Journal of Memory and Language 63 (2010) 476–488

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Conjunction errors in recognition: Emergent structure and metacognitive control

Michael F. Verde

School of Psychology, University of Plymouth, Plymouth PL4 8AA, England, United Kingdom

ARTICLE INFO

Article history:
Received 7 August 2009
revision received 21 July 2010

Keywords:
Recognition
Memory
Conjunction errors
Metacognition
False memory

ABSTRACT

Pictures consisted of component halves from either the same photographic scene (SS) or different scenes (DS). SS pictures possessed the emergent property of forming a continuous scene. In Experiment 1, conjunction lures were created by rearranging halves from studied DS pictures into new DS or SS lures. Despite equally familiar components, conjunction errors were dramatically reduced for SS lures. In Experiment 2, judged frequency of single components within SS and DS probes was equivalent, ruling out differential context effects on component familiarity. In Experiment 3, strengthening the encoding of DS pictures only affected responses to DS probes, indicating separate decision rules for probe types. In Experiment 4, equal numbers of SS and DS pictures were studied, ruling out differential base rates as the source of the effect. The influence of emergent structure on conjunction errors was attributed to metacognitive decision rules.

Introduction

False memories often have their roots in real experiences. In everyday life, for example, we are prone to confusing the names and faces of recent acquaintances, or intermingling the details of similar past events. A form of composite memory studied widely in the laboratory is the *conjunction error*, the false recognition of a novel probe whose components were previously encountered in different contexts. In the investigation of conjunction errors, stimuli are units composed of a pair of elements. These units might be meaningless, arbitrary pairings of syllables or symbols. Alternatively, the elements might form meaningful wholes such as facial features that together form a face or morphemes that form compound words such as *earthquake* and *silkworm*. For the memory test, conjunction lures are created by taking elements from previously studied units and recombining them into novel units. If *earthquake* and *silkworm* were originally studied, falsely recognizing *earthworm* is an example of a conjunction error. Conjunction lures tend to be falsely recognized at much higher rates than items made up of completely novel elements.

Theoretical accounts based on the study of both item and pair recognition have generally taken an elemental approach, treating conjunction errors as simple aggregations of their component parts. Such an approach provides a useful basic theory that has been applied successfully to a wide range of data. The present study, however, demonstrates the limitations of an elemental approach by showing that emergent structure can have dramatic affects on the propensity to make conjunction errors. Gestalt principles have established the power of emergent structure to shape perceptual experience (Köhler, 1929). Higher-order structure, both perceptual and conceptual, is also known to have an important influence on learning (Ceraso, 1985; Köhler, 1929; Thorndike, 1931; Underwood & Schulz, 1960). The influence of structure that emerges only at retrieval is less well understood, particularly with regard to recognition.

The present investigation was directed at retrieval-based accounts of conjunction errors, and therefore the focus was on emergent structure at retrieval. It will be argued that a key to understanding the influence of emergent structure lies in metacognitive theories of decision-making.

E-mail address: michael.verde@plymouth.ac.uk

0749-596X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2010.08.002

Metacognition refers to the way that people use their beliefs about the characteristics and capabilities of cognitive processes to inform their judgments. In recognition, this can take the form of decision rules based on informative cues such as stimulus fluency and memorability. By altering these cues, emergent structure can alter the rules of recognition. Existing metacognitive theories, however, are less than clear in their predictions about the effect of emergent structure at retrieval. As will be shown, their application to the problem of conjunction errors also provides valuable insight into the complexity of decision strategies in recognition.

Conjunction errors: theory and data

According to defective binding theory, conjunction errors result from inappropriately binding unrelated elements during encoding. Conjunction errors are particularly high in populations known to have binding deficits such as patients with hippocampal damage (Cohen & Eichenbaum, 1993; Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996) and the elderly (Old & Naveh-Benjamin, 2008). They increase in normal participants when pressure is put on cognitive resources during encoding. For example, divided attention and the simultaneous processing of several items in working memory increase the likelihood of miscombining elements from different objects (Kroll et al., 1996; Reinitz & Hannigan, 2001, 2004). Although defective binding can explain some aspects of conjunction errors, other aspects are better accounted for by retrieval processes. Notably, binding errors fail to explain how false memories can come to incorporate novel elements in addition to previously studied ones (Jones, Jacoby, & Gellis, 2001). The decision factors investigated in the present study also require a retrieval-based explanation. For this reason, the remainder of the discussion focuses on theories that place the source of error in the retrieval stage.

Component familiarity theory is rooted in the dual-process view that two distinct retrieval processes operate during recognition (for review, see Yonelinas, 2002). The familiarity process quickly compares the retrieval cue to the contents of recent memory, producing evidence of even a partial match. Familiarity establishes the presence or absence of matching elements but provides no information about their association or context. The recollection process, on the other hand, is a slower, discriminative search for specific details of episodic association and context. According to the component familiarity account, memory conjunction errors are the product of the familiarity process. Suppose that the studied items *earthquake* and *silkworm* are stored in memory. When the lure *earthworm* is shown at test, its constituent components *earth* and *worm* feel familiar because each partially matches one of these memory representations. The feeling of familiarity that one gets from a probe is a monotonically increasing function of the number of previously encountered components. Consistent with this, conjunction lures containing two studied components (*earthworm*) produce more false alarms than lures containing one studied and one novel component (*ringworm*), which in turn produce more false alarms than lures containing two novel components (*ringtone*) (Jones et al., 2001; Lampinen, Odegard, & Neuschatz, 2004; Reinitz, Morrissey, & Demb, 1994; Rubin, Van Petten, Glisky, & Newberg, 1999).

While familiarity is a source of conjunction errors, recollection can reduce the error rate via a recall-to-reject strategy in which recalling one of the originally encoded items (*earthquake*) is used as a basis for rejecting the lure. Evidence for this strategy comes from manipulations of repetition and response deadlines. Study repetition is thought to increase both familiarity and recollection. Thus, additional repetitions both increase the familiarity of components but also enhance the ability to recollect the original items. Familiarity is a fast process, and the effect of repetition on familiarity is observed as an increase in the rate of conjunction errors when a short response deadline is imposed at test. At longer deadlines, however, the slower recollection process becomes available. Study repetition then has the opposite effect of decreasing conjunction errors as recollection is used to oppose false recognition based on familiarity. The use of a recall-to-reject strategy seems to be common with compound words (Jones, 2005; Jones & Jacoby, 2001; Lampinen et al., 2004) but not necessarily other types of materials (Jones & Bartlett, 2009).

When dealing with pairs of discrete objects, recognition can require discriminating studied pairs from pairs of nonstudied items (novel lures) or from pairs of studied items that have been rearranged into novel pairings (conjunction lures). Although the latter is commonly referred to as associative recognition, both types of discrimination will be referred to jointly as pair recognition. This allows parallels to be drawn between item and pair recognition which, despite superficial differences, have produced remarkably similar findings and theories with regard to conjunction errors. Studies of pair recognition typically find that conjunction lures produce more false alarms than novel lures (Gallo, Sullivan, Daffner, Schacter, & Budson, 2004; Healy, Light, & Chung, 2005; Kelley & Wixted, 2001; Light, Patterson, Chung, & Healy, 2004; Rhodes, Castel, & Jacoby, 2008). Lures containing one old and one novel item produce fewer false alarms than conjunction lures but more than novel lures (Castel & Craik, 2003; Cohn & Moscovitch, 2007). This pattern in which false alarms are a monotonically increasing function of the number of previously encountered components suggests that, as in item recognition, conjunction errors can be attributed to component familiarity. Moreover, just as in item recognition, recall-to-reject seems to be a common strategy used to reduce error rates. A frequent but puzzling observation in pair recognition is that the false alarm rate of conjunction lures can be unaffected by the number of times component items are repeated at study (Gallo et al., 2004; Kelley & Wixted, 2001; Light et al., 2004; Verde & Rotello, 2004). This can be explained by assuming that increasing repetition has two effects. It increases component familiarity, which increases the tendency toward false recognition, but it also increases the ability to recollect the original objects which can be used to reduce false recognition. These opposing tendencies can lead to no apparent effect of repetition on conjunction errors. Notably, repetition does increase false alarms to conjunction lures at short response deadlines when recollection is not yet available (Light et al., 2004).

