Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/19/$12.00

2019, Vol. 45, No. 5, 834–850
http://dx.doi.org/10.1037/xlm0000594

RESEARCH REPORT

Homogeneity of Item Material Boosts the List Length Effect in Recognition Memory: A Global Matching Perspective

Martin Brandt
University of Mannheim

Ann-Kathrin Zaiser
Saarland University

Martin Schnuerch
University of Mannheim

Kinnell and Dennis (2012) showed that the list length effect in recognition memory is only observed for homogeneous stimulus material. On the basis of the global matching model MINERVA 2 (Hintzman, 1986, 1988), we offer a theoretical explanation for this finding. According to our analysis, homogeneous material immunizes against the disruptive influence of preexperimental items, which might mask the intralist interference predicted by global matching models for familiar heterogeneous material. We tested our approach in three experiments. In Experiment 1, we found list length effects for homogeneous photographs of flowers and landscapes. In Experiment 2 and 3, we presented heterogeneous photographs of scenes (Experiment 2) and faces (Experiment 3). List length effects were only found if these photographs were homogenized by the use of image-processing filters. We further show that our explanation is also in line with the results of Dennis and Chapman (2010) who found an inverse list length effect. Overall, our results provide evidence for a global matching account of familiarity.

Keywords: global matching models, interference in memory, list length effect, recognition memory

Memory performance decreases with the length of the study list. Although this so-called list length effect, first documented by Strong (1912), has been demonstrated repeatedly (e.g., Brandt, 2007; Cary & Reder, 2003; Gillund & Shiffrin, 1984; Gronlund & Elam, 1994; Yonelinas, 1994), there is an ongoing debate on the underlying mechanisms in a list length study. A promising way to address these processes is the application of formal models of human memory. One class of formal models claims that every item stored in long-term memory contributes to the memory signal elicited by a test item. These models are referred to as global memory models, global matching models, or item noise models. Prominent members of this class are SAM (Gillund & Shiffrin, 1984), MINERVA 2 (Hintzman, 1986, 1988), TODAM2 (Murdock, 1993, 1997), the Matrix model (Humphreys, Bain, & Pike, 1989), and REM (Shiffrin & Steyvers, 1997). At the core of all these models lies the assumption that a greater number of items in memory increases the noise in the memory signal of test items and, therefore, memory performance declines with the number of items studied (Clark & Gronlund, 1996).

This article was published Online First July 9, 2018.

Martin Brandt, Department of Psychology, University of Mannheim; Ann-Kathrin Zaiser, Department of Psychology, Saarland University; Martin Schnuerch, Department of Psychology, University of Mannheim.

Correspondence concerning this article should be addressed to Martin Brandt, Department of Psychology, University of Mannheim, D-68131 Mannheim, Germany. E-mail: brandt@psychologie.uni-mannheim.de

In contrast to item noise models, context noise models such as the BCDMEM (Dennis & Humphreys, 2001) claim that item information is represented orthogonally and the only source of interference is the information about the episodic context in which the item was presented. Consequently, increasing the number of items in a study list should not change the memory signal for a test item and, hence, no list length effect is predicted. Indeed, although there is notable empirical support for the list length effect, there are also numerous studies showing a null list length effect (e.g., Buratto & Lamberts, 2008; Dennis & Humphreys, 2001; Kinnell & Dennis, 2012).

Having two model classes with different predictions on the list length effect seems to be the ideal situation for deciding between these conflicting models empirically. Unfortunately, this turns out to be quite complicated. Kinnell and Dennis (2011) showed that there are many potential confounds in the design of list length studies that could lead to inconsistent findings. Following their argumentation, an observed list length effect cannot necessarily be attributed to interference processes (as assumed by item noise models) but might occur because of certain specifics of the experimental design and the data-collection procedure. For example, in list length studies it naturally takes longer to study the long list compared to the short list. Consequently, if the duration of the retention interval is not controlled for, items in a short list will on average have a shorter retention interval than items in a long list. In this case, a list length effect might occur even without any additional interference processes, simply due to the longer retention interval for items studied in the long list. On the other hand, observing a null list length effect might always be associated with low statistical power because the effect size of the list length effect is generally small (e.g., Annis, Lenes, Westfall, Criss, & Malmberg, 2015). We will show that this is especially the case if the list length effect is studied in a within-subjects design. Moreover, there is a methodological controversy on how to analyze data from a list length experiment. Applying a Bayesian analysis to their experiments, Dennis, Lee, and Kinnell (2008) favor the context noise model for recognition experiments. This conclusion was challenged by Annis et al. (2015) who showed that the data of Dennis et al. (2008) do not allow for discriminating between the models.

Considering all these complications, the list length effect seems to be of little theoretical importance. However, there is a puzzling finding that might shed some new light on the controversy: Kinnell and Dennis (2012) studied the list length effect for different materials controlling for all potential confounds outlined by Kinnell and Dennis (2011). The authors found a small list length effect for pictures of faces and fractals, whereas no list length effect could be observed for word pairs and photographs of scenes. Discussing these results, Kinnell and Dennis (2012) admit 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" (p. 324). Thus, interitem similarity seems to be a necessary condition for the list length effect to occur. The BCDMEM (Dennis & Humphreys, 2001) would be compatible with these results if item representations were allowed not to be strictly local. In this case, the model could explain some interference on the level of item information because item representations would partly be overlapping. Pure item noise models, on the other hand, have to explain why there is no list length effect for verbal material and photographs of scenes and why the observed list length effect is rather small for faces and fractals. While Kinnell and Dennis (2012) do not offer such an explanation, we will provide one, based on the global matching model MINERVA 2 (Hintzman, 1986, 1988) as an example for an item noise model. We argue that our explanation of the findings of Kinnell and Dennis (2012) will give rise to some new insights concerning the conflicting results in the list length effect research in general.

We will first introduce the basic mechanisms of the MINERVA 2 model and elaborate on how the list length effect can be explained within this framework. We will then analyze what happens when study items are similar to each other, that is, stimulus material becomes more homogeneous. In a last step, we will demonstrate the effect of extralist items on heterogeneous and homogeneous item material, respectively. Based on these analyses, we will provide an explanation for the results of Kinnell and Dennis (2012) and we will test this new approach within a series of three experiments.

MINERVA 2: The Basic Model

In MINERVA 2, every item $\mathbf{T}$ is represented as a vector of M features with every feature $T_j \in \{-1, 0, 1\}$, $j = 1 \ldots M$ (Hintzman, 1986, 1988). According to Brandt (2007), we can further simplify the model assuming that $T_j \in \{-1, 1\}$, which allows for the introduction of a probability parameter s defining the similarity of two items $\mathbf{T}$ and $\mathbf{U}$ by giving the probability that two features in the same position of the item representations are identical, that is, $P(T_j = U_j) = s$. In the model, long term memory is conceptualized as a matrix $\mathbf{M}$ with N rows representing single episodes and M features representing the attributes in the item vector of a single episode. A new item $\mathbf{T}$ is learned by copying the attributes of the item into a new row i in the memory matrix $\mathbf{M}$ with the following rules: $P(M_{i,j} = T_j \mid T_j \neq 0) = L$, $P(M_{i,j} = 0 \mid T_j \neq 0) = 1 - L$, $P(M_{i,j} = -T_j \mid T_j \neq 0) = 0$, and $P(M_{i,j} = 0 \mid T_j = 0) = 1$. The parameter L is usually referred to as a learning parameter. In memory retrieval, the memory matrix $\mathbf{M}$ is probed with a test item $\mathbf{T}$ and two memory signals arise: In the echo-intensity process, the test item $\mathbf{T}$ activates every memory representation (i.e., every row in $\mathbf{M}$) in parallel and the amount of activation is a function of the similarity of the test item to the memory representation:

$$A_i(\mathbf{T}) = \left[\frac{1}{M}\sum_{j=1}^{M}(T_j \cdot M_{i,j})\right]^3 \tag{1}$$

Most important and central to the model is the assumption that these local activations $A_i(\mathbf{T})$ are added up to a global response to the retrieval cue:

$$A(\mathbf{T}) = \sum_{i=1}^{N} A_i(\mathbf{T}) \tag{2}$$

$A(\mathbf{T})$ is referred to as *echo intensity* and in recognition tests it is usually interpreted as memory strength or a feeling of familiarity. In old/new recognition tests, a decision rule borrowed from signal detection theory is applied (e.g., Macmillan & Creelman, 2008): A response criterion c is postulated. If $A(\mathbf{T}) > c$, an "old" response is given, otherwise participants respond with "new" to the test item $\mathbf{T}$. In a two-alternative forced choice test, two items $\mathbf{T}$ and $\mathbf{U}$ are presented and participants are asked to select the item that was presented in a preceding study phase. In this case, no response criterion is necessary. If $A(\mathbf{T}) > A(\mathbf{U})$, participants select item $\mathbf{T}$, otherwise they select item $\mathbf{U}$.

