal material. *Psychonomic Science, 19,* 203–204. http://dx.doi.org/10.3758/BF03328775

Reddy, B. G., & Bellezza, F. S. (1983). Encoding specificity in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 167–174. http://dx.doi.org/10.1037/0278-7393.9.1.167

Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review, 5,* 597–614. http://dx.doi.org/10.3758/BF03208837

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95,* 318–339. http://dx.doi.org/10.1037/0033-295X.95.3.318

Riefer, D. M., & Batchelder, W. H. (1991). Statistical inference for multinomial processing tree models. In J. P. Doignon & J. C. Falmague (Eds.), *Mathematical psychology: Current developments* (pp. 313–335). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4613-9728-1_18

Ross, B. H., & Bower, G. H. (1981). Comparisons of models of associative recall. *Memory & Cognition, 9,* 1–16. http://dx.doi.org/10.3758/BF03196946

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal-detection model. *Psychological Review, 111,* 588–616. http://dx.doi.org/10.1037/0033-295X.111.3.588

Rowland, C. A., Littrell-Baez, M. K., Sensenig, A. E., & DeLosh, E. L. (2014). Testing effects in mixed- versus pure-list designs. *Memory & Cognition, 42,* 912–921. http://dx.doi.org/10.3758/s13421-014-0404-3

Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of memory* (pp. 375–447). New York, NY: Academic Press. http://dx.doi.org/10.1016/B978-0-12-521350-9.50017-6

Starns, J. J., & Hicks, J. L. (2008). Context attributes in memory are bound to item information, but not to one another. *Psychonomic Bulletin & Review, 15,* 309–314. http://dx.doi.org/10.3758/PBR.15.2.309

Starns, J. J., Hicks, J. L., Brown, N. L., & Martin, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition, 36,* 1–8. http://dx.doi.org/10.3758/MC.36.1.1

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition, 14,* 411–443. http://dx.doi.org/10.1080/13506280500195250

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. http://dx.doi.org/10.1016/0010-0285(80)90005-5

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373. http://dx.doi.org/10.1037/h0020071

Tversky, B. (2001). Spatial schemas in depictions. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 79–112). Cambridge, MA: The MIT Press.

Tversky, B., & Zacks, J. M. (2013). Event perception. In D. Riesberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 83–94). New York, NY: Oxford University Press.

Vandekerckhove, J., Matzke, D., & Wagenmakers, E. J. (2015). Model comparison and the principle of parsimony. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford handbook of computational and mathematical psychology* (pp. 300–318). New York, NY: Oxford University Press.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11,* 192–196. http://dx.doi.org/10.3758/BF03206482

Appendix

Frequencies of the Five Response Categories for the Unrelated and Related Pairs in the Five Experimental Condition Reported by Bellezza, Elek, and Zhang (2016)

	Response category					
Condition	Cue and response in correct locations	Cue and response in reversed locations	Only cue word in correct location	Only response word in correct location	Neither cue nor response in correct location	Total
Study 1, unrelated words in pairs						
Correct recall	603 (603.0)	86 (86.0)	38 (38.0)	44 (44.0)	86 (86.0)	—
Incorrect recall	195 (195.0)	98 (98.0)	221 (221.0)	165 (149.3)	432 (447.7)	1968
Study 2, unrelated words in pairs						
Correct recall	288 (288.0)	29 (29.0)	23 (23.0)	11 (11.0)	29 (29.0)	—
Incorrect recall	236 (236.0)	103 (103.0)	173 (173.0)	147 (155.0)	473 (465.0)	1512
Study 2, related words in pairs						
Correct recall	543 (543.0)	86 (86.0)	19 (23.2)	27 (27.0)	74 (69.8)	—
Incorrect recall	119 (119.0)	83 (83.0)	124 (124.0)	114 (109.3)	323 (327.7)	1512
Study 3, letter cues in pairs						
Correct recall	597 (597.0)	51 (51.0)	36 (36.0)	32 (32.0)	50 (50.0)	—
Incorrect recall	245 (245.0)	105 (105.0)	173 (173.0)	147 (163.8)	508 (491.2)	1944
Study 3, word cues in pairs						
Correct recall	771 (771.0)	172 (172.0)	63 (63.0)	80 (80.0)	187 (187.0)	—
Incorrect recall	133 (133.0)	76 (76.0)	100 (100.0)	90 (90.5)	272 (271.5)	1944

Note. The numbers in parentheses are the frequencies produced by the hybrid model. The values in the column "Cue and response in reversed locations" are not included in the column "Neither cue nor response in correct location."

Received January 11, 2017
Revision received April 30, 2018
Accepted May 5, 2018 ■