# The role of stimulus type in list length effects in recognition memory

Angela Kinnell · Simon Dennis

**Abstract** While many studies have investigated the list length effect in recognition memory, few have done so with stimuli other than words. This article presents the results of four list length experiments that involved word pairs, faces, fractals, and photographs of scenes as the stimuli. A significant list length effect was identified when faces and fractals were the stimuli, but the effect was nonsignificant when the stimuli were word pairs or photographs of scenes. These findings suggest that the intrastimulus similarity is what dictates whether list length has a significant effect on recognition performance. As is the case with words, word pairs and photographs of scenes are not sufficiently similar to generate detectable item interference.

**Keywords** Recognition · List length effect

The list length effect is the finding that recognition performance is superior for items that are part of a short list at study than for items that were part of a long list. The issue of whether there is a genuine list length effect in recognition memory or whether this is a finding that has resulted from the influence of confounding variables in past studies is critical, because the list length effect can provide a test of models of recognition memory. There are two main

A. Kinnell (✉)
School of Psychology, University of Adelaide,
Adelaide, South Australia, Australia
e-mail: angela.kinnell@gmail.com

A. Kinnell
Division of Business, University of South Australia,
Adelaide, South Australia, Australia

S. Dennis
Department of Psychology, Ohio State University,
Columbus, OH, USA

groups of recognition memory models: item noise models and context noise models. *Item noise models* posit that interference in recognition memory originates from the other items presented with the test cue at study. In *context noise models*, there is no interference from other list items, but rather, from the previous contexts in which the test item has been seen. Consequently, item noise models predict a significant effect of list length on recognition performance, but context noise models do not. Thus, investigating the nature of the list length effect is critical to evaluating models of recognition memory.

The list length effect finding has been well replicated in the recognition memory literature (e.g., Bowles & Glanzer, 1983; Cary & Reder, 2003; Gronlund & Elam, 1994; Murnane & Shiffrin, 1991; Shiffrin, Ratcliff, Murnane, & Nobel, 1993; Strong, 1912; Underwood, 1978). However, in recent years, a number of studies have not identified a significant effect of list length on recognition performance (e.g., Buratto & Lamberts, 2008; Dennis & Humphreys, 2001; Dennis, Lee, & Kinnell, 2008; Jang & Huber, 2008; Kinnell & Dennis, 2011). Dennis and Humphreys argued that the studies that had identified a significant list length effect had done so because of a failure to adequately control for a number of potential confounds that could lead to a spurious effect. These confounds were retention interval, attention, displaced rehearsal, and contextual reinstatement.

Retention interval is related to the amount of time that passes between when an item is studied and tested. Typically, short-list study items have a shorter retention interval, but this can be lengthened with a period of filler activity, so that the retention interval is equivalent to that with a long list. The filler activity can come either before (proactive design) or after (retroactive design) the short list. In addition, the long-list test items come from either the end (proactive design) or the beginning (retroactive design) of

the long study list, corresponding to whether the filler for the short list came before or after the presentation of the items (Cary & Reder, 2003; Dennis & Humphreys, 2001; see Fig. 1). The retroactive design control for retention interval can also lead to differential rehearsal of list items, with the filler period allowing for rehearsal of short-list items, with no equivalent opportunity for long-list items. Making the filler activity as engaging as possible in an attempt to discourage rehearsal can help to counteract this difference (Cary & Reder, 2003; Dennis & Humphreys, 2001).

The amount of attention paid to list items is also likely to differ between short and long lists, with boredom more likely to occur in the latter (Underwood, 1978). When the proactive design has been used to control for retention interval and the items at the end of the long study list are being tested, differences in performance based on list length may be magnified. This potential confound can be controlled by using an encoding task that requires a response during study (Cary & Reder, 2003; Dennis & Humphreys, 2001).

Finally, there can also be differences between short and long lists in the reinstatement of the study context at the onset of the respective test lists. This is of particular concern when the retroactive design has been used as a control for retention interval and the filler activity immediately precedes the short-list test, but the long list is immediately followed by its test. This difference may mean that participants are more inclined to reinstate the short-study-list context at test, having just been engaged in a completely different activity, while they may see no such need in the long-list test, having just viewed the list itself. The study-list context may vary with the passage of time, particularly for the long list. Consequently, the end-of-list context may be significantly different from the start-of-list context, which, in the retroactive design, is associated with the items on which they will be tested. This may lead to a spurious list length effect finding. However, these differences can be controlled by presenting participants with an

additional period of filler activity before presentation of each test list (Cary & Reder, 2003; Dennis & Humphreys, 2001; see Fig. 2).

Kinnell and Dennis (2011) reviewed the relative effects of these four potential confounds and concluded that, while there may be no way to completely control for their effects, steps can be taken to minimise the influence of the potential confounds. They recommended the use of the retroactive design along with a pleasantness rating task at study, together with an extended (8-min) period of filler activity prior to the onset of each test list, as the best way to limit the effect of the potential confounds. Under these conditions, they found no effect of list length using words as stimuli.

## The list length effect and stimuli other than words

Many studies have investigated the list length effect in recognition memory, but almost all of these studies have used words as the stimuli. The exceptions include the very first list length study, that of Strong in 1912, which used newspaper advertisements; the work of Brandt (2007), which involved random checkerboard patterns; and the associative recognition experiments of Criss and Shiffrin (2004b), which included faces; all other published list length studies to date have used words as the stimuli. Both Strong and Criss and Shiffrin (2004b) found significant differences between list lengths for their advertisement and face pair stimuli, and Brandt found that while short-list performance was significantly better than long-list performance, long-list performance was actually superior to that with a medium-length list (though not significantly). Strong did not control for any of the four potential confounds noted by Dennis and Humphreys (2001), Brandt's study included a short (30-s) period of filler prior to each test list, and Criss and Shiffrin's (2004b) experiment included controls for retention interval, displaced rehearsal, and attention. Thus, there has not been a fully controlled list length experiment using a stimulus other than words. Therefore, it remains unclear whether the nonsignificant effect of list length on recognition performance for words when potential confounds
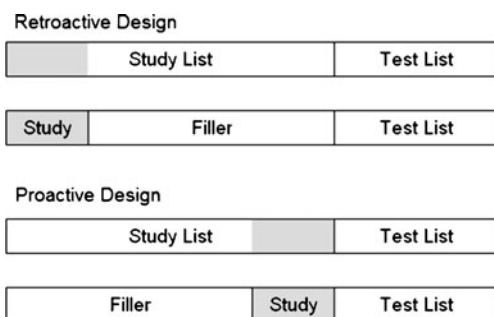


Fig. 1 Experimental design when controls for retention interval are implemented. The shading indicates the sections of the study lists that are tested
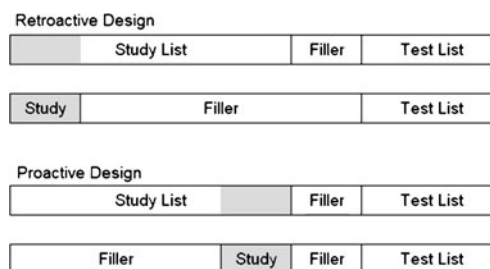


Fig. 2 Experimental design when controls for retention interval and contextual reinstatement are implemented. The shading indicates the sections of the study lists that are tested

are controlled is specific to these stimuli. Dennis and Humphreys conjectured that there may be something special about words: They are unitised stimuli and, because they are very commonly encountered, they have distinctive representations and are not easily confused with each other. As Greene (2004, p. 261) noted, "all words contain strong semantic features that make each one unique and distinctive." It may be that other, less unitised, stimuli behave differently. Thus, the aim of the experiments presented in this study was to examine whether the list length effect is evident when stimuli other than words are used.