In addition to the echo-intensity process, there is a second retrieval process in MINERVA 2. When memory is probed with an item $\mathbf{T}$, this so-called echo-content process results in a reconstructed vector $\mathbf{C}$ defined as the sum over all attributes of all items represented in memory, weighted with the local activations of the memory representations as defined in equation (1):

$$C_j(\mathbf{T}) = \sum_{i=1}^{N} M_{i,j} \cdot A_i(\mathbf{T}). \tag{3}$$

In applications of MINERVA 2 to recognition experiments, it is usually assumed that only the echo-intensity process is relevant. Nevertheless, one should keep in mind that MINERVA 2 is not a single-process model, as the echo-content process could easily be integrated when it comes to the modeling of recognition experiments. For example, it would be possible to differentiate between item- and context-specific features of an item (e.g., the REM.4 model in Shiffrin & Steyvers, 1997). If memory is probed only with item-specific attributes, the echo-content process may be used to retrieve the context-specific attributes of the study phase. This approach would resemble the context-retrieval process assumed in the BCDMEM (Dennis & Humphreys, 2001). Despite the possibility to integrate a second process when modeling recognition experiments with MINERVA 2, in this paper we keep with the common assumption that only the echo-intensity process is involved in recognition.

List Length Effect in the Basic Model

To model list length experiments, some simplifying assumptions need to be made. We start with the basic model as it is usually applied to list length experiments in recognition memory (e.g., Clark & Gronlund, 1996; Gronlund & Elam, 1994; Hintzman, 1988) and add some elaborations of the basic model in the following sections.

Let us assume that there are N_1 items to be learned in the short list and N_2 items to be learned in the long list, each of which has M attributes that are drawn randomly and independently from $\{-1, 1\}$. Note that this is equivalent to the assumption that for any two items **T** and **U** the similarity parameter equals $s = .5$. We further assume that all these items are learned with a fixed learning parameter L.

Furthermore, we assume that the memory matrix **M** is empty before the study phase. Consequently, after the study phase, **M** contains exactly N_1 or N_2 memory representations, depending on which list was learned. In the test phase, we either probe memory with a learned item (i.e., a target item) or a new randomly and independently drawn item (i.e., a distractor). Brandt (2007) presented closed-form equations for $E[A(\mathbf{T})]$ and $Var[A(\mathbf{T})]$ such that it is easy to compute the memory performance predicted by the model, given the parameters L, N_1, N_2, and M. Typically, the number of attributes M is fixed to a sufficiently large value, which will be 40 throughout this paper. Because it is assumed that only items from the study phase are entered into the memory matrix **M**, memory performance is solely determined by the learning parameter L and the number of items in the study phase, that is, N_1 and N_2, respectively. For any given length N of a study list, the expected value of the echo-intensity distribution of a target item **T** is

$$E[A(\mathbf{T})] = \sum_{i=1}^{N} E[A_i(\mathbf{T})] = E[A(\mathbf{T}|s = 1)] + (N-1)E[A(\mathbf{T}|s = .5)] \tag{4}$$

where $E[A(\mathbf{T}|s = 1)]$ is the expected value of the local activation of the target item's memory representation (cf., equation 1) and $E[A(\mathbf{T}|s = .5)]$ is the expected value of the local activation of any noncorresponding memory representation. Because $E[A(\mathbf{T}|s = .5)] = 0$, $E[A(\mathbf{T})]$ only depends on the local activation of the memory representation of the target item which is an increasing function of L. On the other hand, the expected value of the echo-intensity distribution of distractors is always 0, as there are no corresponding memory representations for distractor items. Consequently, expected values of both the target and the distractor distributions are independent of the length of the study list.

The variance of the echo-intensity distribution for an item T is defined as:

$$VAR[A(\mathbf{T})] = \sum_{i,j=1}^{N} COV[A_i(\mathbf{T}), A_j(\mathbf{T})] \tag{5}$$

Because for independent items (i.e., $s = .5$) all covariances in equation (5) are 0 for $i \neq j$, the variance of the echo intensity is simply the sum of the variances of all local activations. For nontrivial cases (i.e., $L \neq 0$ or $L \neq 1$) these variances are positive. Hence, the variances of the echo-intensity distributions for both target and distractor items increase with the number of items in the study list. As a consequence, we have constant expected values and increasing variances as a function of the number of studied items. Thus, both distributions increasingly overlap with increasing list length, and memory performance decreases (see Gronlund & Elam, 1994, for a numeric example). In other words, MINERVA 2 predicts a list length effect in this situation.

Homogeneity of the Stimulus Material

Kinnell and Dennis (2012) found the list length effect only for relatively homogeneous material, so it is important to look at the predictions derived from MINERVA 2 for homogeneous material. Homogeneous study lists can be modeled in the same fashion as word lists in a DRM paradigm (e.g., Roediger & McDermott, 1995) are constructed: We start by drawing a prototype item **Q** that will not be part of the study list. In the next step, we draw N items **R** with the restriction

$$P(Q_j = R_j) = h \text{ for all } j \in M \tag{6}$$

where Q_j and R_j represent corresponding attributes in the two items. The parameter h defines the similarity of an item to the prototype and therefore is a measure of homogeneity of the study list. The basic model is a special case with $h = .5$, that is, all items are independent. Note that the similarity of two specific items $\mathbf{R}_1$ and $\mathbf{R}_2$ is given by $s = h^2 + (1 - h)^2$. Homogenizing the list influences both the expected values and variances in a list length paradigm. The expected value of the echo-intensity distribution for targets is defined as:

$$E[A(\mathbf{T})] = E[A(\mathbf{T}|s = 1)] + (N-1)E[A(\mathbf{T}|s = h^2 + (1-h)^2)] \tag{7}$$

Because $E[A(\mathbf{T}|s = h^2 + (1 - h)^2)]$ is greater than 0 for $s > .5$, it is clear that the overall expected value of a target distribution increases with N. The same is true for the distractor distribution. Nevertheless, the difference between the expected values of the target and the distractor distribution is independent of the number of items in the study list: It is always $E[A(\mathbf{T}|s = 1)] - E[A(\mathbf{T}|s = h^2 + (1 - h)^2)]$ and thus remains independent of list length.

The variances of both target and distractor distributions are also heavily influenced by the parameter h. Not only the variances of local activations increase with h, but also their covariances now become positive (for $h > .5$). These covariances additionally increase the variances of the echo-intensity distributions of targets and distractors. With respect to the list length paradigm, the basic mechanism remains the same: The differences between the expected values of target and distractor distributions are constant for different lengths of study lists, but longer lists lead to greater variances. In short, the basic mechanism for the list length effect is independent of the homogeneity of the stimulus material. However, the effect can be slightly more pronounced because there are now $N \cdot (N - 1)$ covariances of local activations that add to the variances of the echo-intensity distributions of both targets and distractors (cf., equation 5). Nevertheless, this small additional increase in the overlap of the distributions cannot explain why

Kinnell and Dennis (2012) found the list length effect only for homogeneous material.

Influence of Extralist Items

In the basic model, we assumed that only items of the current study list are represented in memory. This somewhat naive assumption might be justified, considering that an episode is not only composed of item information but also of the episodic context the item is presented in. If the learning context is different from all other previous contexts and it is available as a retrieval cue, the features of the learning context can effectively reduce the influence of previously encountered episodes (Hintzman, 1986). On the other hand, in a typical memory experiment, both the items studied (e.g., single words or pictures of objects) as well as the context (e.g., item presentation on a computer screen) are not really unique, especially if several lists are learned in a within-subjects design. Nevertheless, list discrimination experiments show that the context even slightly changes if several lists are learned in one session (e.g., Hintzman & Waters, 1970). On the other hand, if the learning context is very special (as e.g. in the famous underwater learning experiments by Godden & Baddeley, 1975), preexperimental episodes can probably be ignored because they do not match this special learning context. Nevertheless, in many experiments it seems plausible to assume that similar episodes prior to the study list contribute to the retrieved memory signal in global memory models. This extension of the basic model has been discussed under the label *continuous memory* (e.g., Murdock & Kahana, 1993), the influence of extralist items (e.g., Gronlund & Elam, 1994), or recently, background noise (Osth & Dennis, 2015). Modeling the list length paradigm in MINERVA 2, the influence of extralist items can easily be implemented by adding N_e independent items to the memory matrix **M**. Because these extralist items were studied earlier, it seems reasonable to assume that some attributes might have been forgotten. In the model, this is equivalent to assuming a smaller learning parameter (L_{extra}) for extralist items.

Note that taking into account extralist items is similar to the concept of context noise in the BCDMEM (Dennis & Humphreys, 2001). This model also assumes that targets have been presented in different contexts before the learning episode. After learning, it is assumed that participants try to retrieve the learning context given the item under study as a retrieval cue to compare it with the reinstated context for a recognition decision. The retrieved context is a composite of the preexperimental contexts and the actual learning context. In that sense, previously encountered items add noise to the retrieved context because they might have been learned in different contexts. In MINERVA 2, the memory signal is also distorted by the previous targets learned in different contexts. But additionally, the memory signal for a test item is also influenced by other items presented in the same context.