Metacognitive decisions

A recognition judgment depends on two types of decisions: what should be used as evidence, and what level of evidence is sufficient to conclude that something has been previously encountered. Metacognitive beliefs can play a role in both types of decisions. The decision about what to use as evidence would seem straightforward: evidence is the information retrieved from memory. However, it has become clear that people also draw upon many sources of nonmnemonic information. They may do so because memory is often incomplete or ambiguous, is less accessible than other information, or because they have learned to trust certain cues as good indicators of past experience. A well-known example is perceptual fluency, the ability to perceive and process something quickly and easily. Prior experience aids perceptual identification (Jacoby & Dallas, 1981), and the corollary of this is that ease of perception may be a good indicator of prior experience. People seem to be aware of this relationship and commonly use fluency as a cue to infer familiarity. Evidence for this comes from the fact that people are prone to recognition errors when fluency is manipulated independently of prior experience. For example, Whittlesea, Jacoby, and Girard (1990) obscured recognition probes with varying degrees of visual noise and found that probes with a low degree of noise were not only more quickly identified but were more likely to be called ‘‘old.” Perceptual clarity has often been observed to enhance familiarity (Goldinger, Kleider, & Shelley, 1999; Kleider & Goldinger, 2004; Whittlesea & Leboe, 2003).

As important as the evidence itself is the criterion for what constitutes a sufficient level of evidence. A common assumption is that people base their recognition criterion on ease of discrimination: they impose a stricter standard of evidence when discrimination is easy but become more lenient when discrimination is difficult. These standards reflect a conservative or liberal bias respectively. How do people judge discriminability? A good deal of evidence suggests that people are guided by expectations of memorability. Encoding strength offers an obvious cue for memorability. Hirshman (1995; see also Stretch & Wixted, 1998a) manipulated encoding strength via study duration and repetition and found that participants responded more conservatively when study lists were stronger on average. Less direct cues of memorability have also been noted. Schacter, Israel, and Racine (1999) presented two groups with an auditory list. In one group, each word was accompanied by its written form. In the other group, each word was accompanied by its pictorial representation. Studying words with pictures improved memory but also induced a more conservative bias as evident in fewer false alarms. Schacter et al. proposed that participants in the picture group were less willing to recognize probes that failed to evoke the expected memories for visual detail. In other words, distinctive qualities can lead to expectations of stimulus memorability that influence later recognition criteria.

Related to criterion placement is the issue of criterion flexibility. Do people use a single criterion throughout a test or adjust their criterion depending on the nature of the probes? For example, if low and high discriminability targets are intermixed within a test, do people constantly shift their criterion depending on the type of probe? Such flexibility seems to occur under limited conditions. Studies manipulating encoding strength by varying study duration or repetition have typically observed the use of only a single criterion when strong and weak targets were tested together, even when the target types were differentiated using semantic or visual cues (Morrell, Gaitan, & Wixted, 2002; Stretch & Wixted, 1998b; Verde & Rotello, 2007). Singer and Wixted (2006) reported similar behavior when strength was manipulated by varying study-test delay, but only when the delays were relatively short. Very long (2-day) delays resulted in a higher rate of false alarms for the weak, delayed items. Although this was interpreted as a criterion shift, other investigators have suggested that increases in false recognition of semantically related lures (Singer and Wixted used category exemplars as targets and lures) over periods of 1–2 days is the result of individuating details decaying more rapidly than gist memory (Payne, Elie, Blackwell, & Neuschatz, 1996; Thapar & McDermot, 2001). This alternative explanation for the results in the delay condition illustrate the difficulty of knowing when a criterion shift has occurred when different classes of items potentially differ in terms of strength or quality of memory (e.g., Dobbins & Kroll, 1998). However, some studies have shown that within-list criterion shifts can occur for classes identical in memorability as long as there are strong, overt cues available (Singer, 2009; Verde & Rotello, 2007).

Emergent structure

Images of natural scenes were divided into halves and the half-images used as components to create a new set of pictures (Fig. 1). The new pictures were of two types: *same-scene* (SS) pictures consisted of halves from the same original scene, while *different-scene* (DS) pictures consisted of halves from two different scenes. The components of SS pictures together formed a single, continuous scene, and it was hypothesized that the emergent sense of ‘‘wholeness” would have psychological consequences. Stimulus unitization is known to benefit learning (for review, see Ceraso, 1985). For example, Asch, Ceraso, and Heimer (1960) showed that two elements, shape and outline texture, were better learned when presented as a single object than when presented side-by-side. Although participants in the present study were encouraged during encoding to treat each picture as a single image, it might be easier to bind SS components together as a unit which should improve memory for them. Also, SS pictures might be more easily encoded with verbal labels (e.g., ‘‘temple”) that provide an alternative route to retrieval. These potential effects of emergent structure on encoding were not of present interest.

The experiments were instead designed to examine the possibility that the sense of wholeness might influence decision-making at retrieval, with particular consequences for the false recognition of conjunction lures. Fig. 1 summarizes the basic stimuli and design of the experiments.

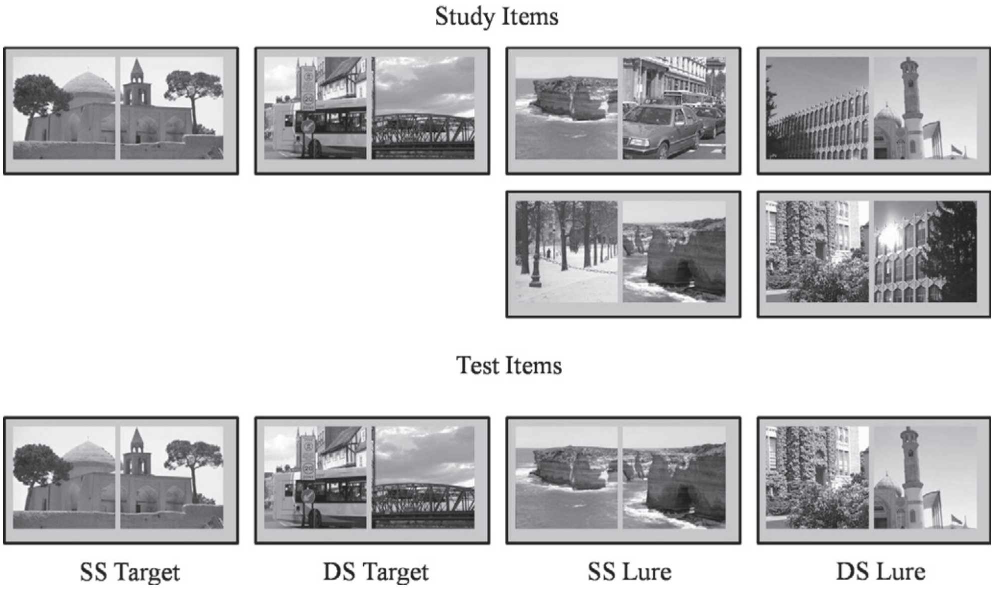

Fig. 1. Example same-scene (SS) and different-scene (DS) stimuli for the design used in Experiments 1, 3, and 4. Actual pictures were in color. (Images courtesy of the Ground Truth database (2009).)