## Experiment 1: Word pairs

One variation on using single words as stimuli is to pair two words together, as in the associative recognition paradigm. Three experiments have manipulated list length in an associative recognition design, with divided results. Clark and Hori (1995) and Nobel and Shiffrin (2001) identified significant list length effects, while Criss and Shiffrin (2004b) did not. None of these experiments controlled for all of the potential confounds listed above.

In Clark and Hori's (1995) experiment, list length was manipulated between 34 and 100 pairs (list length ratio of 1:3). The authors implemented a 45-s retention interval period of mental arithmetic prior to each test list, which could also be considered a control for contextual reinstatement. Nobel and Shiffrin (2001) manipulated list length between 10 and 40 pairs in their first experiment (list length ratio of 1:4), while in their third experiment, list length was either 10 or 20 pairs (list length ratio of 1:2). They also controlled for contextual reinstatement by including a period of mental arithmetic for 26 s prior to each test list, but this period may not have been of sufficient length to encourage contextual reinstatement following the long list (Dennis et al., 2008). Criss and Shiffrin's (2004b) experiment involved participants being presented with three types of lists: pairs of faces (FF), pairs of words (WW), or one word and one face paired together (WF). For lists of word pairs, list length was varied between 20, 30, and 40 pairs (list length ratio of 1 : 1.5 : 2). Participants were asked to rate the degree of association between the two items that made up each pair, which could act as a control for attention. This study used a within-list manipulation of list length that would also have provided controls for retention interval and displaced rehearsal.

There may also be something special about using the associative recognition design, rather than a single-item recognition experiment. Hockley (1991, 1992) carried out a series of experiments comparing forgetting rates in the recognition of single words and of word pairs. The consistent finding was that the rate of forgetting was greater for single words than for word pairs. He concluded that memory for word pairs is "more resistant to the effects of decay, interference from intervening events, or both" than is memory for single items (Hockley, 1992, p. 1328). However, Weeks, Humphreys, and Hockley (2007) later argued that Hockley's (1991, 1992) finding of flat forgetting curves did not show a complete absence of forgetting, but instead suggested that interference was affecting targets and distractors in the same way. Nevertheless, it may be that there is no significant effect of list length when the stimuli are word pairs, simply because they are resistant to the influence of interference from other study pairs.

A partial replication of Criss and Shiffrin's (2004b), Clark and Hori's (1995), and Nobel and Shiffrin's (2001) studies was the basis of the present experiment using word pairs and the associative recognition design. Controls for all potential confounds were introduced, and for consistency with previous experiments (e.g., Cary & Reder, 2003; Kinnell & Dennis, 2011; Nobel & Shiffrin, 2001), the list length ratio was set at 1:4. The primary aim of Experiment 1 was to determine whether word pairs behave differently from single words and whether there is a significant list length effect when the former are used as stimuli.

### Method

*Participants* The participants were 40 first-year psychology students at Ohio State University, who participated in exchange for course credit.

*Design* A 2 × 2 factorial design was used. The factors were List Length (short or long) and Word Frequency (low or high; note that "word frequency" refers to the frequency of both items in the pair, so two high-frequency words made up a high-frequency pair, and two low-frequency items made up a low-frequency pair). The word frequency manipulation was included in this experiment for consistency with previous experiments conducted by the authors. However, the nature of the word frequency effect in associative recognition is not as clear as in single-item recognition. In single-item recognition, performance for low-frequency words is superior to that for high-frequency words, while in associative recognition, the reverse is generally true (Clark, 1992; Clark & Burchett, 1994; Clark, Hori, & Callan, 1993, Exps. 1 and 2; Clark & Shiffrin, 1992); however, some studies (e.g., Clark et al., 1993, Exp. 3; Hockley, 1992) have not identified a significant effect of word frequency. Both were within subjects manipulations.

*Materials* The stimuli for this experiment were 240 five- and six-letter words from the *Sydney Morning Herald* Word Database (Dennis, 1995). Half of the words were high

frequency (100–200 occurrences per million), and half were low frequency (1–4 occurrences per million). Words were randomly paired with another word of the same frequency. All word pairs were randomly assigned to lists for each participant, and no participant saw the same pair of words twice, with the exception of targets.

*Procedure* Upon arrival, participants were given an overview of the experiment and completed a practice session with the sliding-tile puzzle activity that would be used throughout the experiment during retention intervals. Participants completed two study lists, one short (24 word pairs) and one long (96 word pairs), with the list presentation order counterbalanced for participants. Each pair appeared on screen for 3,000 ms. High-frequency pairs comprised half of the list, and the other half were low-frequency pairs. The two words of each pair were presented side by side on a computer screen.

During study, participants were asked to rate how related the two words were to each other on a 6-point Likert scale (1 = *unrelated*, 6 = *related*) by clicking on the appropriate number displayed below the word pair. They were instructed to make this rating while the word pair was on screen, and to move on to and rate the next pair should they miss giving a rating.

The experiment had a retroactive design as a control for retention interval. This meant that the short list was followed by a 3-min 36-s filler period with the sliding-tile puzzle, and that the first 8 word pairs from both the short and long lists were included as the targets at test. One word was taken from each of the next 16 word pairs at study to create rearranged pairs at test. Each of the words in a rearranged pair was presented in the same screen position (left or right) where it had been presented as part of its original pair. Test pairs were presented in a random order.

Participants were given a 20-s warning before the start of the test list. They were instructed to respond "yes" if they recognised a word pair from the study list and to respond "no" if they did not. Responses were recorded when the participant clicked on the appropriate button, which was displayed on screen below the word pair. Test lists comprised eight target pairs and eight rearranged pairs. There was no time limit for responding, and there were no missing data.

In addition to the control for retention interval following the short lists, participants completed 8 min of sliding-tile puzzle filler before the onset of each test list (a control implemented to encourage contextual reinstatement).

Results

*Within-subjects versus between-subjects analysis* This experiment, and those to follow in this article, was designed with list length as a within-subjects manipulation, in order to ensure that it had the greatest experimental power possible and was consistent with the method of previous studies in the area. However, it was later noted that the use of a within-subjects design might itself mask list length effects (Kinnell & Dennis, 2011). Indeed, inspection of the within-subjects analyses revealed statistically significant interactions between list length and list order for all but one of the experiments presented in this article [the interaction was not significant in the present experiment: $F(1, 38) = 0.01$, $p > .05$].

The first potential reason for these significant interactions was that for the first list studied, each participant was naive, whereas all participants were experienced by the second list. For example, a participant who sees the long list first might expect the second list to be of equal length. This expectation could influence the study strategy they adopted; for example, they might prepare to spread their attention over another long list of items and, as a result, pay less attention to the second list than a participant who began with a short list. Similarly, a participant who began with a short list might also expect the second list to be short and attend to items in the same manner. In the retroactive condition, this focused attention at the start of the long list would be on precisely the items on which they would later be tested and would positively influence performance on the long list when it was viewed second.

Furthermore, a participant who viewed the short list second would already have seen 80 long-study-list items plus 40 test items prior to the start of the second (short) list. Consequently, this participant might have been bored prior to the start of the second list and therefore have been less likely to pay attention to the short study list than if it had been viewed first. This would again favour performance on the long list when viewed second. Finally, the use of the within-subjects design meant that by the end of the second study list, all participants would have seen an equal number of study list items, irrespective of whether that second list was long or short. This might also have made finding a significant effect of list length less likely.