Given that extralist items are independent of the items in the study list and that the item material is heterogeneous (i.e., $h = .5$), the effect of taking into account extralist items is straightforward: Every single extralist item adds a small amount to the variances of both the target and the distractor distributions without affecting the corresponding expected values. Thus, extralist items impair the discriminability of targets and distractors and, more importantly, dramatically reduce the list length effect in MINERVA 2. To illustrate the latter, remember that in MINERVA 2 the list length effect is considered a result of a greater increase of variances of the echo-intensity distributions in the long-list condition compared to the short-list condition. If the number of extralist items is sufficiently high, this difference between the variances vanishes. In other words: Without extralist items, additional items in the long-list condition have a relatively great impact on the variances of the underlying distributions. But with an increase in the number of extralist items, the impact of a relatively small number of additional items in the longer study list is negligible and no list length effect is predicted.

Let us have a look at the impact of extralist items on homogeneous stimulus material. At a first glance, the effect seems identical to the case of heterogeneous material: Every single extralist item adds a small amount to the variances of echo-intensity distributions. But this impact is much smaller compared to the impact on heterogeneous material because both the target and the distractor distributions have much greater variances even without the effect of extralist items due to the influence of covariances. To be precise, given these greater variances, the *relative* impact of extralist items is much smaller. Thus, assuming a reasonable amount of extralist items, the list length effect is still predicted for homogeneous material, as illustrated in Figure 1.

In short, using homogeneous item material immunizes the model predictions against the disturbing influence of extralist items. However, note that as the number of extralist items becomes very large, MINERVA 2 no longer predicts a list length effect even for homogeneous material. Also, the prediction only holds if extralist items are unrelated to the study items. If homogeneous item material is used in the study list, extralist items from the same category nevertheless have a great impact on the variances of the underlying echo-intensity distributions. This scenario might be quite common in within-subjects designs. If all items stem from a common category, items from previously studied lists might interfere with the current list.

In summary, there are three sources of interference in a global matching model. First, there is interference caused by items in the actual study list, which may be labeled *intralist interference*. Second, we may also consider interference caused by items from lists previously presented in the experiment, which may be referred to as *interlist interference*. This source of interference is only important in within-subjects designs where all items stem from the same homogeneous category. Third, extralist items, that is, items from all previously encountered episodes, might inflate the variances of the underlying distributions and, hence, mask the list length effect. As we have outlined above, the influence of this *extralist interference* can be reduced by using homogeneous item material.

Given these three sources of interference, we can now offer an explanation for the effect of the stimulus material reported by Kinnell and Dennis (2012) from a global matching perspective: Both word pairs (Exp. 1) and photographs of scenes (Exp. 4) are very familiar material and thus should be prone to the influence of preexperimental encounters.[1] Therefore, the null list length effect

[1] Note that only individual words can be assumed to be familiar, whereas word pairs might have never been learned in common context and the associative information might even be unique. It is an open question how the preexperimental frequency of individual words influences associative recognition in a list length paradigm. We thank Amy H. Criss for this remark.

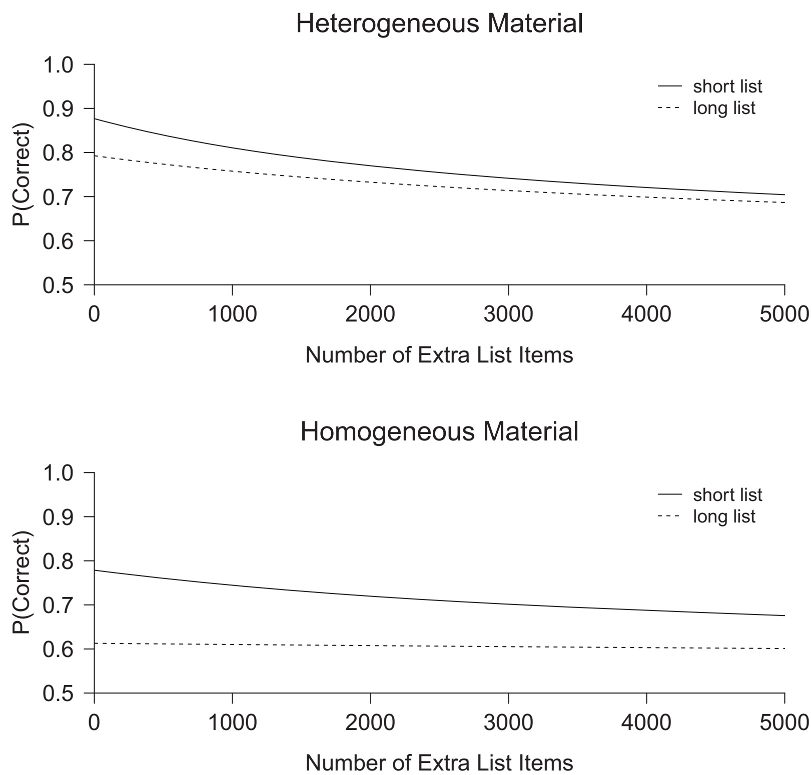

Figure 1. Influence of extralist items on the list length effect on heterogeneous (top graph) and homogeneous (lower graph) material in MINERVA 2. Note that with homogeneous material (a) performance decreases, (b) the list length effect is more pronounced, and (c) the influence of extra list items is reduced. Model parameters are: $M = 40$, $L = .3$, N(short list) = 20, N(long list) = 80, h(heterogeneous) = .5, h(homogeneous) = .8. Extralist items are assumed to be heterogeneous, independent from the list items, and learned with $L_{extra} = .1$.

for this type of material could eventually be attributed to the strong influence of extralist interference. In Experiment 2, Kinnell and Dennis (2012) used pictures of faces which is also very familiar material and should therefore lead to problems with extralist interference. However, contrary to the word and scene material, pictures of faces are relatively homogeneous, such that the impact of extralist items is reduced and a list length effect is expected, just as it was observed. In Experiment 3, fractals were used and a strong list length effect emerged, because fractals are unfamiliar objects one certainly does not come across very often outside an experimental context. Moreover, because fractals are highly homogeneous, no disturbing influence of extralist interference is expected. Therefore, a list length effect for fractals is also expected from a global matching perspective.

An aspect that remains critical for the global matching approach is that the observed effects (Exp. 2 and 3) are relatively small compared with predictions from model simulations. We argue, however, that the small impact of list length on recognition performance is attributable to interlist interference. Because Kinnell and Dennis (2012) always used a within-subjects design, that is, presenting participants with a long and a short study list in a first and a second test block, respectively, performance in the second block might be influenced by length of the study list in the first block. Note that for homogeneous material, the influence of extralist items in the first block should be negligible. This is not true, however, for the second block because participants already studied items from the same homogeneous category in the first block. Studying a short list in the second block implies that participants have studied a long list in the first block, and vice versa. Thus, if the short-list condition is in the second block, there is a stronger impact of interlist interference compared with the long-list condition. As a consequence, only a small or even a null effect is to be expected in the second block. By analyzing the within-subjects design, we average over a large (dependent on the concrete number of studied items) effect in the first block and a small or even nonexisting effect in the second block. This results in an overall smaller effect size compared to between-subjects designs, which are usually used in simulations. Thus, from a global matching perspective, a relatively small effect size for homogeneous item material would be expected in a within-subjects design. However, we predict a large effect when only the first block is analyzed in the sense of a between-subjects design and a small or a null effect when analyzing only the second block in a between-subjects manner. Note that this specific prediction is bound to the assumption of a global matching perspective. In the BCDMEM (Dennis & Humphreys, 2001), neither extralist items nor interlist items should interfere with the items. Thus, the separate analysis of the first and the second block is a further differential test of both types of models.

In three experiments, we will test these theoretical explanations for the findings of Kinnell and Dennis (2012). In Experiment 1, we use pictures as stimulus material. But unlike Kinnell and Dennis (2012, Exp. 4), we specifically use homogeneous pictures (i.e., pictures of certain landscapes and flowers, respectively). We expect a small list length effect that is mainly driven by performance in the first block. In the second experiment, we use the same material for both conditions but directly manipulate the degree of homogeneity. In one condition, we use heterogeneous pictures of scenes and expect no list length effect. In the other condition, we use exactly the same pictures but homogenize them by applying special image-processing filters, therefore expecting a list length effect to occur. In Experiment 3, we use photographs of faces. For this material, Kinnell and Dennis (2012) found a list length effect. In contrast to them, however, we use very heterogeneous faces (i.e., different gender, emotional expressions, ages, races, etc.), which is why we expect no list length effect for this material. Applying the same image-processing filters as in Experiment 2, we again homogenize the pictures and, in turn, expect a list length effect.