Participants studied a mixture of SS and DS pictures. For the recognition test, lures were created by taking halves from some of the studied pictures and rearranging them in novel ways to create new pictures. A critical aspect of the design was that the components were always taken from DS pictures. This meant that the components of DS and SS conjunction lures had been encoded identically. The only thing that differentiated DS from SS lures was the manipulation of emergent structure at retrieval.

Controlling the encoding of lure components meant that any effect of emergent structure could be tied to retrieval processes or decision-making at retrieval. Three alternative hypotheses are suggested by the research reviewed earlier. The first, the component familiarity hypothesis, arises directly from the component familiarity theory of conjunction errors in item and pair recognition which predicts that picture familiarity will be an aggregate of the familiarity of the individual components (the picture halves). Because the halves were encoded identically, the components of DS and SS lures have equivalent familiarity and should therefore produce equal rates of false alarms. Component familiarity is essentially the null hypothesis that predicts no special effect of emergent structure on decision-making and on conjunction error rates. This basic prediction is potentially complicated by two issues, the role of recollection and encoding specificity, both of which will be considered in due course.

The fluency hypothesis arises from metacognitive theories about what people consider to be evidence of past experience. Intuitively, the "wholeness" of SS lures should confer a feeling of processing fluency. As noted earlier, fluency is often taken as evidence of familiarity. This might take the form of perceptual fluency. The integration of the components into a perceptually identifiable entity might proceed more quickly and easily when the components converge in a common direction. Alternatively, the common convergence might impart a sense of subjective congruity which itself can feel fluent, in the way that text in uniform font feels more fluent and familiar than text with letters in alternating font (Whittlesea & Leboe, 2003). SS pictures might also produce conceptual fluency in that the components converge on a coherent, semantically-identifiable unit. This might be likened to the way that words feel more fluent and familiar when embedded in semantically predictive sentences (Whittlesea, 1993, 2002). Whatever its source, the fluency hypothesis predicts that misattributed fluency will lead to higher rates of conjunction errors for SS lures. Indirect support for the fluency hypothesis can be found in a study by Greene and Tussing (2001), who examined the effect of inter-item similarity on the recognition of word pairs. They found that false alarms increased when a semantic relationship existed between the words of a pair, consistent with the notion that semantic coherence produces a sense of fluency. In one experiment (Experiment 4), participants were led to believe that they had seen a word list presented below the threshold of consciousness. The recognition test probes were in fact all novel pairs. Related pairs produced more false alarms than unrelated pairs even when no encoding was involved, demonstrating a fluency effect purely at retrieval.

The memorability hypothesis arises from metacognitive theories about how people choose their recognition criterion. People tend to set a higher standard of evidence when they expect materials to be more memorable. Previous studies of perceptual unitization and learning suggest that SS pictures should be objectively more memorable than DS pictures. If the emergent property of perceptual continuity causes people to believe that SS pictures are generally more memorable, and if this cue is powerful enough to encourage flexible criterion shifts within the test, conjunction error rates for SS and DS lures should differ. Given equally familiar components, a more conservative criterion for SS lures will result in lower rates of conjunction errors relative to DS lures.

Experiment 1

The purpose of Experiment 1 was to test the basic predictions of the component familiarity hypothesis, according to which there should be no effect of emergent structure and thus no difference in the way that DS and SS lures are judged. Metacognitive theories of decision-making offer two alternative hypotheses which do predict an effect of emergent structure. The fluency hypothesis predicts that SS lures will produce more errors. The memorability hypothesis predicts that DS lures will produce fewer errors. The fact that fluency and memorability lead to contradictory predictions is interesting in itself. Metacognitive decision rules have generally been studied in isolation to one another. The existence of circumstances in which different rules conflict underscores the need to understand how people coordinate the different decision strategies available to them.

Component familiarity theory is often placed within the larger framework of dual-process theory. The recollection process, which may contribute to recognition alongside familiarity, has the potential to complicate the predictions of the component familiarity hypothesis. For example, when faced with conjunction lures, people could adopt a recall-to-reject strategy in which they attempt to recall the original picture containing the lure components. Memory for the original picture would then serve as evidence to reject the lure. It should be noted, however, that in Experiment 1 SS and DS lure components were studied under identical conditions as parts of two different DS pictures. Therefore, recall-to-reject should be equally successful in both conditions and cannot be responsible for any differences in conjunction error rates.

Method

Participants

Twenty-seven undergraduates from the University of Plymouth participated for course credit.

Materials and design

A set of 204 digital color photographs of natural scenes were selected from the Ground Truth database (2009) and other royalty-free sources (additional photos were used for filler items). Each 160 × 320 pixel scene was divided into two 160 × 160 pixel halves. Each study and test picture consisted of two scene halves placed side-by-side (see Fig. 1), measuring 17.5 cm × 7.3 cm and subtending 16.3° of visual angle horizontally in the center of a 17″ flatscreen monitor. For the remainder of the discussion, the term *scene* will refer to one of the original photographs, while *picture* will refer to a single study or test stimulus composed of component halves.

Two study-test blocks were created, each with unique study and test lists. Each study list contained 48 pictures, with two additional fillers placed at the beginning and end of the list. One-fourth of the pictures (SS pictures) consisted of halves from the same photographic scene. The remainder (DS pictures) consisted of halves from different scenes. Both halves of a scene always appeared in the same study list, and no photos were reused across study lists. Each test list contained 48 studied targets and 48 conjunction lures, with two additional fillers placed at the beginning as practice trials. Conjunction lures were created by taking one half from each of two studied pictures and pairing them to form a new picture. The placement on the left or right side of the picture was preserved. Studied pairs were assigned to one of four conditions depending on their status at test. In the SS target condition, halves from the same scene were paired at study and again at test. In the DS target condition, halves from different scenes were paired at study and again at test. In the SS lure condition, halves from different scenes were paired at study, and the halves were rearranged at test to create new pictures with halves from the same scene. In the DS lure condition, halves from different scene scenes were paired at study, and the halves were rearranged at test to create new pictures with halves from different scenes. There were an equal number of probes from each condition in the test list. The assignment of pictures to condition and list-position was uniquely randomized for each participant. List creation, stimulus presentation, and response collection were computer-controlled.

Procedure

The session consisted of two study-test cycles. At the beginning of the study phase, participants were told that they would study a series of pictures for an upcoming memory test, and that each picture was made up of two halves, sometimes from the same scene and sometimes from different scenes. They were instructed to think of each one as a single picture, regardless of whether the halves "matched" or "mismatched." The pictures were then presented individually on the computer screen, each for 3500 ms followed by a 500 ms blank interval.

During the test phase, each trial began with a fixation line displayed in the center of the screen for 500 ms. This was replaced with the test probe, which remained until both responses were made. Participants first indicated whether the probe was an old picture from the study list or a new picture (using the/and z keys respectively). They then made a confidence rating on a 6-point scale, ranging from 1 = very sure new to 6 = very sure old (using the 1–6 keys at the top of the keyboard). A 1500 ms blank interval concluded the trial. On-screen prompts showing the mapping of keys to response categories appeared below the stimulus display area.