To address this potential problem, two sets of analyses were conducted on the data in each experiment. A within-subjects analysis was carried out, as well as a between-subjects analysis using only the data from the first list studied by each participant. The results from both analyses are reported for each experiment. For all analyses presented, the results produced $F < 1$ and $p > .05$ unless otherwise noted.

*List length* Figure 3 shows the effect of list length on $d'$, while Table 1 displays the hit and false alarm rate data. Corrections were made to all hit and false alarm rates in this study prior to calculation of $d'$ to avoid infinite values of this statistic when recognition performance was at ceiling or floor. As suggested by Snodgrass and Corwin (1988), the
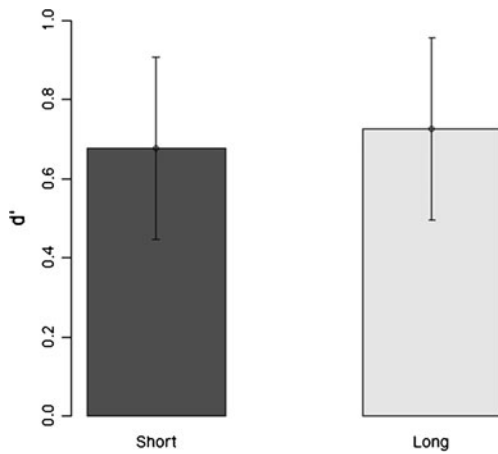
**Fig. 3** Mean $d'$ values for short and long lists in Experiment 1 (word pairs). Bars represent 95% confidence intervals (within-subjects analysis)



**Fig. 4** A significant effect of word frequency was identified only for the false alarm rate in the word pair data. There was no significant effect on the hit rate. Bars represent 95% confidence intervals (within-subjects analysis). *Difference significant at the $p < .05$ level

corrections were made by adding a value of 0.5 to the hit and false alarm counts and adding 1 to the number of target and distractor items.

In both the within- and between-subjects analyses, 2 (length) × 2 (frequency) repeated measures ANOVAs did not yield statistically significant effects of list length on $d'$, hit rate, or false alarm rate.

*Word frequency* There was a significant main effect of word frequency on $d'$, $F(1, 39) = 10.17$, $p = .003$, $\eta_p^2 = .21$, and on the false alarm rate, $F(1, 39) = 17.04$, $p < .001$, $\eta_p^2 = .30$, with a higher false alarm rate for low-frequency pairs (Fig. 4). However, there was no significant effect of word frequency on the hit rate.

*Response latency* One-way ANOVAs yielded nonsignificant effects of list length on the median response latencies for correct and incorrect responses in both the within- and between-subjects analyses; see Table 2. Medians were used to minimise the effect of outlier response latencies.
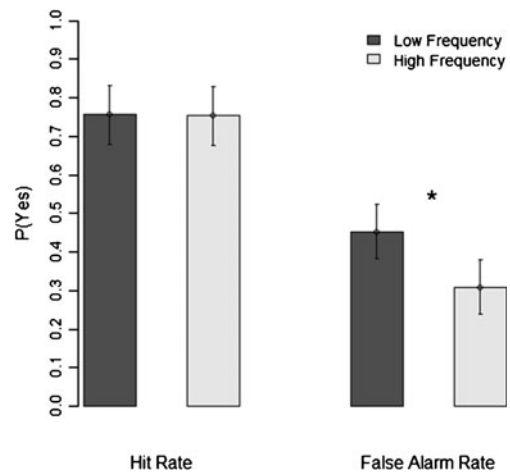
Discussion

The results of Experiment 1, using word pairs as the stimuli, were consistent with the findings of Criss and

**Table 1** Mean hit and false alarm rates for short and long lists in Experiment 1 (word pairs)

| | Hit Rate | | False Alarm Rate | |
|---|---|---|---|---|
| | Short List | Long List | Short List | Long List |
| High frequency | .72 (.22) | .79 (.22) | .28 (.27) | .34 (.27) |
| Low frequency | .73 (.30) | .78 (.23) | .46 (.33) | .45 (.31) |

Standard deviations appear in parentheses.

Shiffrin (2004b) for word pairs, but were in contrast to those of Clark and Hori (1995) and Nobel and Shiffrin (2001). No significant list length effect was identified in either the accuracy or response latency data in the within- or between-subjects analyses. The list length ratio in this experiment was stronger than in the studies of Criss and Shiffrin (2004b) or Clark and Hori, or in Nobel and Shiffrin's Experiment 1. The larger list length ratio would make it more likely that we would identify a significant effect of list length, but this was not the case. However, the increase in the list length ratio was likely offset by the introduction of controls for the four potential confounds.

It should also be acknowledged that using the retroactive design to control for retention interval might have had an influence on the results of this experiment. While the first 8 pairs of each study list were included as targets at test, the distractor pairs were created using one word from each of the next 16 pairs, meaning that they had been more recently encountered than the targets. This could perhaps have affected recognition performance. However, the situation was the same for both the short and long lists and would not have had an impact on the list length findings.

The word frequency effect was identified in the $d'$ and false alarm rate data, with better performance for high-frequency pairs, consistent with some previous research (Clark, 1992; Clark & Burchett, 1994; Clark et al., 1993, Exps. 1 and 2; Clark & Shiffrin, 1992). However, the effect was nonsignificant in the hit rate data, consistent with other previous findings (Clark et al., 1993, Exp. 3; Hockley, 1992). It appears that there is a high-frequency advantage, but the mirror pattern that is often observed in single-item recognition (Glanzer & Adams, 1985) does not appear to be robust in associative recognition.

**Table 2** The mean of median response latencies (in milliseconds) for word pairs in Experiment 1

Standard deviations appear in parentheses.

|  | Within-Subjects Analysis | | Between-Subjects Analysis | |
|---|---|---|---|---|
|  | Short List | Long List | Short List | Long List |
| Correct responses | 1,551 (389) | 1,502 (400) | 1,609 (498) | 1,385 (271) |
| Incorrect responses | 2,110 (1,008) | 2,032 (739) | 2,359 (1,121) | 1,892 (720) |

## Experiment 2: Faces

In Experiment 2, images of novel faces were used as the stimuli. While we are used to seeing many faces in our everyday lives, these particular examples had never been encountered before by the participants. Criss and Shiffrin's (2004b) associative recognition experiment revealed that words behave differently from faces, in that the list length effect was identified for the latter but was not significant for words.

Much research has been devoted to identifying differences in the processing, encoding, and retrieval of words and faces. It has been suggested that words and faces are processed in the same way, as a series of parts (e.g., Martelli, Majaj, & Pelli, 2005). However, others have argued that faces are processed holistically, while words are processed as a series of parts that together make up the whole (e.g., Farah, Wilson, Drain, & Tanaka, 1998). Thus, a nonsignificant effect of list length when the stimulus is words does not necessarily mean that there will also be a nonsignificant effect when faces are used as the stimuli.

Chalmers (2005) noted that faces have nameable features (e.g., eyes, nose, and lips), but that these features are common to every face, making the stimuli difficult to describe in a unique way. This, too, is different from words, which, even with different combinations of features, combine to form a unique word that is easy to describe. When it comes to a recognition memory test, this may mean that more overlap occurs in the encoding of faces than of words. As Jacoby and Dallas (1981) noted, the processing of such verbal stimuli as words is more fluent than that of nonverbal stimuli, such as images of novel faces. Thus, adding extra faces to a study list might result in greater interference than adding extra words, and might make a significant list length effect more likely to result for faces.