Figure 2. Example of the material used in Experiment 1. Stimuli were either photographs of landscapes (A) or photographs of blossoms (B). See the online article for the color version of this figure.

Experiment 1

In Experiment 1, we refer to Experiment 4 reported by Kinnell and Dennis (2012). The authors found no list length effect for photographs of everyday scenes. Because these photographs were rather heterogeneous (the authors list a classroom, a library, and a beach), we wanted to investigate whether the list length effect can be found with more homogeneous pictures. To do so, we used two stimulus types, that is, photographs of landscapes and photographs of flowers. Both types of stimuli are not exactly pictures of scenes but they clearly are homogeneous, such that a list length effect should emerge for this type of material.

Method

Participants. In total, 50 people participated in Experiment 1, of which 88% were female. The mean age of participants was 21.84 years, ranging from 18 to 56 years. Almost all participants were students of psychology at the University of Mannheim and received course credit in return for their participation.

Materials. Overall, participants were presented with 170 photographs of landscapes and 170 photographs of flowers. All landscape photographs showed a horizon, but no clearly distinguishable, concrete objects, such as people, animals, or buildings, were depicted. The pictures of flowers did not show the whole plants but single blossoms on a rather monochrome background instead. Figure 2 shows examples of both stimulus types. Pictures in the learning phase were presented in 400×400 pixels size. In the two-alternative forced choice test, pictures were scaled down to 300×300 pixels.

Design. We used a $2 \times 2 \times 2 \times 2 \times 2$ factorial design with the factors List Length (short vs. long), Material (landscape vs. flower), Order Material (landscape-flower vs. flower-landscape), Order Landscapes (short-long vs. long-short), and Order Flowers (short-long vs. long-short). The factors List Length and Material were manipulated within subjects. Dependent variables were the proportion of correct answers in a two-alternative forced choice recognition test and the respective response latencies. All factor levels were randomly assigned to participants.

Procedure. To make participants familiar with the tasks of the experiment, they passed a short practice session at the beginning. For the practice memory task, we used 20 close-up photographs of everyday objects (e.g., a bottle) to minimize interference with the photographs of landscapes and flowers. In the practice session, participants studied 10 photographs. After learning, a perceptual discrimination task was implemented as a filler task. This task was the same as in the actual experiment. Finally, recognition performance was tested using a two-alternative forced choice task. For every trial, a photograph of a learned item was paired with a photograph of a distractor item. Both items were shown side by side and subjects indicated with the arrow keys which item was studied in the learning phase. Positions of the target and distractor items were determined randomly for every trial. Participants could repeat the practice session until they were familiar with the procedure.

After the practice session, participants worked through four blocks, each consisting of a study phase, the perceptual discrimination filler task, and the two-alternative forced choice recognition task. Participants were always presented with one type of material in the first two blocks (i.e., either pictures of blossoms or landscapes) and the other one in the last two blocks. Within one type of material, the order of list lengths was randomly determined. Short lists consisted of 20 and long lists of 80 photographs. In the study phase, items were presented with a 200-ms interstimulus interval for 900 ms, during which the screen was left blank. The filler task following long study lists lasted at least 30 seconds. After short study lists, the filler task lasted at least 90 seconds in order to compensate for the enlarged study phase for long lists. In the two-alternative forced choice recognition task, the first 20 items of each study list were tested in random order. For long lists, we tested 10 randomly selected additional items from the study list after testing the first 20 items. This was done to mask the fact that the last items in the long study list were not of interest, and avoid confounding effects. These 10 additional items were not included in the analysis.

Participants worked through the recognition task at their own pace. After each block, there was a break without any filler activity for one minute.

Results

Analyzing the full five-factorial design we found that landscapes (M = .80) were better remembered than flowers (M = .76), $F(1, 42) = 10.52$, $p < .01$, $\eta_p^2 = .20$. This difference in performance presumably reflects the fact that pictures of blossoms are more homogeneous and therefore harder to differentiate. The main effect of Order Material was not significant, $F(1, 42) = 3.65$, $p >$ .05. Because all interactions of the factors Material and Order Material with the critical independent variable List Length were not significant, we decided to analyze the different types of material separately.

Landscapes.

Within-subjects analysis. We first analyzed the subset of data with the within-subjects factor Length and the between-subjects factor Order. Overall recognition performance was slightly better in short lists (M = .82) compared to long lists (M = .78), $F(1, 48) = 7.81$, $p < .01$, $\eta_p^2 = .14$. The main effect of Order Landscapes was not significant, $F(1,48) = 1.75$, $p = .19$. Importantly, the interaction Length × Order Landscapes was significant, $F(1, 48) = 12.87$, $p < .01$, $\eta_p^2 = .21$. This interaction is not surprising if the role of interlist interference is taken into account: Let us first look at the condition where the short list is learned first. Here, we expect small intralist interference and no interlist interference, as there is no preceding study list. For the long list, which is learned afterward, we expect small interlist interference and large intralist interference. Thus, in this condition the interlist interference adds to the intralist interference and the list length effect should be strong. If the long list is learned first, there is large intralist interference for the long list and, again, no interlist interference. For the short list, however, there is small intralist interference but large interlist interference because the long list was studied beforehand. Accordingly, the interlist interference in this condition might mask differences in intralist interference between short and long lists, and the list length effect should be at least reduced. As can be seen from Table 1, this is exactly the pattern observed in Experiment 1 (see also Figure 3, for a graphical representation). There is a rather huge list length effect if the short list is studied first, $t(44) = 3.30$, $p < .01$, $d = 0.97$, and no list length effect if the long list is studied first, $t(52) = -0.54$.

Between-subjects analysis. As noted before, in a within-subjects analysis there is no unbiased test for the critical effect of intralist interference. Therefore, we followed Kinnell and Dennis (2012) and also analyzed the data in two separate between-subjects analyses for the first and the second block, respectively. Because there is no interlist interference in the first block, only extralist interference and intralist interference should affect performance. Moreover, because—according to MINERVA 2—homogeneous material should minimize the influence of extralist variance, we expect a list length effect due to intralist variance in the first block. In the second block, however, interlist interference works against the list length effect, that is, for short lists, intralist interference is low and interlist interference is high, while the reverse pattern applies for long lists. Beginning the recognition test in Block 2, the number of previously studied items (i.e., in both blocks) is identical for both list length conditions within a block. Therefore, we expect no list length effect in the second block. Looking at Table 1, this means comparing data from Columns 1 and 4 for the first block and Columns 2 and 3 for the second block, respectively (see also Figure 3). As expected, we found a large list length effect for the first block, $t(48) = 2.86$, $p < .01$, $d = 0.81$, and no list length effect for the second block, $t(48) = -0.56$.

Response latencies. Although participants were not instructed to respond as fast as possible, it is important to analyze response latencies in interference paradigms because a potential decrease in performance might not be attributable to interference processes but the result of a speed–accuracy trade-off. Response time analyses showed that this interpretation can be ruled out. The means of the individual median response times are shown in Table 2. The results were in line with the accuracy analysis: There was a main effect of List Length, $F(1, 48) = 14.20$, $p < .01$, $\eta_p^2 = .23$, showing that response times for short lists (M = 1633.80 ms) were faster than for long lists (M = 1811.29 ms). The significant interaction of the factors List Length and Order, $F(1, 48) = 4.11$, $p < .05$, $\eta_p^2 = .08$, demonstrated that the difference between short and long list was more pronounced if the short list was studied first. These results are perfectly in line with results from the accuracy analysis. The main effect of Order Landscapes was not significant ($F < 1$).

Flowers.

Within-subjects analysis. The mean portions of correct answers for the flowers subset are shown in Table 1. Performance for short lists (M = .77) was only numerically better than performance for long lists (M = .74), $F(1, 48) = 1.99$, $p = .17$. More impor-

Table 1

Mean Percent Correct Answers (and Standard Errors) in Two-Alternative Forced Choice Recognition Tests in Experiments 1–3 Depending on Material, Length of Lists, and Order of the List Lengths Studied

		Order of lists			
		Short–Long list length		Long–Short List Length	
Experiment	Materials	Short	Long	Short	Long
Exp. 1	Landscapes	.87 (.015)	.79 (.015)	.78 (.014)	.77 (.011)
	Flowers	.84 (.018)	.75 (.025)	.71 (.030)	.73 (.018)
Exp. 2	Homogeneous scenes	.89 (.015)	.85 (.017)	.86 (.024)	.83 (.027)
	Heterogeneous scenes	.91 (.015)	.91 (.017)	.92 (.011)	.90 (.014)
Exp. 3	Homogeneous faces	.90 (.021)	.84 (.021)	.81 (.027)	.84 (.022)
	Heterogeneous faces	.94 (.011)	.90 (.018)	.91 (.018)	.94 (.017)

Note. Standard errors were estimated between subjects. For repeated measurement analyses the corresponding error term might be smaller.