Results and discussion

Preliminary analysis revealed no significant effect of study-test cycle on the variables of interest in this or subsequent experiments, and this factor is not considered further. Because there were two recognition responses, a binary "old" vs. "new" judgment followed by a confidence rating, hit and false alarm rates could be calculated from either response. In this and subsequent experiments, the results from the two responses were virtually identical. Only performance based on confidence ratings was analyzed as these also produced the more detailed indexes of accuracy and bias (however, binary judgment performance is reported in Appendix Table A1 for comparison). Hit and false alarm rates were calculated from the confidence ratings by assigning the ratings 1–3 to the "new" category and the ratings 4–6 to the "old" category (Table 1). Analysis of variance (ANOVA) with item type (studied target vs. conjunction lure) and picture type (SS vs. DS) as within-subjects, repeated measures factors indicated significant main effects of item type, $F(1, 26) = 105.89$, $p < .001$, $\eta^2 = .42$, and picture type, $F(1, 26) = 28.03$, $p < .001$, $\eta^2 = .14$, as well as an interaction between item type and picture type, $F(1, 26) = 83.40$, $p < .001$, $\eta^2 = .17$. As would be expected, targets were more likely to be called "old" than lures. The hit rates for DS and SS targets did not differ statistically. However, the false alarm rate was significantly lower for SS lures, $t(26) = 3.38$, $p < .001$, $d = 2.29$. Finally, the confidence ratings were used to construct receiver operating characteristics (ROCs) for each participant and condition, and these in turn yielded indexes of accuracy, A_z, and bias, c_a (Macmillan & Creelman, 2005) (Table 2). Recognition of SS probes was significantly more accurate, $t(26) = 8.79$, $p < .001$, $d = 1.81$, and also showed a more conservative bias, $t(26) = 5.39$, $p < .001$, $d = 1.47$.

The results were consistent with the prediction of the memorability hypothesis that a stricter criterion of evidence would be applied to SS probes. This is seen in both the lower false alarm rate for SS lures and the higher value of the bias measure, c_a, for SS probes. The fact that error rates for SS and DS lures differed despite equivalent component familiarity also indicates that participants dynamically shifted their criterion in response to the perceptual nature of each picture. In other words, participants may have considered some types of stimuli to be more memorable than others. Pictures that are subjectively "whole" may be deemed more memorable, and this subjective memorability cannot be understand in terms of component familiarity because it is an emergent property of the picture components. Participants used this metacognitive insight to suppress false recognition of SS conjunction lures by adopting a conservative response bias.

Although the results quite clearly allow the rejection of the null hypothesis based on component familiarity theory, the failure to support the fluency hypothesis requires further examination. It could be that despite intuition, the emergent structure of SS pictures did not enhance processing fluency. An often used measure of fluency is response latency (e.g., Jacoby & Dallas, 1981). The latency of accurate responses to the initial binary "old" vs. "new" judgment is shown in Table 3. Hits were much faster for SS than for DS targets, $t(26) = 5.83$, $p < .001$, $d = 0.88$. Correct rejections were also faster for SS than for DS lures, $t(26) = 7.51$, $p < .001$, $d = 1.21$. If response latency is any indication, emergent structure did enhance the processing fluency of SS pictures. It is possible, of course, that the effect of fluency on familiarity was hidden by the opposing effect of memorability. The implications of this possibility are considered later. For the moment, the simple conclusion is that there was no apparent support for the fluency hypothesis.

Table 1
Recognition hit and false alarm rates.

	Studied targets				Conjunction lures			
	SS		DS		SS		DS	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Experiment 1	.67	.03	.65	.03	.15	.03	.53	.03
Experiment 3	.66	.04	.56	.03	.19	.03	.46	.04
Experiment 4	.61	.03	.61	.02	.18	.03	.50	.03

Notes: Picture components were from the same scene (SS) or different scenes (DS). Lure components had always been originally studied as part of a DS picture.

Table 2
Recognition accuracy (A_z) and bias (c_a).

	Accuracy (A_z)				Bias (c_a)			
	SS		DS		SS		DS	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Experiment 1	.78	.02	.59	.02	.30	.08	−0.22	.06
Experiment 3	.78	.03	.58	.02	.23	.07	−0.02	.06
Experiment 4	.76	.02	.59	.02	.31	.06	−0.13	.05

Notes: Picture components were from the same scene (SS) or different scenes (DS). Lure components had always been originally studied as part of a DS picture.

Experiment 2

A memorability heuristic can account for the lower rate of conjunction errors with SS pictures. That is to say, participants responded more strictly to "whole" pictures based on a metacognitive belief about the memorability of such items. However, the principle of encoding specificity offers an alternative, retrieval-based explanation. It is well established that recognition can be influenced by changes in study-test context (Light & Carter-Sobell, 1970; Tulving & Thomson, 1973), and this means that component familiarity may also depend on study-test context. Jones, Brown, and Atchley (2007) provided evidence for this in an experiment in which the morpheme components of conjunction lures (*crossbow*) were studied either as parts of other compound words (*crossroad*, *rainbow*) or as single words (*cross*, *bow*). Conjunction error rates were higher when the components originally appeared as parts of other compound words. The potential ef-

Table 3
Experiment 1: recognition latency (ms).

Hits (targets)				Correct rejections (lures)			
SS		DS		SS		DS	
M	SE	M	SE	M	SE	M	SE
2295	128	2930	148	2484	126	3452	178

Note: Response latency for accurate responses to binary "old" vs. "new" judgments.

fect of different encoding contexts on components, like that observed by Jones et al., was ruled out by design in the present study. SS and DS lure components were always studied as part of DS pictures to ensure that the encoding context was identical for both conditions. What differed was the test context. At one level, the two conditions were identical in that both SS and DS lures involved a change in study-test context for their components. However, it might be that there is some abstract quality that DS contexts share such that transfer from one DS context to another constitutes less of a contextual shift than transfer from a DS to an SS context. If this resulted in diminished familiarity of the components of SS lures, component familiarity theory would not necessarily be incompatible with the lower rates of conjunction errors observed in Experiment 1.

There is good reason to believe that judgments of studied frequency rely on the same mnemonic information as familiarity (Hintzman, Curran, & Caulton, 1995; Malmberg, Holden, & Shiffrin, 2004). In Experiment 2, just as in the previous experiment, all tested components were studied as parts of DS pictures. Each picture appeared once, twice, or four times in the list. The study procedure was identical in all other respects to that of Experiment 1. The presentation of test probes was also similar in that they were displayed as DS or SS pictures. However, each probe contained one studied component and one non-studied component (the latter was always taken from the same scene as a component that had appeared in the study list). During the test, the studied component was indicated and participants were asked to judge the frequency that it (and it alone) had appeared during study. Judged frequency was used as an indirect measure of single-component familiarity. Of interest was whether the perceptual context provided by the non-judged component would influence the familiarity of the judged component. If transfer from a DS to an SS picture constitutes a greater contextual shift than transfer from one DS picture to another, frequency judgments should be lower for SS than DS probes.

Method

Participants

Twenty-seven undergraduates from the University of Plymouth participated for course credit.