Alternatively, it has been argued that words and faces behave in similar manners. Xu and Malmberg (2007) conducted an associative recognition study in which the type of stimuli used for the pairs was manipulated between subjects. The lists were made up of word pairs, face pairs, pseudoword pairs, or Chinese character pairs. Xu and Malmberg found that the pattern of results for words resembled that for faces, with pseudowords and Chinese characters behaving differently from both. They proposed that words and faces behave in the same way because they are more commonly encountered in everyday life than are pseudowords and Chinese characters (by non-Chinese speakers).

The aim of Experiment 2 was to investigate whether there was a significant effect of list length on recognition performance when novel faces were used as the stimulus.

Method

*Participants* A group of 40 first-year psychology students from Ohio State University participated in this experiment. They received course credit for their participation.

*Design* Length (short or long) was the only factor manipulated in this experiment and was manipulated within subjects.

*Materials* The stimuli in this experiment were 140 colour images of faces taken from the AR Face Database (Martínez & Benavente, 1998; see Fig. 5 for examples). Half of the images were of males and half of females. All images were 460 × 460 pixels in size and were randomly assigned to lists, with no image appearing twice.

*Procedure* The procedure of Experiment 2 resembled that of Experiment 1, with some key differences. The procedure was first described to participants, who then completed a practice session of the sliding-tile puzzle activity that would be used throughout the experiment. Participants completed



**Fig. 5** Examples of face stimuli from the AR Face Database (Martínez & Benavente, 1998) used in Experiment 2. Half of the images were of females, half were of males

two study lists, one short (20 items) and one long (80 items), with the list presentation order counterbalanced across participants. Each face appeared on screen for 4,000 ms. Male faces made up half of each list, and the other half were female faces. Test lists comprised 20 targets and 20 distractors. All faces were presented in the middle of the computer screen.

During study, participants were asked to rate the pleasantness of each image on a 6-point Likert scale (1 = *least pleasant*, 6 = *most pleasant*) by clicking on the appropriate number displayed below the image. They were instructed to make this rating while the image was being displayed on screen and to move on to and rate the next image should they miss giving a rating in the allotted time.

The experiment had a retroactive design as a control for retention interval. The short list was followed by a 4-min period of sliding-tile puzzle filler, and the first 20 face images from the long list were included as the targets at test.

Participants were given a 15-s warning before the start of the test list. Using the yes/no recognition paradigm, participants were instructed to respond "yes" if they recognised a face image from the study list and "no" if they did not. Responses were recorded when the participant clicked on the appropriate button displayed on screen. There was no time limit for responding, and there were no missing data.

Contextual reinstatement was facilitated following both lists by including an 8-min period of sliding-tile puzzle before the onset of each test list, in addition to using the puzzle as a control for retention interval.

Results

*Length × Order interaction* A 2 × 2 mixed design ANOVA yielded a statistically significant interaction between list length (short vs. long) and list order (long–short vs. short–long), $F(1, 38) = 15.87$, $p < .0001$, $\eta_p^2 = .30$.

*List length* Figure 6 shows the effect of list length on $d'$, and Table 3 shows the hit and false alarm rate data. A one-way repeated measures ANOVA yielded a statistically significant effect of list length on $d'$ in the within-subjects analysis, $F(1, 39) = 6.53$, $p = .01$, $\eta_p^2 = .14$, and a marginally significant effect in the between-subjects analysis, $F(1, 38) = 3.53$, $p = .07$, $\eta_p^2 = .08$. This was driven by a significant effect of list length on the false alarm rates under both the within-subjects, $F(1, 39) = 12.16$, $p = .001$, $\eta_p^2 = .24$, and between-subjects, $F(1, 38) = 4.56$, $p = .04$, $\eta_p^2 = .12$, analyses. The effect of list length on the hit rate was not statistically significant in either analysis.

*Response latency* One-way repeated measures ANOVAs did not yield significant effects of list length on the median
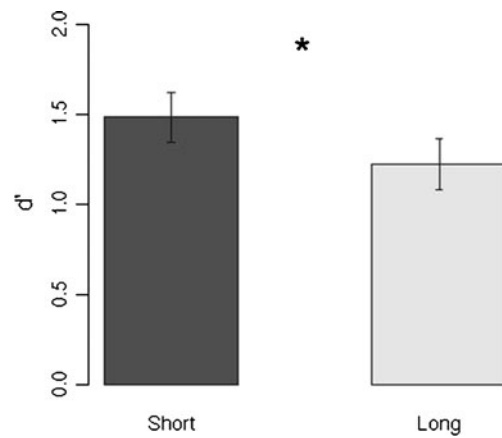


**Fig. 6** Mean $d'$ values for short and long lists in Experiment 2 (faces). Bars represent 95% confidence intervals (within-subjects analysis). [*]Difference significant at the $p < .05$ level

response latencies for correct or incorrect responses in either the within- or the between-subjects analyses; see Table 4.

Discussion

Consistent with the results of Criss and Shiffrin (2004b), a statistically significant effect of list length was identified on recognition performance when unfamiliar faces were used as the stimuli. However, this was in contrast to the nonsignificant effects of list length reported in several previous studies (e.g., Dennis & Humphreys, 2001; Dennis et al., 2008; Kinnell & Dennis, 2011) and in Experiment 1. It seems that, contrary to the work of Xu and Malmberg (2007), there is something different about words and faces. The different findings may be attributable to possible differences in the processing of faces and words—that is, holistically or as a combination of parts, respectively (see Farah et al., 1998).

Alternatively, the hypothesis of Chalmers (2005), that there is a lack of unique ways in which one can describe and encode faces, may explain the different results. This difficulty in encoding face stimuli may result in greater overlap in the representations of the faces that appeared at study versus a similar list composed of words. Thus, the addition of more faces to a study list would lead to greater interference than would the addition of other words to the

**Table 3** Mean hit and false alarm rates for the short and long lists in Experiment 2 (faces)

|  | Hit Rate | False Alarm Rate |
|---|---|---|
| Short list | .79 (.15) | .25 (.16) |
| Long list | .80 (.14) | .34 (.17) |

Standard deviations appear in parentheses.

**Table 4** The mean of median response latencies (in milliseconds) for faces in Experiment 2

| | Within-Subjects Analysis | | Between-Subjects Analysis | |
|---|---|---|---|---|
| | Short List | Long List | Short List | Long List |
| Correct responses | 1,169 (162) | 1,212 (188) | 1,180 (164) | 1,214 (191) |
| Incorrect responses | 1,325 (631) | 1,346 (406) | 1,269 (397) | 1,300 (340) |

Standard deviations appear in parentheses.

study list—hence, the significant list length effect finding in the present case. This greater interference might also be reflected in the false alarm rates, which were high compared with those in previous experiments using single words as the stimuli (e.g., Kinnell & Dennis, 2011), consistent with the idea that "yes" responses are more probable when the stimuli are nonverbal (e.g., Greene, 2004; Whittlesea & Williams, 2000; but see Xu & Malmberg, 2007).

While the face stimuli in this experiment were unfamiliar, in that it was highly unlikely that a participant would have ever seen the faces in any other context, participants were certainly familiar with looking at faces in everyday life. Thus, there had been many previous instances of witnessing faces with varying degrees of similarity to those presented at study. The impact on the list length effect of stimuli that are encountered less often—images of fractals—was investigated next.