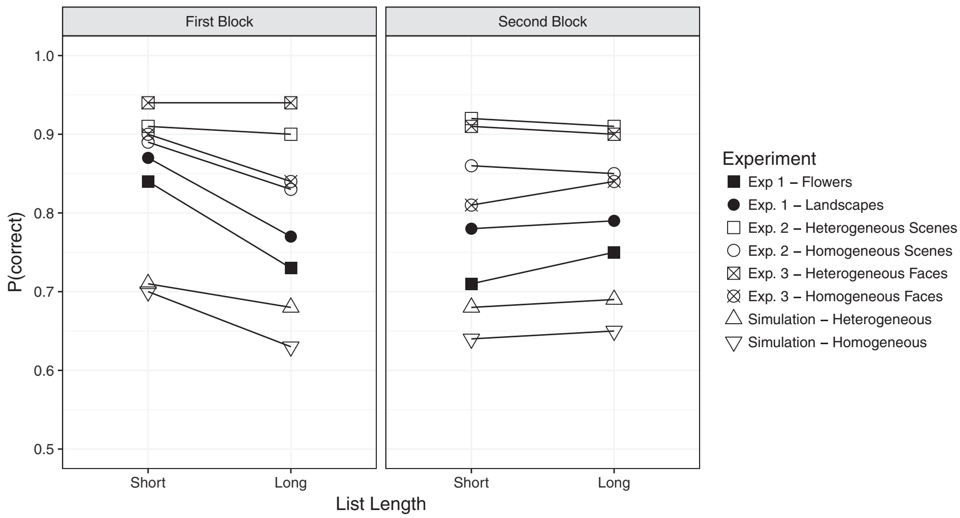

Figure 3. Summary of the results of Experiment 1–3 and a simulation of MINERVA 2. The data of the simulation show the means of 500 simulated experiments with 80 participants respectively and the same design as in Experiments 2 and 3. The parameters for the simulation were $M = 40$, $L(\text{items}) = .3$, $L(\text{extra items}) = .1$, $N(\text{short list}) = 20$, $N(\text{long list}) = 80$, $h(\text{heterogeneous}) = .5$, $h(\text{homogeneous}) = .8$, and $N(\text{extra list}) = 5000$. Note that there is a pronounced list length effect only for homogeneous material in the first block in both the empirical data as well as in the simulation. In the second block, there is no list length effect irrespective of the type of study material.

tantly, we again found a significant List Length × Order interaction, $F(1, 48) = 6.37$, $p < .05$, $\eta_p^2 = .12$. Post hoc tests showed that this interaction means that there is a clear list length effect when the short list is studied first, $t(46) = 2.62$, $p < .05$, $d = 0.73$, but not when the long list is studied first, $t(46) = -0.66$. Note that considering the effect of interlist interference, this pattern is again exactly in line with the predictions made by MINERVA 2. In contrast to the pictures of landscapes, with pictures of flowers we found a significant main effect of Order, $F(1, 48) = 8.45$, $p < .01$, $\eta_p^2 = .15$, showing that the overall performance was better when the short list was studied first ($M = .79$), compared with the condition where the long list was studied first ($M = .72$). We had not expected to find this main effect. However, it could be attributed to a decrease in motivation after studying the long list with flowers. Because this is the most difficult condition in the experiment, subjects might have become a little frustrated and slightly less engaged in the following list.

Between-subjects analysis. As pointed out above, the most sensible analysis to test for a list length effect is to look at the first block only. For the flowers, we found a strong list length effect in the first block, $t(48) = 4.09$, $p < .01$, $d = 1.16$, with better performance in the short list ($M = .84$) than in the long list ($M = .73$). In the second block, intralist interference was masked by interlist interference, caused by items encountered during the first block. Consequently, we found no list length effect in the second block, $t(48) = -1.17$, $p = .25$.

Response latencies. Response latencies for the flowers may be obtained from Table 2. There was a significant effect of list length,

Table 2
Mean of the Median Response Latencies (and Standard Errors) in Milliseconds in Two-Alternative Forced Choice Recognition Tests in Experiments 1–3 Depending on Material, Length of Lists, and Order of the List Lengths Studied

		Order of lists			
		Short–Long list length		Long–Short List Length	
Experiment	Material	Short	Long	Short	Long
Exp. 1	Landscapes	1603.39 (101.07)	1888.57 (116.92)	1659.70 (59.60)	1745.46 (99.56)
	Flowers	1614.20 (86.52)	1704.48 (90.04)	1571.57 (94.38)	1682.54 (71.92)
Exp. 2	Homog. scenes	1683.03 (70.47)	1763.49 (47.01)	1660.65 (84.92)	1734.78 (78.87)
	Heterog. scenes	1403.94 (61.66)	1419.25 (48.70)	1373.09 (53.26)	1496.20 (53.11)
Exp. 3	Homog. faces	1470.08 (71.43)	1829.10 (95.90)	1847.35 (128.20)	1789.02 (98.87)
	Heterog. faces	1285.93 (68.80)	1492.82 (74.84)	1333.67 (55.72)	1252.63 (43.42)

Note. Standard errors were estimated between subjects. For repeated measurement analyses the corresponding error term might be smaller.

$F(1, 48) = 5.85$, $p < .05$, $\eta_p^2 = .11$, with faster responses for short lists ($M = 1591.18$ ms) compared to long lists ($M = 1692.63$ ms). The effect of Order Flowers and the interaction of Order Flowers and List Length were not significant ($Fs < 1$). Therefore, a speed–accuracy trade-off can again be ruled out.

Discussion

Overall, the pattern found in Experiment 1 is quite consistent and in line with predictions of MINERVA 2. First, analyzing the within-subjects data, there is only a small list length effect for the landscape data and no list length effect for the flower data. In some sense, this pattern reflects the ambiguous situation found in the literature concerning the list length effect. Although the analysis of a within-subjects design is quite common, from a global matching perspective it might be misleading. This is because there should be an influence of interlist interference in the second list. Taking interlist interference into account, one would expect a list length effect when the short list is studied first, but no or at least a reduced list length effect when the long list is studied first. This is exactly the pattern we found for the landscape and the flower pictures. Interestingly, the effect of order of list lengths is usually not reported at all in list length experiments.

An unbiased view on the effect of intralist interference is obtained if one looks at the first block for each type of material only. Assuming that there is only negligible interference between flowers and landscapes, there cannot be any interlist interference in the first blocks. Moreover, from a global matching perspective, interference from preexperimental items (i.e., extralist items) should be small for homogeneous material like landscapes and flowers. Indeed, we found a strong and consistent performance advantage for short lists. The size of the list length effect in the first block is comparable to the effect size predicted by MINERVA 2 (see Figure 1).

If the second block is analyzed separately, predictions of MINERVA 2 change dramatically. For the short list, there is a small amount of intralist interference and a large amount of interlist interference. For the long list, MINERVA 2 predicts a large amount of intralist interference and a small amount of interlist interference. Given that the context of the first and the second list is highly similar (as it is the case in our study and in most other studies), intralist and interlist interference should add to a comparable amount of interference for both short and long list in the second trial. Therefore, no list length effect is expected to be found in the second trial, which is exactly the pattern we found in Experiment 1.

The third source of interference discussed is extralist interference, that is, the influence of preexperimental learning episodes. We argued that from a global matching perspective, the influence of extralist interference is reduced for homogeneous material, and the critical effect of intralist variance can be observed. The fact that we observed a list length effect in the first trial can be seen as indirect evidence for this line of argumentation. Nevertheless, it is clear that the homogeneity of the material can only reduce the influence of memory representations that do not stem from the same type of stimuli. The missing list length effect in the second blocks shows that learning the same type of material in preceding blocks has a huge impact on performance in the current block. Obviously, this interlist interference is a special kind of extralist interference, which cannot be reduced by the homogeneity of the material.

Although all results of Experiment 1 are in line with the predictions of a global-matching model, we still cannot be sure that homogeneity is the critical variable that minimizes the influence of extralist interference, because we used material that is homogeneous according to some face validity only. Therefore, we want to explicitly manipulate the homogeneity of the stimulus material in Experiment 2.

Experiment 2

In Experiment 2 we wanted to further test the impact of extralist items on homogeneous and heterogeneous item material. To support our notion that homogeneity is the crucial variable underlying the occurrence of the list length effect, we directly manipulated the homogeneity of our material. The experiment was again based on Experiment 4 by Kinnell and Dennis (2012). Using a rather heterogeneous set of photographs of scenes, the authors found no list length effect. We wanted to replicate this finding in one condition using heterogeneous material as well. In the other condition, however, we modified exactly the same photographs using image-processing filters, leading to an increase of the similarity between the pictures, and therefore an increase of the homogeneity of our material. Using such homogenized material, we expected to find a list length effect, particularly when only the first block is analyzed, which is not affected by interlist interference. For the second block, we did not expect a list length effect for either the heterogeneous or the homogeneous scenes. See Figure 4 for the model predictions based on a simulation of MINERVA 2.