Materials and design

For each participant, a set of 90 scene photographs was randomly selected from the pool used in Experiment 1 (additional photos were used for filler items). The study list contained 45 unique pictures, each composed of two halves from different scene photos, with two additional fillers placed at the beginning and end of the list. Only one of the halves from each scene image appeared in the study list. Thirty pictures were designated as critical: 10 were repeated 1×, 10 were repeated 2×, and 10 were repeated 4× in the study list. The remaining 15 filler pictures were repeated x1. In total, the study list consisted of 85 pictures including repetitions. The test list consisted of 60 pictures, each created by taking one component half from a critical studied picture and pairing it with an unstudied component (the latter were always taken from scene photos of critical items). The placement of studied halves on the left or right side of the picture was preserved. Four additional fillers were placed at the beginning of the test as practice trials. Test items were divided evenly into two conditions. In the same scene (SP) condition, both halves of the picture were from the same scene photo. In the different scene (DP) condition, both halves were from different scene photos. For each test probe, a frequency judgment was made for the studied half only. The non-studied half was never judged but only served as a perceptual context. Study position (left or right half) and study repetition (1×, 2×, or 4×) were balanced between the two conditions. The assignment of pictures to condition and list-position was uniquely randomized for each participant.

Procedure

The instructions and procedure of the study phase were identical to Experiment 1. The test procedure was also similar save for the type of memory judgment. Each test trial began with a fixation line displayed in the center of the screen for 500 ms. A picture was then presented and the prompt "how many times?" appeared below either the left or the right half of the picture. Participants were to respond with the number of times that the indicated half had appeared in the study list using the number keys 0–9. The non-indicated half was to be ignored. A 1500 ms blank interval followed each response.

Results and discussion

Frequency judgments are shown in Table 4. ANOVA with study repetition (1×, 2×, 4×) and picture type (SS vs. DS) as repeated measures factors produced a main effect of repetition, $F(1, 26) = 160.15$, $p < .001$, $\eta^2 = .80$. Neither the main effect of picture type, $F(1, 26) = 2.08$, $p = .16$, nor the interaction between repetition and picture type, $F(1, 26) = 0.50$, $p = .61$, was statistically reliable. Frequency judgments clearly and rather accurately increased with the number of times the tested component appeared in the study list. However, whether this component appeared was tested as part of an SS or DS picture did not significantly influence the judgments. One can conclude, based on the assumption that familiarity and judgments of recent frequency draw on a common mnemonic source, that component familiarity is not more adversely

Table 4
Experiment 2: frequency judgments.

Probe type	Study presentations					
	1		2		4	
	M	*SE*	*M*	*SE*	*M*	*SE*
Same scene (SS)	1.06	.14	1.97	.14	3.19	.21
Different scene (DS)	1.13	.14	2.03	.15	3.40	.20

Note: Frequency judgment for the previously-studied component of the test probe, which was always originally studied as part of a DS picture. Probe type indicates whether the other, non-tested component of the test probe belonged to the same (SS) or different (DS) scene.

affected by transfer to an new SS context than a new DS context. This argues against the idea that context-driven changes in component familiarity were behind the reduced rate of conjunction errors for SS lures in Experiment 1.

Experiment 3

People tend to respond more conservatively when recognition is easier. However, they seem reluctant to shift their criterion dynamically in response to trial-to-trial changes in discriminability unless provided with salient cues. The findings of Experiment 1 suggests that picture continuity acted as such a cue, overriding component familiarity and leading participants to use different criteria for SS and DS probes. It could be that people ignore internal cues (component familiarity) when external cues (emergent structure) are available, or that they are generally disinclined to attend to more than a single cue. These possibilities were examined in Experiment 3 by adding a second manipulation of memorability alongside the manipulation of emergent structure. Strength manipulations such as study duration or repetition, which alter only the internal memory representation, are known to exert considerable influence on response bias (Hirshman, 1995; Stretch & Wixted, 1998a; Verde & Rotello, 2007). How participants deal with strength and emergent structure cues together would provide insight into the nature and sophistication of their decision strategies.

Experiment 3 was identical to Experiment 1 in all respects save for the strength of the study list. In the experiment, many of the DS pictures in the study list (specifically, those that would later be used to create DS targets and lures in the test list) were shown three times while the other pictures were shown only once. This increased the average strength of DS items and also increased the average strength of the list as a whole. If participants are fairly primitive in their ability to use multiple cues, attending to only one at a time, then there should be an effect of only strengthening or only emergent structure but not both. If this were the case, the experiment would reveal which of the cues has precedence, indicating which is the one that people consider most useful. Some theorists have proposed that cues have an order of priority (Strack & Bless, 1994). On the other hand, if people utilize multiple cues but fail to integrate them, the strength manipulation might be expected to produce only a general, non-specific bias. If this were the case, a comparison between Experiments 1 and 3 should reveal differences in the hit and false alarm rates for both SS and DS probes. However, the effect on SS probes is particularly diagnostic for this hypothesis. In both experiments, the pictures used to create SS probes were studied exactly once and should be similarly memorable. However, if the increase in average study list strength induced a general conservative bias in Experiment 3, then SS hit and false alarm rates should be lower in this experiment. A final possibility is that people are quite sophisticated in their ability to use multiple cues in an integrated fashion. If participants are able to use truly separate decision rules for SS and DS pictures, then the selective strengthening of DS pictures in the study list should only affect the recognition criterion for DS test probes. Thus, there should be no difference in the hit and false alarm rates for SS pictures in the two experiments. A conservative bias shift specifically for DS pictures should, however, result in lower DS hit and false alarm rates compared to Experiment 1.

Method

Participants

Twenty-seven undergraduates from the University of Plymouth participated for course credit.

Materials and design

These were identical to Experiment 1 save for one change in the construction of the study list. Pictures that were later tested as DS targets or whose components were rearranged into DS lures were presented 3× in the study list. The remaining pictures were shown only 1×. This increased the length of each study list from 48 to 96 pictures including repetitions.

Procedure

The procedure was identical to that of Experiment 1.

Results and discussion

Hit and false alarm rates were calculated from the confidence ratings (Table 1). ANOVA with item type (studied target vs. conjunction lure) and picture type (SP vs. DS) as within-subjects, repeated measures factors indicated statistically significant main effects of item type, $F(1, 26) = 53.35$, $p < .001$, $\eta^2 = .41$, and picture type, $F(1, 26) = 8.31$, $p = .008$, $\eta^2 = .04$, as well as an interaction between item type and picture type, $F(1, 26) = 50.92$, $p < .001$, $\eta^2 = .16$. Targets were more likely than lures to be called "old." The hit rate was higher for SS targets, $t(26) = 2.60$, $p = .015$, $d = .53$. The false alarm rate was much lower for SS lures, $t(26) = 6.61$, $p < .001$, $d = 1.62$. ROC analysis of confidence ratings was used to calculate indexes of accuracy, A_z, and bias, c_a (Table 2). Recognition of SS probes was significantly more accurate, $t(26) = 6.43$, $p < .001$, $d = 1.52$, and also showed a more conservative bias, $t(26) = 3.00$, $p = .006$, $d = 0.72$.