## Experiment 3: Fractals

A fractal is "a geometrical figure in which an identical motif repeats itself on an ever diminishing scale" (Lauwerier, 1991, p. xi). As a consequence of their being novel stimuli, the encoding of fractals may be negatively affected by the lack of readily available appropriate labels with which to tag and encode the individual examples (Curran, Schacter, Norman, & Galluccio, 1997; Gardiner & Java, 1990), as may have also been the case with Brandt's (2007) checkerboard stimuli. With faces, there are common features that can be described in different ways—for example, big nose, blonde hair, and blue eyes. With fractals, it is not as apparent what features can be described—for example, a certain pattern or shape may be present in some but not all fractal images, while a nose and mouth are present on every face. This may lead to a longer description of each image being employed and to greater overlap in encoding than is the case with words.

This difficulty in encoding, together with these stimuli being less commonly encountered than faces, means that the fractal images may be subject to more interference, in that they are more confusable with each other than are images of faces, especially since recognition performance for faces is considered to be quite high (e.g., Bahrick,

Bahrick, & Wittlinger, 1975). Experiment 3 aimed to investigate whether the list length effect was evident when fractals were used as the stimuli.

Method

*Participants* The participants in this experiment were 40 first-year psychology students at Ohio State University. They each received course credit in exchange for their participation.

*Design* This experiment had a 2 × 2 factorial design and the factors were List Length (short or long) and Fractal Type (circle or leaf). Both were within-subjects manipulations.

*Materials* The stimuli used in this experiment were 140 colour images of fractals, each 600 × 400 pixels in size. Half of the fractals were classed as *circle fractals*, which involved a circular shape as the centrepiece, and the remainder were termed *leaf fractals* (see Fig. 7 for examples), which involved leaf-like shapes scattered across the image. All images were randomly assigned to lists, with no images except targets appearing twice.

*Procedure* The procedure for this experiment was identical to that of the previous experiment, with the only difference being the stimuli used. Each list was made up of equal numbers of both circle and leaf fractals.
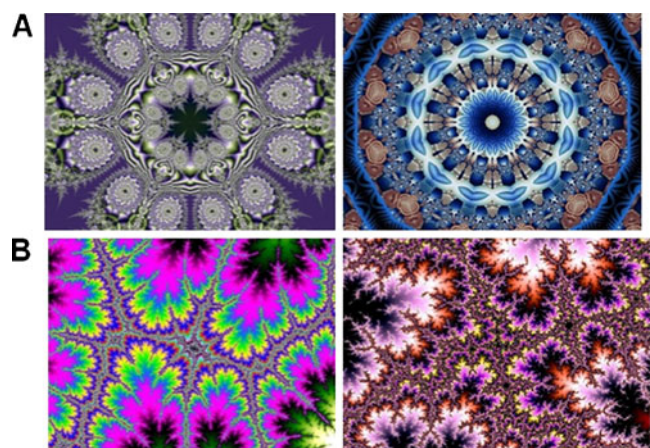


**Fig. 7** (A) Two examples of "circle" fractals and (B) two examples of "leaf" fractals

Results

*Length × Order interaction* A 2 × 2 mixed design ANOVA yielded a statistically significant interaction between list length (short vs. long) and list order (long–short vs. short–long), $F(1, 38) = 4.71$, $p = .04$, $\eta_p^2 = .11$.

*List length* Figure 8 and Table 5 present the effect of list length on $d'$ and the hit and false alarm rate data, respectively. A 2 × 2 (Length × Fractal Type) repeated measures ANOVA did not yield a statistically significant effect of list length on $d'$ in the within-subjects analysis; however, when the between-subjects analysis was conducted, a significant effect was identified, $F(1, 38) = 7.40$, $p = .01$, $\eta_p^2 = .16$. In both analyses, there was no significant effect of list length on the hit rate, but the effect on the false alarm rate was significant in both the within-subjects, $F(1, 39) = 10.86$, $p = .002$, $\eta_p^2 = .22$, and between-subjects, $F(1, 38) = 4.84$, $p = .03$, $\eta_p^2 = .11$, analyses.

*Fractal type* The 2 × 2 ANOVA also yielded a statistically significant effect of fractal type (leaf vs. circle) on $d'$ in both the within-subjects, $F(1, 39) = 13.39$, $p < .001$, $\eta_p^2 = .26$, and between-subjects, $F(1, 38) = 17.43$, $p < .001$, $\eta_p^2 = .31$, analyses, with better performance for circle fractals. The effect of fractal type on the false alarm rate was also statistically significant in the within-subjects, $F(1, 39) = 12.30$, $p = .001$, $\eta_p^2 = .24$, and between-subjects, $F(1, 38) = 17.97$, $p < .001$, $\eta_p^2 = .32$, analyses. The effect of fractal type on hit rates was nonsignificant for both types of analysis (see Table 5 for the hit and false alarm data).

*Response latency* One-way repeated measures ANOVAs yielded statistically significant effects of list length on the median response latencies for correct responses and incorrect responses in both the within-subjects [correct, $F(1, 39) = 17.85$,
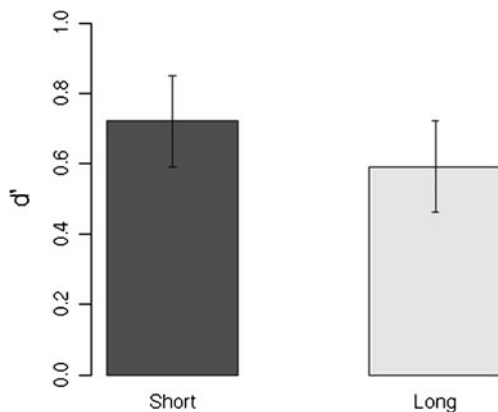
**Table 5** Mean hit and false alarm rates for fractal data in Experiment 3 (fractals)

|  | Hit Rate | | False Alarm Rate | |
|---|---|---|---|---|
|  | Short List | Long List | Short List | Long List |
| Circle fractals | .66 (.21) | .69 (.22) | .30 (.21) | .37 (.25) |
| Leaf fractals | .62 (.26) | .69 (.19) | .40 (.21) | .54 (.25) |

Standard deviations appear in parentheses.

$p = .0001$, $\eta_p^2 = .31$; incorrect, $F(1, 39) = 24.29$, $p < .0001$, $\eta_p^2 = .38$] and between-subjects [correct, $F(1, 38) = 7.31$, $p = .01$, $\eta_p^2 = .16$; incorrect, $F(1, 38) = 5.06$, $p = .03$, $\eta_p^2 = .12$] analyses; see Table 6.

Discussion

When the results of the present experiment were analysed using the within-subjects data, a statistically significant effect of list length on the response latency data for correct and incorrect responses was identified, but this was not reflected in the accuracy data (with the exception of the false alarm rate). Analysis of the between-subjects data resulted in significant effects of list length in both the accuracy and response latency data. This discrepancy highlights the importance of analysing and reporting both sets of data. Taken together, the results suggest a significant effect of list length on recognition performance for fractals.

Overall, recognition performance in this experiment was low. This decreased performance was likely the result of overlapping representations of the fractal images. As Chalmers (2005) noted with respect to faces, there is a shortage of unique ways in which complex stimuli of this type can be described and uniquely encoded into memory. When this is the case, the addition of extra items to the list has a more detrimental effect on performance, as compared with word lists. As was the case with the faces in the previous experiment, the false alarm rates for fractals were higher than those in previously published studies (e.g., Kinnell & Dennis, 2011). Brandt (2007) also noted that the list length differences in his checkerboard list length study were driven by an increase in false alarm rates. These results are another example of a tendency for participants to provide more "yes" responses for nonverbal stimuli (Greene, 2004; Whittlesea & Williams, 2000; but see Xu & Malmberg, 2007).