Method

Participants. Overall, 123 participants took part in Experiment 2, of which 80.5% were female. Mean age of participants was 22.02 years, ranging from 18 to 44 years. Almost all participants were students of psychology at the University of Mannheim and received course credit in return for their participation. Four sub-

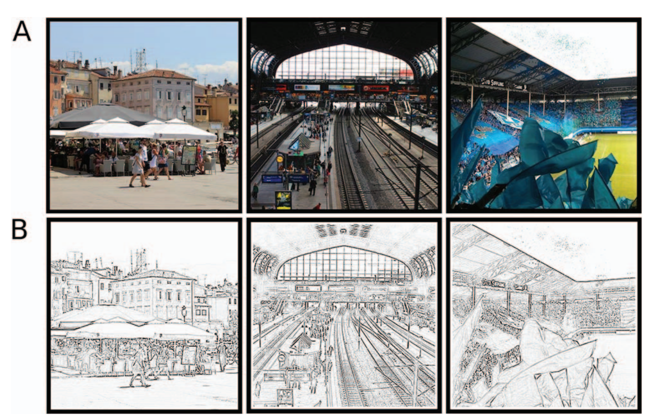

Figure 4. Example of the material used in Experiment 2. Stimuli were photographs of scenes (A) in the heterogeneous condition. In the homogeneous conditions (B) the same pictures were used, but the photographs were experimentally homogenized by applying image-processing filters. See the online article for the color version of this figure.

jects were excluded from analysis because their memory performance did not differ significantly from chance level. Because overall performance was very high, it is likely that these participants did not follow task instructions.

Materials. The stimuli were 150 photographs showing a wide variety of scenes, for example, traffic jam, a swimming hall, a market, or a fitness center, to name just a few. In the heterogeneous condition, these photographs were left in their original shape. In the homogeneous condition, the same pictures were shown, but this time we applied the *pencil sketch* and the *high dynamic range* image-processing filters implemented in the software Picasa to all images. As a consequence, all pictures looked more similar to each other compared with the heterogeneous condition, although they still displayed exactly the same content. Figure 4 shows examples of the stimuli in the heterogeneous and homogeneous condition. As in Experiment 1, pictures in the study phase were presented in 400×400 pixels size. In the two-alternative forced choice test, pictures were scaled down to 300×300 pixels.

Design. We used a $2 \times 2 \times 2$ factorial design with the factors List Length (short vs. long), Material (homogeneous vs. heterogeneous), and Order (short-long vs. long-short). The factor List Length was manipulated within subjects. Dependent variables were the proportion of correct answers in a two-alternative forced choice recognition test and response latencies. All factor levels were randomly assigned to participants.

Procedure. The procedure was the same as in Experiment 1, except that every participant only passed two blocks in total. Because recognition for heterogeneous picture material is usually very high, we reduced the presentation time to 700 ms in the heterogeneous condition. In the homogeneous condition, pictures were each presented for 900 ms.

Results

Within-subjects analysis. Means of correct answers for all conditions are shown in Table 1 (see also Figure 3, for a graphical representation). Memory performance for short lists ($M = .90$) was better than memory performance for long lists ($M = .87$), $F(1, 155) = 5.29$, $p < .05$, $\eta_p^2 = .04$. Moreover, recognition performance for items of heterogeneous lists was higher ($M = .91$) than for items of homogeneous lists ($M = .86$), $F(1, 115) = 15.81$, $p < .05$, $\eta_p^2 = .14$. There was no main effect of Order, $F(1, 115) = 1.17$, $p > .05$. As stated above, we expected to find a list length effect for homogeneous but not for heterogeneous material. However, neither the interaction of Length and Material was significant, $F(1, 115) = 1.50$, $p > .05$, nor the two other two-way interactions (greatest $F = 1.41$). Of particular interest in this analysis is the three-way interaction of Order, Length, and Material. We expected the Order x Length interaction to be evident for homogeneous lists (thereby replicating the results from Experiment 1) but not for heterogeneous lists. Somewhat unexpectedly, this three-way interaction was not significant ($F < 1$). Nevertheless, it again might be insightful to look at the first and second block separately.

Between-subjects analysis. In the first block, only extralist interference and intralist interference affect memory performance. We hypothesized that extralist interference is quite large for heterogeneous items and therefore intralist interference should be masked by the huge amount of extralist interference. Homogeneous items on the other hand should be at least partially immune to the influence of extralist interference. Thus, we expected a list length effect for homogeneous but not for heterogeneous items. A simple main effect analysis of the effect of Length nested in Material confirmed this predictions. First, there was a significant main effect of Material, $F(1, 115) = 5.40$, $p < .05$, $\eta_p^2 = .05$, with better performance for heterogeneous items ($M = .91$) than for homogeneous items ($M = .86$). More importantly, we found a significant list length effect for homogeneous material, $F(1,115) = 7.58$, $p < .01$, $\eta_p^2 = .06$, but not for heterogeneous material ($F < 1$).

In the second block, interlist interference comes into play. Again, for heterogeneous material the additional impact of interlist interference should be small, considering the supposedly large impact of extralist interference. For homogeneous items, interlist interference is more critical because homogeneity can protect only from the influence of preexperimental learning episodes if items are independent of the study material. For the short list length in the second block there should be a greater amount of interlist interference (because the long list was studied in the first block) and only a small amount of intralist interference. For the long list, there should be only a small amount of interlist interference and a large amount of intralist interference. Therefore, the list length effect should be absent for heterogeneous material and absent or at least reduced for homogeneous material. Again, this prediction was confirmed with a simple main effects analysis. The main effect of Material was significant, $F(1, 115) = 14.53$, $p < .01$, $\eta_p^2 = .11$, but both nested main effects of Length were not significant (both Fs < 1).

Response latencies. Means of median response times for Experiment 2 are shown in Table 2. There was a significant main effect of Material, $F(1, 115) = 25.23$, $p < .01$, $\eta_p^2 = .18$, caused by faster response times for the heterogeneous list ($M = 1425.57$ ms) compared with the homogeneous list ($M = 1713.13$ ms). Also, participants responded faster to items of the short list ($M = 1526.05$ ms) than to items of the long list ($M = 1605.41$ ms), $F(1, 115) = 7.18$, $p < .01$, $\eta_p^2 = .06$. Neither the main effect of Order nor any of the interactions were significant (greatest $F = 1.09$). This pattern of response latencies disqualifies the interpretation that performance differences are due to a speed–accuracy tradeoff. On the contrary, longer response latencies were associated with lower memory performance.

Discussion

The results of Experiment 2 further confirm our notion that homogeneity is the key to understand list length effects in recognition memory. Overall, there was only a small list length effect. The between-subjects analyses showed that this effect is mainly driven by the first block of the homogeneous condition. Analyses of the second block further demonstrated that homogeneity alone is not enough to explain a list length effect. In the homogeneous condition, there is no longer a list length effect in the second block. This result replicates the finding from Experiment 1 where we argued that homogeneity might minimize the influence of unrelated extralist items, but it cannot protect from interlist interference when items from the same stimulus material were studied before. In the heterogeneous condition, there was no list length effect in the first nor in the second block. The most important aspect of

Experiment 2 concerns the manipulation of homogeneity: Participants in both the heterogeneous and homogeneous condition studied the same set of pictures, only in the homogeneous condition image-processing filters were used to make the pictures appear more similar. In Experiment 3, we will demonstrate that this manipulation of homogeneity also works with a different type of stimulus material.

Experiment 3

To investigate to which extent the findings from Experiment 2 are applicable to other stimuli and to further test the impact of extralist items on homogeneous versus heterogeneous stimulus material, we conducted a third experiment. In their second experiment, Kinnell and Dennis (2012) found a list length effect using pictures of faces. Like the photographs of scenes used in our Experiment 2, faces are a very familiar material in the sense that we see faces (and pictures of faces) every day. But unlike the photographs of scenes, the pictures of faces used by Kinnell and Dennis (2012) were very homogeneous, as they all showed faces of middle-aged Caucasian adults (of both genders) on a neutral background. Just as in our previously reported findings, we argue that homogeneity of the faces is the crucial factor that allowed Kinnell and Dennis (2012) to find a list length effect with this stimulus material. If this argumentation is correct, it should be possible to *switch off* the list length effect using more heterogeneous face material and again *switch on* the effect by homogenizing the material with image-processing techniques as used in Experiment 2.

Method

Participants. In Experiment 3, 107 participants were tested, of which 82.2% were female. Mean age of participants was 21.69 years, ranging from 18 to 37 years. Almost all participants were students of psychology at the University of Mannheim and received course credit for their participation. Four subjects were excluded prior to the analysis because their performance did not significantly exceed chance level.