Replicating the critical finding of Experiment 1, conjunction error rates were considerably reduced for SS pictures. The results differed from those of Experiment 1 solely with respect to DS hit and false alarm rates, which were both lower in this experiment. Formal analysis consisted of comparing first SS recognition, and then DS recognition, between the experiments. Beginning with SS recognition, a 2 (item type: target vs. lure) × 2 (Experiment: 1 vs. 3) ANOVA produced the expected main effect of item type, $F(1, 26) = 207.67$, $p < .001$, $\eta^2 = .67$. There were no statistically reliable effects of Experiment or the interaction between Experiment and item type. In other words, there was no discernable difference in the recognition of SS pictures between the two experiments. The same could not be said for DS pictures. A 2 (item type) × 2 (Experiment) ANOVA produced the expected main effect of item type, $F(1, 26) = 18.65$, $p < .001$, $\eta^2 = .11$. There was also a main effect of Experiment, $F(1, 26) = 5.47$, $p = .023$, $\eta^2 = .06$, but no interaction between Experiment and item type. DS hit and false alarm rates were both lower in Experiment 3. The reduction was similar for both rates, consistent with a conservative shift in bias. This was confirmed by an analysis of A_z and c_a. There was no statistical difference in accuracy between the experiments. However, c_a indicated a more conservative bias in Experiment 3, $t(26) = 2.33$, $p = .024$, $d = 0.63$. The failure to observe an improvement in DS accuracy despite additional study repetitions might seem odd at first glance. However, repetition would be expected to increase the familiarity of both targets and lures, and this apparently resulted in no net change in accuracy. The expected tendency of increased familiarity to also increase hit and false alarm rates was also apparently completely overshadowed by the opposite trend of a conservative criterion shift.

The fact that the selective strengthening of DS pictures in the study list only affected responding to DS test probes suggests that participants were able to integrate their use of strength and perceptual continuity cues. Emergent structure allowed the separation of pictures into two memorability classes. Participants used different decision rules for each class, applying them in a dynamic fashion depending on the nature of each recognition probe. The decision rules were sophisticated enough to account for not only the perceptual memorability but also the average mnemonic strength of each class. This last point offers an important insight into the nature of the memorability heuristic. The heuristic does no consist solely of simple, static rules about the memorability of various classes of objects. The rules are adaptive and draw on multiple sources of information.

Experiment 4

The purpose of Experiment 4 was to address two remaining alternatives to the memorability hypothesis used to account for the findings so far. The first has to do with the fact that there were three times as many DS pictures as SS pictures in the study lists of Experiments 1 and 3. This allows for two alternative decision rules that could have resulted in the reduction of SS conjunction errors. The first is a variant of the memorability hypothesis. The decision that SS and DS pictures differed in memorability may have been based not on perceptual characteristics but rather on the relative rarity of the two types of pictures. Rare things are more salient, leading people to think of them as more memorable (Strack & Bless, 1994). A second possibility is that the decision rule had little to do with subjective memorability at all. Information about the base rates of targets and lures in a recognition test is known to influence response bias (Ratcliff, Sheu, & Gronlund, 1992; Strack & Förster, 1995). The smaller proportion of SS pictures in the study list may have led participants to believe that there was also a smaller proportion of SS targets in the test list. This mistaken belief would have encouraged a more conservative bias for SS probes. In order to rule out these alternative explanations, the study lists in Experiment 4 included an equal number of SS and DS pictures.

As noted earlier, the recall-to-reject strategy of recalling the studied pictures in which the lure components originally appeared should be equally successful for SS and DS lures. However, an alternative recall-to-reject strategy is possible. If one were to recall all of the studied SS pictures, the absence of the SS lure among them would be grounds for rejection. The study lists of Experiments 1 and 3 contained 12 SS pictures. The success of this strategy would require recalling all or most of these. It is doubtful that this happened; if participants were capable of this level of recall, it is puzzling that SS hit rates were not higher than they were. Nonetheless, the number of SS pictures in Experiment 4 was increased to 36, a number large enough to prevent the success of, and likely deter any attempt to use, a recall-all-SS strategy.

Method

Participants

Twenty-seven undergraduates from the University of Plymouth participated for course credit.

Materials and design

These were identical to Experiment 1 save for one change in the construction of the study list. An additional 24 SS pictures were added to the study list so that the number of SS and DS pictures was the same. This increased the length of each study list from 48 to 64 pictures.

Procedure

The procedure was identical to that of Experiment 1.

Results and discussion

Hit and false alarm rates were calculated from the confidence ratings (Table 1). ANOVA with item type (studied target vs. conjunction lure) and picture type (SS vs. DS) as within-subjects, repeated measures factors indicated statistically significant main effects of item type, $F(1, 26) = 126.79$, $p < .001$, $\eta^2 = .43$, and picture type, $F(1, 26) = 39.96$, $p < .001$, $\eta^2 = .14$, as well as an interaction between item type and picture type, $F(1, 26) = 42.70$, $p < .001$, $\eta^2 = .16$. Targets were more likely to be called "old" than lures. The hit rates for DS and SS targets did not differ statistically. However, the false alarm rate was significantly lower for SS lures, $t(26) = 8.56$, $p < .001$, $d = 2.36$. Confidence ratings were used to calculate indexes of accuracy, A_z, and bias, c_a (Table 2). Recognition of SS probes was significantly more accurate, $t(26) = 6.73$, $p < .001$, $d = 1.60$, and also showed a more conservative bias, $t(26) = 6.43$, $p < .001$, $d = 1.59$.

Replicating previous results, conjunction error rates were considerably reduced for SS pictures. However, performance was similar to Experiment 1. This was examined formally by comparing first SS recognition, and then DS recognition, between Experiments 1 and 4. Comparisons of hit and false alarm rates, accuracy, and bias revealed no significant differences between the experiments. Two conclusions can be drawn from these results. First, the reduction of SS conjunction errors despite the equal number of SS pictures in the study list rules out decision rules effects based on the rarity of SS pictures. Second, the large number of SS pictures in the study list make it extremely unlikely that participants could have used a strategy of recalling all studied SS pictures to reject SS lures.

General discussion

Component familiarity theory has proven very successful in explaining the phenomenon of conjunction errors in item recognition (Jones & Jacoby, 2001; Jones et al., 2001). The approach has been equally influential in theories of pair recognition (Castel & Craik, 2003; Kelley & Wixted, 2001). The present study suggests that while component familiarity may provide a good basic understanding of the phenomenon, a more complex view must consider that objects are often more than the sum their parts. Components can take on emergent properties that dramatically affect the recognition of a object. This was observed here when pictures composed of two components belonging to a single, continuous scene produced much lower rates of false recognition than pictures composed of unrelated components despite the fact that the components were equally familiar in both cases. A similar point was recently made by McKone and Peh (2006) in their study of conjunction errors in faces. They observed that studied faces were recognized more often than rearranged-feature lures when the faces were shown upright, but studied and rearranged faces were equally likely to be called "old" when shown upside-down. McKone and Peh reasoned that the strongly configural nature of facial processing occurred only when faces were viewed normally; only component processing occurred when faces were inverted. Their findings illustrate the mediating role of perception in recognition. The present study expands on this point in two important respects. First, the configural processing of faces is known to be supported by specialized brain areas (Maurer, Le Grand, & Mondloch, 2002). The use here of scenic pictures emphasizes the point that effects of emergent structure are not domain specific. Second, the ability of emergent structure to support discrimination between targets and lures, as observed by McKone and Peh, is likely due to encoding (the ability to store holistic configurations) or retrieval (the ability to rely on recollection more than familiarity). The dissociation in false recognition of different types of lures, as observed here, points to a different source of influence, one residing in post-retrieval decision processes.