**Experiment 4: Photographs of scenes**

The experimental results to this point have suggested that, when controls for potential confounds are in place, no significant effect of list length is identified for single words



**Fig. 8** Mean $d'$ values for short and long lists in Experiment 3 (fractals). Bars represent 95% confidence intervals (within-subjects analysis)

**Table 6** The mean of median response latencies (in milliseconds) for fractals in Experiment 3

|  | Within-Subjects Analysis | | Between-Subjects Analysis | |
| --- | --- | --- | --- | --- |
|  | Short List | Long List | Short List | Long List |
| Correct responses | 1,018 (347) | 1,226 (204) | 1,011 (351) | 1,256 (202) |
| Incorrect responses | 980 (372) | 1,253 (268) | 975 (369) | 1,194 (231) |

Standard deviations appear in parentheses.

(see, e.g., Kinnell & Dennis, 2011) or for word pairs in an associative recognition paradigm. There is, however, a significant list length effect when novel faces and fractal images are used as the stimuli. The final experiment of this study was designed in an attempt to fall somewhere between these two types of stimuli and to help establish the boundary conditions between the two.

Shepard (1967) looked at recognition performance on lists of words (540 items), sentences (612 items), and pictures (612 items). In a forced choice recognition paradigm involving 68 test pairs, Shepard found that performance for all three types of stimuli was high, and he noted that discriminability was at its best when the "stimuli were meaningful, colored pictures" (1967, p. 159). The meaningfulness of the stimuli was revisited by Chalmers (2005), who noted that this may, in part, explain differences in performance on novel faces, very low-frequency (novel) words, and pictures of complex scenes. Thus, colour photographs of complex scenes were used as the stimuli in Experiment 4. These were visual in nature, should be meaningful to the participants, and should elicit high recognition performance.

Photographs of this kind are also more nameable than the images of either faces or fractals, and participants should be better able to encode them uniquely (Chalmers, 2005). This should also improve performance and allow us to ascertain whether it is the visual nature of faces and fractals or the difficulty of unique encoding that has resulted in the contradictory list length results for these stimuli and for words.

## Method

*Participants* A group of 40 first-year psychology students from Ohio State University participated in this experiment in exchange for course credit.

*Design* List length (short or long) was manipulated within subjects in Experiment 4.

*Materials* The stimuli for this experiment were 140 different colour photographs of everyday scenes—for example, images of a library interior, a beach, and a classroom (see Fig. 9 for examples). Each photograph was 800 × 600 pixels in size.

*Procedure* The procedure for this experiment largely followed that of the previous two experiments. The main difference was the duration of presentation of the items at study. A pilot study was carried out in which the photographs of scenes were presented for 3,000 ms at study, as in the previous experiments, but performance was at ceiling. Therefore, in an effort to reduce performance, the rate of presentation of the stimuli at study was cut to 500 ms with a 250-ms interstimulus interval, during which time the screen was blank. As a consequence of the shortened rate of presentation, it was not possible to request a response to an encoding task while the stimuli were on screen, as had been the case in all previous experiments. Given the short overall duration of the present experiment and the results of Kinnell and Dennis (2011), which had revealed that the encoding task does not alter the list length effect finding, this did not seem problematic in terms of the controls for the potential list length effect confounds. This experiment used the retroactive experimental design, with the first 20 items of the long list included as targets at test. An additional 45 s of filler activity followed the short list, in order to equate the retention interval with that of the long list. All other details were as in Experiments 2 and 3.

**Fig. 9** Examples of photographs used as stimuli in Experiment 4 (photographs)

Results

*Length × Order interaction* A 2 × 2 repeated measures ANOVA yielded a statistically significant interaction between list length (short vs. long) and list order (long–short vs. short–long), $F(1, 38) = 20.83$, $p < .0001$, $\eta_p^2 = .35$.

*List length* Figure 10 shows the effect of list length on $d'$, and Table 7 presents the hit and false alarm rate data. One-way ANOVAs did not yield significant effects of list length on recognition performance, as measured by $d'$, in either the within- or between-subjects analyses. The effect of list length on the hit rate, $F(1, 39) = 2.09$, $p = .16$, and the false alarm rate was also nonsignificant. Despite the nonsignificance of these results, the means for $d'$ and the hit rate were in the direction opposite the traditional notion of the list length effect, with long-list performance superior to that of the short list in this case.

*Response latency* There was no significant effect of list length on the median response latencies for correct responses [within: $F(1, 39) = 2.20$, $p = .15$] or incorrect responses [within: $F(1, 35) = 2.56$, $p = .12$], as revealed by one-way repeated measures ANOVAs on both the within- and between-subjects data; see Table 8.

Discussion

Following the pilot study, in which performance was at ceiling, the reduced presentation time of the photographic stimuli in Experiment 4 successfully brought about deterioration in recognition performance. No significant effect of list length was identified in this experiment in either the accuracy or the response latency data, and regardless of whether the design and analyses were within or between subjects.
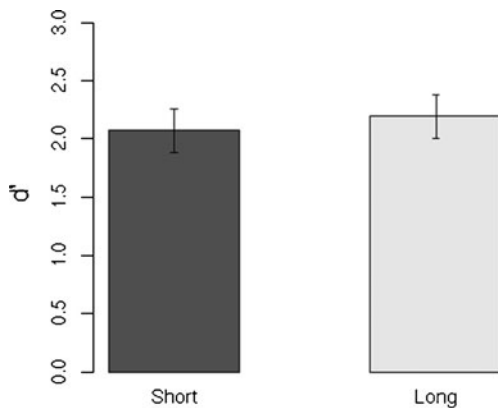


**Fig. 10** Mean $d'$ values for short and long lists in Experiment 4 (photographs of scenes). Bars represent 95% confidence intervals (within-subjects analysis)

**Table 7** Mean hit and false alarm rates for the short and long lists in Experiment 4 (photographs of scenes)

|            | Hit Rate   | False Alarm Rate |
|------------|------------|------------------|
| Short list | .80 (.16)  | .10 (.12)        |
| Long list  | .84 (.14)  | .10 (.12)        |

Standard deviations appear in parentheses.

Contrary to results for the images of faces and fractals that were the stimuli in the previous two experiments, the photographic images in the present experiment were straightforward to describe and to name. There was no overlap in the scenes depicted in the photographs, and these could be described in most cases using just one word—for example, "library," "beach," or "classroom." In this way, the photographs of scenes were similar to words, which might explain why the list length findings for the two classes of stimuli follow the same pattern. In addition, the false alarm rate in the present experiment was similar to those for the word stimuli in Experiment 1 and in Kinnell and Dennis's (2011) experiments, and lower than the false alarm rates when faces and fractals were the stimuli. This finding suggests that, despite the photographs technically being nonverbal stimuli, like faces and fractals, they behave as if they were verbal, suggesting that participants might have been basing their decisions on the verbal labels given to the photographs (Greene, 2004; Whittlesea & Williams, 2000; but see Xu & Malmberg, 2007). On this basis, it seems reasonable to suggest that if all study items and distractors shared the same verbal label—that is, different scenes that could also be labelled, for example, a "beach"—then a list length effect would result. If all of the items were images of beaches, one would presumably have to encode the stimuli, as with faces and fractals, by a lengthy description, which might lead to greater overlap in the representations and a list length effect.