Material. Our stimulus set contained 271 photographs of faces, which represented both genders, different ethnicities, covered a wide age range (including children and elderly people), and expressed different emotional states (although most faces displayed a neutral or happy expression). There were different backgrounds in the pictures; however, we ensured that it was more or less neutral, that is, there were no people or dominating objects in the background. Some of the people in the pictures wore headgear, glasses, or make-up. In the heterogeneous condition, all photographs were shown without any alteration. However, in the homogeneous condition, we again applied the *pencil sketch* and the *high dynamic range* image-processing filters implemented in the software *Picasa* to all images. As a consequence, all pictures appeared to be more similar to each other in the homogeneous condition than they did in the heterogeneous condition, although the actual content of the pictures, that is, the faces, remained the same. Just like in Experiments 1 and 2, pictures in the learning phase were presented in 400×400 pixels size. In the two-alternative forced choice recognition test, pictures were scaled down to 300×300 pixels. For the training phase, we used faces of cartoon figures to minimize interference with the material used in the actual experiment.

Design. We used a $2 \times 2 \times 2$ factorial design with the factors List Length (short vs. long), Material (homogeneous vs. heterogeneous), and Order (short-long vs. long-short). The factor List Length was manipulated within subjects. Dependent variables were the performance of correct answers in a two-alternative forced choice recognition test and response latencies. All factor levels were randomly assigned to participants.

Procedure. The procedure was the same as in Experiment 2, and again, presentation time in the heterogeneous condition was 700 ms and 900 ms in the homogeneous condition with a 200-ms interstimulus interval.

Results

Within-subjects analysis. Means of correct answers for all conditions are shown in Table 1 (see also Figure 3, for a graphical representation). Memory performance for heterogeneous faces ($M = .92$) was better than for homogeneous faces ($M = .85$), $F(1, 101) = 22.37$, $p < .01$, $\eta_p^2 = .18$. There was no main effect of List Length ($F < 1$) but a significant interaction of Length and Order, $F(1, 101) = 13.42.37$, $p < .01$, $\eta_p^2 = .12$. Separate repeated measurements ANOVAs revealed that there was a significant list length effect ($M(short) = .92$, $M(long) = .87$) when the short list was studied first, $F(1, 51) = 13.22.37$, $p < .01$, $\eta_p^2 = .21$, and a no list length effect when the long list was studied first ($M(short) = .86$, $M(long) = .89$), $F(1, 52) = 3.36$, $p = .07$. We expected this pattern to be more pronounced for the homogeneous condition but the three-way interaction including the factor Material was not significant ($F < 1$). All other effects were also not significant (greatest $F = 1.63$).

Between-subjects analysis. Analyzing just the first block in a between-subjects simple main effect analysis, we found a significant effect of Material, $F(1, 101) = 16.65$, $p < .01$, $\eta_p^2 = .14$, showing that memory performance for heterogeneous material ($M = .94$) was better compared to homogeneous material ($M = .87$). More importantly, recognition accuracy for short lists ($M = .90$) was better than for long list ($M = .84$) in the homogeneous condition only, $F(1, 115) = 4.71$, $p < .05$, $\eta_p^2 = .05$. In the heterogeneous condition, this list length effect was not significant ($F < 1$).

In the second block, items from heterogeneous lists were remembered better ($M = .91$) than items from homogeneous lists ($M = .83$), $F(1, 101) = 14.75$, $p < .01$, $\eta_p^2 = .13$. Because in the second block interlist interference masks the effect of intralist interference, we expected the list length effect to be absent in the homogeneous condition. Indeed, the simple main effect of Length was not significant, neither nested in the homogeneous condition, $F(1, 101) = 1.44$, $p = .23$, nor in the heterogeneous condition ($F < 1$).

Response latencies. Means of individual median response times for Experiment 3 are shown in Table 2. Overall, response latencies were faster for items from heterogeneous lists ($M = 1341.26$ ms) than for items from homogeneous lists ($M = 1735.54$ ms), $F(1, 101) = 26.72$, $p < .01$, $\eta_p^2 = .21$. Further, there was a significant main effect of Length, $F(1, 101) = 10.36$, $p < .01$, $\eta_p^2 = .09$, showing that items from short lists were associated with faster responses ($M = 14.81.07$ ms) than items from long lists

(M = 1584.47 ms). Analogously to the accuracy data, there was a significant interaction of Order and Length, $F(1, 101) = 28.33$, $p < .01$, $\eta_p^2 = .22$. Separate ANOVAs for the two Order conditions revealed that there is a huge effect of Length on response times ($M(short)$ = 1374.46 ms, $M(long)$ = 1654.49 ms) when the short list is studied first, $F(1, 51) = 36.14$, $p < .01$, $\eta_p^2 = .42$, but no effect when the long list is studied first, $F(1, 52) = 2.18$, $p = .15$. All other main effects or interactions were not significant (greatest $F = 3.04$). Overall, this pattern is in line with the accuracy data. Manipulations which decrease memory performance (by increasing homogeneity or list length) concurrently increase response latencies. A speed–accuracy trade-off can definitively be ruled out by this pattern in the response latencies.

Discussion

In sum, the results of Experiment 3 confirm the results of Experiment 2. This time, we used stimulus material for which Kinnell and Dennis (2012) found a list length effect. If there was something special about faces, then it should not make a difference whether we used unaltered or manipulated pictures of faces. We argue that there is nothing special about faces per se, but that the homogeneity of the material is critical for the list length effect to occur. Following this line of reasoning, the list length effect should be eliminated when using more heterogeneous pictures of faces. This is exactly what we found in the between-subjects analysis. Furthermore, following the same logic as in Experiment 2, we took the same set of heterogeneous faces and homogenized it by applying a set of image-processing filters. For these homogenized pictures of faces there is a strong list length effect in the first block. In this block, only intralist interference (which varies as a function of list length) and interference from extralist items (which is constant for different list lengths) affects participants' performance. Because, according to MINERVA 2, homogeneity reduces the influence of extralist items, the differences of intralist interference, that is, the effect of differing list lengths, should be observable. This situation changes dramatically in the second block, where interlist interference a) varies for different list lengths and b) works against the effect of intralist interference. As a consequence, the list length effect disappears even in the homogeneous condition in Block 2. To summarize, our results strongly support our assumption that homogeneity of the stimulus material is the crucial variable in the occurrence of a list length effect and not the type of material per se.

General Discussion

The list length effect is a controversial phenomenon. Empirically, several studies replicated the effect (e.g., Brandt, 2007; Cary & Reder, 2003; Gillund & Shiffrin, 1984; Gronlund & Elam, 1994; Yonelinas, 1994), whereas others failed to do so (e.g., Buratto & Lamberts, 2008; Dennis & Humphreys, 2001; Kinnell & Dennis, 2012). Theoretically, item noise models predict the effect attributable to intralist interference, whereas context noise models negate any intralist interference and therefore predict a null list length effect. Recently, Kinnell and Dennis (2012) showed that the effect only holds for certain item material. Further, they speculated that item interference only emerges when interstimulus similarity (i.e., the homogeneity of the stimulus material) is high enough. Based on their research, we offered a theoretical explanation for their results by means of the item noise model MINERVA 2.

According to our approach, preexperimental memory traces increase the variances of echo-intensity distributions of both old and new test items. These increases in the variances of the underlying distributions mask the relatively small increase in variances that is attributable to a longer study list with heterogeneous material. However, studying homogeneous material changes the situation: Increasing the homogeneity of the stimulus material will enlarge the variances of echo-intensity distributions. As a consequence, the relative impact of the additional variances due to extralist items is reduced, that is, homogeneous material immunizes against the impact of extralist items such that intralist interference can be observed. Note, however, that only the impact of unrelated extralist items is reduced for homogeneous stimulus material. In a multilist within-subjects design, items from previously learned lists of the same stimulus category will reduce the list length effect for items in late blocks irrespective of the homogeneity. In within-subjects analyses, the list length effect therefore is predicted to be reduced not only for heterogeneous but also for homogeneous material. Analyzing the first and the second block of a within-subjects design separately (i.e., between subjects), our approach suggests a stronger list length effect for the first block and a reduced list length effect for the second block, compared to within-subjects analyses.

We tested our approach in three experiments. In Experiment 1, we used photographs of flowers and landscapes to address the homogeneity hypothesis. With both kinds of material we found the predicted interaction of list length and order of length studied. In between-subjects analyses, the list length effect was significant only in the first block but not in the second block. In Experiments 2 and 3, we tested the homogeneity hypothesis more thoroughly. In Experiment 2, we used a set of photographs of heterogeneous scenes comparable with those used by Kinnell and Dennis (2012, Exp. 4). By applying image-processing filters to these photos, we experimentally homogenized the material. As predicted, we replicated the null list length effect in the heterogeneous condition but found a list length effect for homogeneous material. Experiment 3 was conducted in reference to the results of Experiment 2 by Kinnell and Dennis (2012). Using pictures of faces, they found a significant list length effect. We hypothesized that this effect resulted from the homogeneous nature of the faces used in their experiment and should be eliminated by using more heterogeneous faces. Indeed, we found no list length effect for heterogeneous faces in Experiment 3. However, when we applied the same image-processing filters to the pictures as we did in Experiment 2, we indeed found a list length effect for the same set of photos in the between-subjects analysis of the first block. In all experiments, response time analyses confirmed that the list length effects we found were not simply attributable to a speed–accuracy trade-off.