Recognition judgments depend fundamentally on how people choose to interpret retrieved information. These choices are shaped by metacognitive beliefs. For example, expectations about stimulus memorability help to determine recognition criteria. When items are thought to be particularly memorable, people demand a higher standard of evidence before accepting something as "old." Strengthening memory by repetition or increased duration at encoding leads to more conservative responding (Hirshman, 1995; Singer, 2009; Stretch & Wixted, 1998b; Verde & Rotello, 2007). Improving the encoded quality of memory with more elaborate processing or rich episodic details has a similar effect (Bodner & Lindsay, 2003; Israel & Schacter, 1997; Schacter et al., 1999). The present findings show that the perceptual continuity that emerges when picture components belong to a common scene is yet another cue of memorability. The subjective link between picture continuity and memorability may be based on a direct evaluation of the memory representations of SS pictures. Alternatively, perceptual continuity may tap into a general belief that perceptually fluent objects are more memorable. People have been shown to be more confident in their future ability to recall words when they are displayed in larger font (Rhodes & Castel, 2008) or spoken more loudly (Rhodes & Castel, 2009). The data offer no way of distinguishing between these possibilities. The important point, however, is that the effect of perceptual continuity is clearly similar to that of other memorability cues.

Knowing the sorts of cues people use to judge memorability is an essential step toward understanding their metacognitive beliefs about recognition. A further step is to determine how people deal with the variety of potential cues available in the environment. The present findings contribute to the sparse body of evidence relevant to this question. The results of Experiment 3 indicate that the ability to integrate multiple cues can be quite sophisticated. Emergent structure and study repetition are both informative about memorability. Participants not only attended to both but also took into account that these cues differed for the two classes of stimuli. SS pictures were studied in the same manner in Experiments 1 and 3. Recognition of SS probes did not differ significantly across the experiments, suggesting that participants applied the same decision rule, one driven by the perceptual properties of that class of items. On the other hand, DS pictures were repeated many more times on average in Experiment 3. This led to a more conservative bias in Experiment 3 that was specific to DS probes and reflected an understanding of their internal strength. In sum, participants used one decision rule for SS probes and another one for DS probes. The fact that this produced adaptive, dynamic criterion-setting within the test is of interest given that people seem reluctant to make trial-to-trial criterion shifts in the absence of particularly salient cues. That emergent structure was able to provoke such behavior speaks to its power as a cue.

The lack of support for the fluency hypothesis was interesting on more than one level. Empirically, the findings seem to defy expectations arising from the large body of evidence that perceptual and conceptual fluency increases feelings of familiarity. One possibility is that, despite intuition, perceptual continuity simply did not enhance fluency. The finding of faster recognition times for SS probes in Experiment 1 argues against this. Identification speed has often been considered an indicator of processing fluency (Jacoby & Dallas, 1981). A second possibility is that enhanced fluency did increase familiarity slightly but the effect was overshadowed by a conservative bias which pushed hit and false alarm rates in the opposite direction. A small, unobservable effect on familiarity seems puzzling given such a dramatic effect on response latency: accurate responses were over 600 ms faster to SS than to DS targets, and nearly 1000 ms faster to SS lures than to DS lures. That the resulting effect on familiarity was so easily overshadowed implies that fluency must be a relatively minor cue—a surprising implication given the robustness of fluency effects observed in other studies. A third possibility is that the nature of the stimuli, being pairs of component items, was not conducive to fluency effects. It is true that most such effects have been observed with single-item stimuli. However, Greene and Tussing (2001), using an associative recognition task, found that a semantic relationship between the two words of a pair produced elevated rates of false alarms just as predicted by fluency. The critical difference between their stimuli and the pictures used here is that words of a pair, even if related, are still likely viewed as two discrete units rather than a single, emergent unit.

A final possibility, the one that seems most plausible, is that participants disregarded fluency in favor of memorability. This is interesting on a theoretical level because it offers insight into the status of fluency as a cue. Two explanations have been offered in the literature for why people sometimes disregard fluency cues. The first is that they identify fluency as coming from a source other than memory (Jacoby & Whitehouse, 1989; Whittlesea et al., 1990). For example, Whittlesea (2002) found that the fluency effect created by embedding a target word in a semantically predictive sentence disappeared when the connection between the word and the sentence was made particularly obvious. This implies that participants understood that a sentence context could create a sense of fluency irrelevant to their recognition judgment. In the present study, participants could have been similarly aware that processing fluency arising from perceptual continuity could be misattributed to familiarity. Although possible, this would indicate a surprisingly high degree of metacognitive awareness.

Another reason for disregarding fluency is offered by Strack and Bless (1994), who proposed that a hierarchy of priority exists for various response strategies. When stimuli are deemed to be very memorable, people adopt a strategy of rejecting probes that fail to evoke distinct memories. When stimuli are not memorable, they recognize that such a strategy is ineffective and rely instead on other inferential cues. In their study, Strack and Bless presented items belonging to a distinctive and non-distinctive category, and also presented recognition lures using the definite article "the" or the indefinite article "a." Because the definite article more strongly suggests a specific encounter, it is known to bias people toward accepting misinformation. Using the definite article did increase false recognition of lures from the non-distinctive category. This did not happen with lures from the distinctive category, presumably because participants relied instead on memorability cues. Although aspects of their findings might be interpreted in terms of retrieval rather than decision processes (Rotello, 1999), the cue priority hypothesis proposed by Strack and Bless remains appealing. Some cues must be more reliable or diagnostic than others, and a reasonable strategy would be to ignore the latter when the former are available. One can interpret the present findings in these terms: despite their ready use of fluency cues, people understand them to be unreliable. They have much more confidence in the diagnosticity of memorability cues and are happy to rely on them exclusively. Although this offers only a basic insight into cue prioritization, the mere fact that people use such a strategy suggests some depth to their metacognitive knowledge.

Global or selective bias?

The memorability heuristic has been described here in terms of a global criterion shift, a general tendency to respond more conservatively to items of a particularly memorable class. However, bias resulting from beliefs about memorability is sometimes discussed in the literature in a somewhat different way. A representative example is the *distinctiveness heuristic*, according to which people reject lures that do not possess distinctive details expected of studied items. Schacter et al. (1999) first used this term in a study which two groups were given lists of semantic associates. The word group heard each word and saw its written form, while the picture group heard each word and saw its pictorial form. Recognition probes were written words that included lures semantically related and unrelated to targets. Both groups showed similar hit rates, but false alarm rates to both types of lures were lower in the picture group. If picture encoding improved memory for studied words, the pattern can be explained by assuming that the picture group adopted a more conservative criterion. Schacter et al. suggested that the picture group expected targets to possess visual imagery details and used the distinctiveness heuristic to reject probes lacking in these details. Such an explanation might be thought of as a more specific version of the rule that people adopt a more conservative criterion when studied items are more memorable, and indeed the data could be explained using this more general rule.

On the other hand, the distinctiveness heuristic is sometimes described not as a general bias but one that selectively affects lures (Dodson & Schacter, 2002). If this is true, one then needs to make additional assumptions to explain how a rule could affect only lures and not targets. For example, one might suggest that in the Schacter et al. (1999) study, the picture group adopted a strong rule by which probes were called "old" only when they evoked pictorial details. Such a rule does not explain the presence of false alarms. One could make the additional assumption that some lures falsely evoked details of imagery. Alternatively, one might suggest that the picture group simultaneously used two different rules. On some proportion of trials, probes were accepted based on familiarity alone, resulting in some false alarms. On the remaining trials, probes were only accepted if they evoked imagery details. However, because some targets would also likely fail to evoke imagery and be rejected, this dual-rule strategy predicts a decrease in hits relative to a single rule, familiarity-only strategy. This necessitates an additional assumption that picture encoding also improved target recognition, bringing hit rates back to the same level as the word group.