**General discussion**

In an attempt to identify and define the boundary conditions of the list length effect, a series of four experiments was conducted, each involving recognition testing of a different stimulus: word pairs, faces, fractals, and photographs of scenes. Controls for the potentially confounding effects of retention interval, attention, displaced rehearsal, and contextual reinstatement were implemented in each case, with the method generally following that of several previous studies (e.g., Cary & Reder, 2003; Dennis & Humphreys, 2001; Dennis et al., 2008, Kinnell & Dennis, 2011). The results varied across experiments. Regardless of whether the within- or between-subjects data were analysed, no

**Table 8** The mean of median response latencies (in milliseconds) for photographs of scenes in Experiment 4

| | Within-Subjects Analysis | | Between-Subjects Analysis | |
|---|---|---|---|---|
| | Short List | Long List | Short List | Long List |
| Correct responses | 1,099 (146) | 1,124 (138) | 1,103 (183) | 1,134 (137) |
| Incorrect responses | 1,299 (484) | 1,426 (457) | 1,464 (625) | 1,385 (329) |

Standard deviations appear in parentheses.

significant effect of list length was identified on recognition performance for word pairs or photographs of scenes in either the accuracy or response latency data. There was a significant list length effect on recognition performance for faces based on the accuracy data in the within-subjects analysis, with the effect being marginally significant in the between-subjects analysis. The within-subjects analysis of response latency data for fractals showed a statistically significant effect of list length that was absent in the accuracy analysis. However, the between-subjects analysis of the fractal data revealed significant effects of list length on both $d'$ and response latency.

The effect sizes for word pairs ($\eta_p^2 = .03$ within subjects, $\eta_p^2 = .01$ between subjects) and photographs ($\eta_p^2 = .02$ within subjects, $\eta_p^2 = .01$ between subjects) were very small. The effect sizes for the faces (mean $\eta_p^2 = .14$ within subjects, .08 between subjects) and fractals (mean $\eta_p^2 = .04$ within subjects, .16 between subjects) stimuli were larger. See Fig. 11. However, all of these effect sizes are small, especially when we consider that they are based on a fourfold increase in list length. If item noise played a
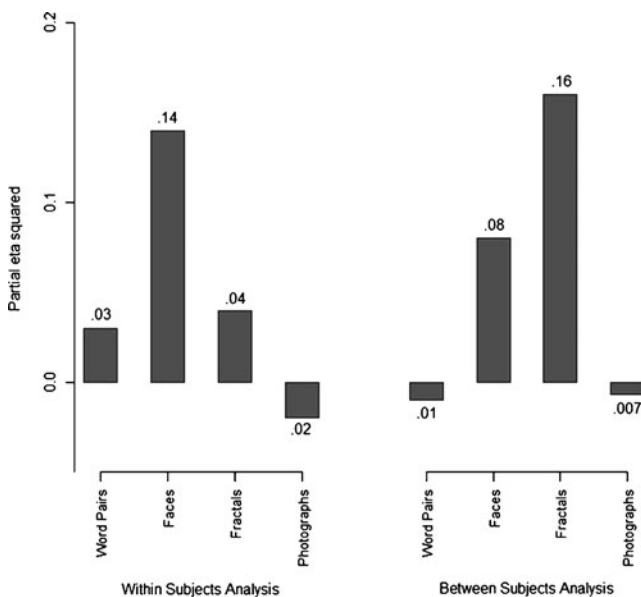
substantial role in recognition memory, a larger effect would be expected, given the four-times increase in interference.

One explanation for the different results for the various stimuli used is the difference in the ease with which the stimuli could be named or labelled, which may also be related to the similarity of the stimulus items to others of the same class. Words are themselves unique labels as stimuli, and the photographic images, while complex, could be easily and uniquely described with a single word and thus might have been coded as verbal stimuli. This ease of labelling reduces the similarity within this type of stimuli. Labelling appears to have been more difficult for the nonverbal face and fractal stimuli in which there was necessarily more overlap in feature descriptions and greater similarity within the stimulus class. Thus, the more of these items that were on the study list, the more problematic it became for the participant to create unique labels for the items. This was reflected in the false alarm rates, which were higher for faces and fractals than for the other stimuli, consistent with the suggestions of Greene (2004) and Whittlesea and Williams (2000; but see Xu and Malmberg, 2007). It might be that the different types of stimuli lie on a continuum, from the nonverbal faces and fractals, through photographs of scenes, to words, which are effectively verbal stimuli.

Another possibility is that the critical dimension of difference between the stimuli was the density of the space in which they were encoded, regardless of how that came about. For instance, words may be more sparsely encoded because they are experienced frequently, as proposed by Xu and Malmberg (2007). But pictures may be more sparsely encoded because they are richer, more meaningful stimuli, rather than because they are more readily coded verbally. Faces and fractals, on the other hand, vary on only a small number of dimensions, and so may overlap to a greater degree.



**Fig. 11** The partial eta-squared effect sizes for each of the experiments in this article. Effect sizes from the within-subjects analyses are presented on the left, and those from the between-subjects analysis are presented on the right. Negative values indicate a situation in which long-list performance was superior to short-list performance

### Implications for memory models

The results of the four experiments presented in this article have consequences for mathematical models of recognition memory and for the item and context noise approaches to interference. The item noise approach posits that interference is generated by the other items that make up the study list (Criss & Shiffrin, 2004a). In contrast, in the context noise approach, the previous contexts in which an item has been seen are the source of interference (Dennis &

Humphreys, 2001). Alternatively, interference may arise both from other list items and from previous contexts (Cary & Reder, 2003; Criss & Shiffrin, 2004a). Item noise models predict a significant effect of list length, but context noise models do not.

The existence of list length effects in recognition memory for fractals and faces implies that one cannot maintain a strict context noise position across all stimulus classes. Item noise does seem to play some role, although the effects are small and not always present. This conclusion was perhaps inevitable. The strict version of the context noise approach relies on the notions that stimuli are effectively coded as local representations and that, at test, one is able to reactivate each representation unerringly. In the limit as stimuli become increasingly similar, this property must break down.

What is surprising in the context of existing models of recognition memory is how similar stimuli must be in order for item interference to become detectable. Global-matching models (Gillund & Shiffrin, 1984; Hintzman, 1984; Humphreys, Bain, & Pike, 1989; Murdock, 1982; Shiffrin & Steyvers, 1997) have relied on the similarity of item representations to explain why lures drawn from the same categories as items on the study list show high false alarm rates (Clark & Gronlund, 1996; Shiffrin, Huber, & Marinelli, 1995). For such stimuli as faces and fractals, this may be the case. However, for words, word pairs, and photographs of scenes, it seems that the degree of overlap of item representations is very small, perhaps too small to generate the large increases in false alarms that are observed. Instead, it may be necessary to assume that people make use of a separate representation of the categories that appear on the list. Participants may not study each item in isolation. Rather, they may form a notion of what kinds of items are on the study list and use this in their recognition decisions, in which case one would expect to see category-based list length effects (as Criss & Shiffrin, 2004a, did for faces). Indeed, such a conclusion would be necessary to account for the inverse list length results reported by Dennis and Chapman (2010; see Maguire, Humphreys, Dennis, & Lee, 2010, for a similar conclusion).

The original global-matching models for recognition (Gillund & Shiffrin, 1984; Hintzman, 1984; Humphreys et al., 1989; Murdock, 1982) were abandoned because they predicted that there should be a list strength effect such that strengthening some items on a list would compromise performance on the other items. With words, a null list strength effect is observed (Ratcliff, Clark, & Shiffrin, 1990). However, with faces, a small list strength effect is found (Norman, Tepe, Nyhus, & Curran, 2008). The amount of interference that is predicted by the global-matching models depends on the degree of overlap of the item representations. The difficulty that the global-matching models face, then, is not necessarily structural, but rather a matter of parameterisation. If one is prepared to assume that for words, word pairs, and photographs of scenes the overlap between items is sufficiently small as to be negligible, but for faces and fractals this is not the case, one could capture the list strength and list length effects and how they change as a function of stimulus class.