Similar selective bias rules might be applied to the present findings. Perhaps memory representations of SS pictures possessed some distinctive feature that differentiated them from DS pictures. When faced with an SS probe, participants could have responded "old" only if this feature was recovered from memory. SS lures could have sometimes spuriously evoked memory for this feature, producing false alarms. Alternatively, participants may have used familiarity on some SS trials and the distinctive feature rule on other trials. The expected decrease in hits could have been balanced by the superior memorability of SS targets. It is difficult to rule out these or similar selective bias rules, not only because they are capable of producing results similar to a global bias but also because formal models of selective bias have yet to be detailed. The simplest explanation for the present results, however, is the one proposed: a conservative criterion shift for SS targets. In any case, the question of global or selective bias is independent of the main conclusion of the study, which is that emergent structure can have a profound effect on the tendency to make conjunction errors.

A. Appendix

Table A1. Recognition hit and false alarm rates from binary judgments.

	Studied targets				Conjunction lures			
	SS		DS		SS		DS	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Experiment 1	.66	.04	.64	.03	.16	.04	.53	.03
Experiment 3	.66	.04	.56	.03	.18	.03	.45	.04
Experiment 4	.62	.03	.61	.02	.17	.02	.49	.03

Notes: On each trial, a binary ("old" vs. "new") judgment preceded the confidence rating. For all experiments, hit and false alarm rates were virtually identical to those derived from confidence ratings (Table 1) and are presented here only for reference.

References

Asch, S. E., Ceraso, J., & Heimer, W. (1960). Perceptual conditions of association. *Psychological Monographs,* 74 (3, No. 490).

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language, 48*, 563–580.

Castel, A. D., & Craik, F. I. M. (2003). The effects of aging and divided attention on memory for item and associative information. *Psychology and Aging, 18*, 873–885.

Ceraso, J. (1985). Unit formation in perception and memory. In G. H. Bower (Ed.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 19, pp. 179–210). New York: Academic Press.

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampus*. Cambridge, MA: MIT Press.

Cohn, M., & Moscovitch, M. (2007). Dissociating measures of associative memory: Evidence and theoretical implications. *Journal of Memory and Language, 57*, 437–454.

Dobbins, I. G., & Kroll, N. E. A. (1998). Distinctiveness and the recognition mirror effect: Evidence for an item-based criterion placement heuristic. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*, 1186–1198.

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

Gallo, D. A., Sullivan, A. L., Daffner, K. R., Schacter, D. L., & Budson, A. E. (2004). Associative recognition in Alzheimer's disease: Evidence for impaired recall-to-reject. *Neuropsychology, 18*, 556–563.

Goldinger, S. D., Kleider, H. M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory & Cognition, 27*, 328–338.

Greene, R. L., & Tussing, A. A. (2001). Similarity and associative recognition. *Journal of Memory and Language, 45*, 573–584.

Ground Truth database. (2009). <http://www.cs.washington.edu/research/imagedatabase/> Retrieved 01.03.09.

Healy, M. R., Light, L. L., & Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 768–788.

Hintzman, D. L., Curran, T., & Caulton, D. A. (1995). Scaling the episodic familiarities of pictures and words. *Psychological Science, 6*, 308–313.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 302–313.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review, 4*, 577–581.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118*, 126–135.

Jones, T. C. (2005). Study repetition and the rejection of conjunction lures. *Memory, 13*, 499–515.

Jones, T. C., & Bartlett, J. C. (2009). When false recognition is out of control: The case of facial conjunctions. *Memory & Cognition, 37*, 143–157.

Jones, T. C., Brown, A. S., & Atchley, P. (2007). Feature and conjunction effects in recognition memory: Toward specifying familiarity for compound words. *Memory & Cognition, 35*, 984–998.

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45*, 82–102.

Jones, T. C., Jacoby, L. L., & Gellis, L. A. (2001). Cross-modal feature and conjunction errors in recognition memory. *Journal of Memory and Language, 44*, 131–152.

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 701–722.

Kleider, H. M., & Goldinger, S. D. (2004). Illusions of face memory: Clarity breeds familiarity. *Journal of Memory and Language, 50*, 196–211.

Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.

Kroll, N. E. A., Knight, R. T., Metcalfe, J., Wolf, E. S., & Tulving, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory and Language, 35*, 176–196.

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

Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 1–11.

Light, L. L., Patterson, M. M., Chung, C., & Healy, M. R. (2004). Effects of repetition and response deadline on associative recognition in young and older adults. *Memory & Cognition, 32*, 1182–1193.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old–new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 319–331.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255–260.

McKone, E., & Peh, Y. X. (2006). Memory conjunction errors for realistic faces are consistent with configural processing. *Psychonomic Bulletin & Review, 13*, 106–111.

Morrell, H. E. R., Gaitan, S., & Wixted, J. T. (2002). On the nature of the decision axis in signal-detection-based models of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1095–1110.

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging, 23*, 104–118.

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

Ratcliff, R., Sheu, C., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99*, 518–535.

Reinitz, M. T., & Hannigan, S. L. (2001). Effects of simultaneous presentation and attention switching on memory conjunction errors. *Journal of Memory and Language, 44*, 206–219.

Reinitz, M. T., & Hannigan, S. L. (2004). False memories for compound words: Role of working memory. *Memory & Cognition, 32*, 463–473.

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 161–168.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*, 615–625.

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review, 16*, 550–554.

Rhodes, M. G., Castel, A. D., & Jacoby, L. L. (2008). Associative recognition of face pairs by younger and older adults: The role of familiarity-based processing. *Psychology and Aging, 23*, 239–249.

Rotello, C. M. (1999). Metacognition and memory for nonoccurrence. *Memory, 7*, 43–63.

Rubin, S. R., Van Petten, C., Glisky, E. L., & Newberg, W. M. (1999). Memory conjunction errors in younger and older adults: Event-related potential and neuropsychological data. *Cognitive Neuropsychology, 16*, 459–488.

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

Singer, M. (2009). Strength-based criterion shifts in recognition memory. *Memory & Cognition, 37*, 976–984.

Singer, M., & Wixted, J. T. (2006). Effect of delay on recognition decisions: Evidence for a criterion shift. *Memory & Cognition, 34*, 125–137.

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language, 33*, 203–217.

Strack, F., & Förster, J. (1995). Reporting recollective experiences: Direct access to memory systems? *Psychological Science, 6*, 352–358.

Stretch, V., & Wixted, J. T. (1998a). Decision rules for recognition memory confidence judgements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1397–1410.

Stretch, V., & Wixted, J. T. (1998b). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1379–1396.

Thapar, A., & McDermot, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition, 29*, 424–432.

Thorndike, E. L. (1931). *Human learning*. New York: Century.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Underwood, B. J., & Schulz, R. W. (1960). *Meaningfulness and verbal learning*. Chicago: Lippincott.

Verde, M. F., & Rotello, C. M. (2004). ROC curves show that the revelation effect is not a single phenomenon. *Psychonomic Bulletin & Review, 11*, 560–566.

Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition, 35*, 254–262.

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1235–1253.

Whittlesea, B. W. A. (2002). Two routes to remembering (and another to remembering not). *Journal of Experimental Psychology: General, 131*, 325–348.

Whittlesea, B. W., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29*, 716–732.

Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *The Journal of Memory and Language, 49*, 62–79.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441–517.