While words typically show a low-frequency advantage (Glanzer & Adams, 1985), in Experiment 1 we found a high-frequency advantage for word pairs. This result is difficult for models, such as the Retrieving Effectively from Memory model (REM; Shiffrin & Steyvers, 1997), that propose that the typical low-frequency advantage derives from the discriminability of item features. In particular, if one models a word pair as the concatenation of vectors representing the individual words, then one would predict no effect.

Dennis and Humphreys (2001) proposed that recognition memory was primarily a context noise process and that this was the primary reason for the typical word frequency effect. The context noise approach suggests that while single words will typically have been seen many times before, unrelated word pairs like those employed in this study are unlikely to have been seen as units in many previous contexts, and therefore will not be subject to context noise. Consequently, there should be no effect. As such, the context noise approach also provides no insight into why there might be a high-frequency advantage.

It has been proposed that associative recognition involves a cued-recall component (Clark et al., 1993; Nobel & Huber, 1993). Participants could use one word to recall its partner on the list, and compare the recalled item to the other item of the test pair. Cued recall shows a high-frequency advantage, so one might suggest that the high-frequency advantage in associative recognition is a consequence of the underlying cued-recall process. However, this is unlikely, as cued recall is subject to list length effects (e.g., Tulving & Pearlstone, 1966), so if it were an integral part of associative recognition we should have observed a list length effect (though cued recall may be subject to the same influence from confounds as recognition memory). Rather, what the data here suggest is that the formation of a word pair representation, or the ability to re-form the same representation at test, is superior for high-frequency words.

## Conclusions

The majority of experiments on recognition memory have involved words, and almost all of the work on list length effects in recognition has used word stimuli. Words are semantically rich, well learned, and typically have been seen in many previous contexts. Previous failures to find list length effects thus might have been a consequence of the choice of words as stimuli.

The present work shows that this is only partially true. Fractals and faces do show small list length effects. However, word pairs and photographs of scenes do not show significant effects. The pattern of results suggests that item interference can play a role in episodic recognition, but that this role is minor and confined to items that are quite similar to each other.

## References

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104,* 54–75.

Bowles, N. L., & Glanzer, M. (1983). An analysis of interference in recognition memory. *Memory & Cognition, 11,* 307–315.

Brandt, M. (2007). Bridging the gap between measurement models and theories of human memory. *Zeitschrift für Psychologie, 215,* 72–85.

Buratto, L. G., & Lamberts, K. (2008). List strength effect without list length effect in recognition memory. *Quarterly Journal of Experimental Psychology, 61,* 218–226.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language, 49,* 231–248.

Chalmers, K. (2005). Basis of recency and frequency judgements of novel faces: Generalised strength or episode-specific memories? *Memory, 13,* 484–498.

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition, 20,* 231–243.

Clark, S. E., & Burchett, R. E. R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition, 22,* 55–62.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60. doi:10.3758/BF03210740

Clark, S. E., & Hori, A. (1995). List length and overlap effects in forced-choice associative recognition. *Memory & Cognition, 23,* 456–461.

Clark, S. E., Hori, A., & Callan, D. E. (1993). Forced-choice associative recognition: Implications for global-memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 871–881.

Clark, S. E., & Shiffrin, R. M. (1992). Cuing effects and associative information in recognition memory. *Memory & Cognition, 20,* 580–598.

Criss, A. H., & Shiffrin, R. M. (2004a). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review, 111,* 800–807. doi:10.1037/0033-295X.111.3.800

Criss, A. H., & Shiffrin, R. M. (2004b). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition, 32,* 1284–1297. doi:10.3758/BF03206319

Curran, T., Schacter, D. L., Norman, K. A., & Galluccio, L. (1997). False recognition after a right frontal lobe infarction: Memory for general and specific information. *Neuropsychologia, 35,* 1035–1049.

Dennis, S. (1995). *The Sydney Morning Herald word database.* Available at http://psy.uq.edu.au/CogPsych/Noetica

Dennis, S., & Chapman, A. (2010). The inverse list length effect: A challenge for pure exemplar models of recognition memory. *Journal of Memory and Language, 63,* 416–424.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* 452–478. doi:10.1037/0033-295X.108.2.452

Dennis, S., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language, 59,* 361–376.

Farah, M. J., Wilson, K.-D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105,* 482–498. doi:10.1037/0033-295X.105.3.482

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18,* 23–30. doi:10.3758/BF03202642

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20. doi:10.3758/BF03198438

Greene, R. (2004). Recognition memory for pseudowords. *Journal of Memory and Language, 50,* 259–267.

Gronlund, S. D., & Elam, L. E. (1994). List-length effect: Recognition accuracy and variance of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1355–1369.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16,* 96–101. doi:10.3758/BF03202365

Hockley, W. E. (1991). Recognition memory for item and associative information: A comparison of forgetting rates. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 227–248). Hillsdale, NJ: Erlbaum.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1321–1330.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory of episodic, semantic, and procedural tasks. *Psychological Review, 96,* 208–233. doi:10.1037/0033-295X.96.2.208

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340. doi:10.1037/0096-3445.110.3.306

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 112–127.

Kinnell, A., & Dennis, S. (2011). The list length effect in recognition memory: An analysis of potential confounds. *Memory & Cognition, 39,* 348–363.

Lauwerier, H. (1991). *Fractals: Images of chaos* (S. Gill-Hoffstadt, Trans.). Princeton, NJ: Princeton University Press. (Original work published 1987).

Maguire, A. M., Humphreys, M. S., Dennis, S., & Lee, M. D. (2010). Global similarity accounts of embedded-category designs: Tests of the global matching models. *Journal of Memory and Language, 63,* 131–148. doi:10.1016/j.jml.2010.03.007

Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision, 5*(1), 58–70.

Martínez, A., & Benavente, R. (1998). *The AR Face Database* (CVC Tech. Rep. #24). Barcelona: Universitat Autònoma de Barcelona, Centre de Visió per Computador.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626. doi:10.1037/0033-295X.89.6.609

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 855–874. doi:10.1037/0278-7393.17.5.855

Nobel, P. A., & Huber, D. E. (1993, August). *Modeling forced-choice associative recognition through a hybrid of global recognition and cued-recall.* Paper presented at the 15th Annual Meeting of the Cognitive Science Society, Boulder, CO.

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 384–413.

Norman, K. A., Tepe, K., Nyhus, E., & Curran, T. (2008). Event-related potential correlates of interference effects on recognition memory. *Psychonomic Bulletin & Review, 15,* 36–43.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156–163. doi:10.1016/S0022-5371(67)80067-7

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 267–287.

Shiffrin, R., Ratcliff, R., Murnane, K., & Nobel, P. (1993). TODAM and the list-strength and list-length effects: Comment on Murdock and Kahana (1993a). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1445–1449. doi:10.1037/0278-7393.19.6.1445

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166. doi:10.3758/BF03209391

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50. doi:10.1037/0096-3445.117.1.34

Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological Review, 19,* 447–462.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5,* 381–391.

Underwood, B. J. (1978). Recognition memory as a function of length of study list. *Bulletin of the Psychonomic Society, 12,* 89–91.

Weeks, C. S., Humphreys, M. S., & Hockley, W. E. (2007). Buffered forgetting: When targets and distractors are both forgotten. *Memory & Cognition, 35,* 1267–1282.

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 547–565.

Xu, J., & Malmberg, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition, 35,* 526–544. doi:10.3758/BF03193292