To see that the pattern of the results of Experiments 1–3 quantitatively match model predictions of MINERVA 2, we ran a simulation of 500 experiments with 80 participants each, implementing the design of Experiments 2 and 3. The empirical results as well as the results of the simulation are shown in Figure 3. Although the performance level is lower in the simulation than in the experiments, the overall pattern is quite clear. There is only a substantial list length effect for homogeneous material in the first block of the study. For heterogeneous material, the effect is very small in the first block. In the second block, there is a null effect for both levels of homogeneity. It is puzzling that we did not find a significant three-way interaction between the factors Length, Homogeneity, and the Order of testing. We argue that this effect is attributable to a lack of statistical power. Given the perfect simulated data shown in Figure 3 we found a significant three-way interaction only in about 10% of the cases. Thus, it is very likely not to find any significant interaction given this sample size. This is a general problem of finding predicted interference effects of item noise models, such as the list length effect or the list strength effect. These effects are statistically small effects and, hence, difficult to detect in empirical data.

Moreover, in our experimental data, the influence of the critical variable homogeneity is rather weak, especially in Experiment 2 and 3. In these experiments, we started with heterogeneous material in order to show that extralist interference eliminates the list length effect. This was confirmed for heterogeneous scenes (Exp. 2) and heterogeneous faces (Exp. 3). We tried to find a manipulation that renders the scenes and faces into more homogeneous material using some picture manipulation. However, such a manipulation can influence the homogeneity only slightly because the semantics of the pictures are not affected. Therefore, it is not surprising that the effect sizes for homogeneity are small and the statistical results are not completely convincing. Our manipulation of the homogeneity—as it was done in Experiments 2 and 3—might be weak, but it points to the potential underlying mechanism instead of focusing on effects for certain stimulus categories. The effect of the order of testing, on the other side, is strong and stable in our data. Note that the order of testing refers to the influence of interlist interference only. Assuming that the learning context does not change greatly from list 1 to list 2 and MINERVA 2 holds, after the second study list, participants all have the same number of items in memory and no list length effect is to be expected in the second block. This argument holds irrespective of the homogeneity of the material. The effect of homogeneity in our account is to constrain the influence of unrelated extralist items, not of items of the same category learned in a previous list. As a consequence, a list length effect in within-subjects designs is hard to find and usually smaller than model predictions, as long as the context of the lists is not manipulated to be very different. If this was the case, homogeneity of the contexts would homogenize every single list and the influence of list 1 items on list 2 items should be partially blocked.

It is also important to note that we adhered to the procedure of Kinnell and Dennis (2012) as close as possible. We additionally tried to eliminate all possible confounds discussed in Kinnell and Dennis (2011). The only structural difference in the experiments was that we implemented a two-alternative forced choice recognition test, whereas Kinnell and Dennis (2012) used a yes/no recognition test. We used the forced choice test because estimators of the signal-detection-based performance measure *d*' are biased when the underlying distributions for old and new test items have unequal variances (e.g., Verde, Macmillan, & Rotello, 2006). Because MINERVA 2 predicts unequal variances, this might be an issue.

We did not investigate the other types of material used by Kinnell and Dennis (2012). In Experiment 3, they found a list length effect for fractals. Following our line of reasoning, this effect should disappear using more heterogeneous fractals. Unfortunately, it is not easy to create heterogeneous fractals. Also, the impact of extralist items should be stronger if items of the same material have frequently been encountered before the experiment. Assuming that participants do not admire the beauty of fractals too often in their spare time, it should be hard to find a substantial impact of extralist items at all. Nevertheless, even with this rather unusual stimulus material, we would predict an interaction of list length and order of study lists in a within-subjects design. This is because even unusual homogeneous material is prone to the impact of interlist interference. Unsurprisingly, Kinnell and Dennis found a significant interaction of these factors in their study. Unfortunately, they neither discuss this effect nor do they present the descriptive data for the first and second block. Based on our theoretical analyses and our empirical results, we predict that the list length effect was stronger in the first block of their experiment than in the second block.

In their first experiment, Kinnell and Dennis (2012) found no list length effect when they used word pairs as stimuli. The reason why we did not examine the influence of homogenizing word material on the list length effect will be discussed in the following section. We will refer to an important study that seems to contradict our approach at a first glance.

The Inverse List Length Effect

Dennis and Chapman (2010) showed that under certain circumstances, recognition accuracy can increase with list length. This so-called *inverse list length effect* does not seem to be compatible with the item noise approach proposed in this paper. This approach can predict a null list length effect if noise from extralist items masks the interference caused by intralist items but an inverse effect seems impossible. However, a closer look at their experimental procedures reveals that this is not the case. In their experiments, Dennis and Chapman used word material of eight taxonomic categories. For the shortest list, participants studied one word, for the medium list they studied three words, and for the long list they studied eight words from each category. For targets, they observed a weak increase in the hit rate. If distractors were unrelated, that is, words were not drawn from the taxonomic categories, the false alarm rate decreased with list length. In other words, comparing targets with unrelated distractors results in an inverse list length effect. It is important to note that this very situation does not reflect a typical list length effect study, because distractors and targets were not drawn from the same stimulus pool. In fact, under the assumption that words from the same category are semantically related, item noise models actually predict exactly this pattern. MINERVA 2, for example, predicts that the variances of both the target and the distractor distributions indeed increase with list length. But, as opposed to a typical list length experiment, the difference of the means between the two echo-intensity distributions is not unaffected in this situation. For unrelated distractors, the mean of the echo-intensity distribution is zero, irrespective of list length. The mean of the target distribution, however, will increase with list length, especially for homogeneous lists. When the length of a homogeneous study list is increased, the number of similar items in memory rises accordingly: There is one matching and seven unrelated memory representations in the short list; one matching memory representation, two similar, and 21 unrelated memory representations in the medium list, and one matching, nine similar, and 70 unrelated memory representations in the long list. Because the local activation of a single memory representation that is similar to the test item is greater than zero (cf., formula 7), increasing mean differences might outperform the increasing variances in this list length experiment. Therefore, the better performance in the longer lists is perfectly in line with the model predictions of MINERVA 2. The situation is different if targets and distractors from the same taxonomic category are compared. In this case, the mean differences of the target and distractor distributions remain constant with increasing list length. Therefore, a typical list length effect should emerge. Fortunately, Dennis and Chapman also tested distractors from the taxonomic categories. Here the false alarm rates increase with list length, and, importantly, this increase is stronger than the increase in hit rates. In other words, a typical list length effect can be observed as it is predicted by the item noise approach.

Dennis and Chapman (2010) argue that there is no representational overlap (i.e., no similarity) in word material because words are usually well learned and therefore highly discriminable. Their empirical argument for this assumption is that list length effects are seldom found with word material. Therefore, the comparison of targets with unrelated distractors seems to be justified. On the other hand, from their point of view it is not easy to explain why false alarms increase with list length for related distractors. Our explanation is different: We agree that it is difficult to observe list length effects with word material. But, according to our approach, this occurs because words are often encountered and therefore frequently represented in memory. Hence, word material leads to strong extralist interference that masks the intralist interference predicted by item noise models. To overcome these problems, one may (a) increase power and (b) use homogenized material. Interestingly, this is exactly what Dennis and Chapman did. First, they increased power by comparing a short and a long list which strongly differed in length by a factor of 10, which is quite unusual. Second, they increased the homogeneity of stimulus material by using semantic categories. As shown in our experiments, in MINERVA 2 this homogeneity of the stimulus material immunizes against the disruptive influence of extralist items.

One critical aspect of the study by Dennis and Chapman (2010) is the fact that the false alarm rate decreases with list length for unrelated distractors. Because—according to MINERVA 2—the mean of the echo-intensity distribution is constant for unrelated distractors and the variance increases with list length, this pattern still seems incompatible with our assumptions. However, because the mean of the target distributions rises, it is plausible that participants adopt a different response criterion in this situation (Hirshman, 1995). Allowing for varying response criteria, we computed the predictions of MINERVA 2 for the experimental procedure of Dennis and Chapman using the following parameters: We assumed 30 features for every item. Homogeneity was set to $h = .75$ within every semantic category. Furthermore, we assumed that there were 20,000 extralist items moderately learned with $L_{extra} = .05$, all unrelated to the categories. The experimental items were learned with $L = .35$. The model predictions qualitatively matched the pattern found by Dennis and Chapman (see Figure 5): There was a moderate increase in the hit rate, a pronounced increase in the false alarm rate for related distractors and a moderate decrease in the false alarm rate for unrelated distractors. Therefore, assuming different response criteria, MINERVA 2 can both model the inverse as well as the typical list length effect in the paradigm of Dennis and Chapman (2010).

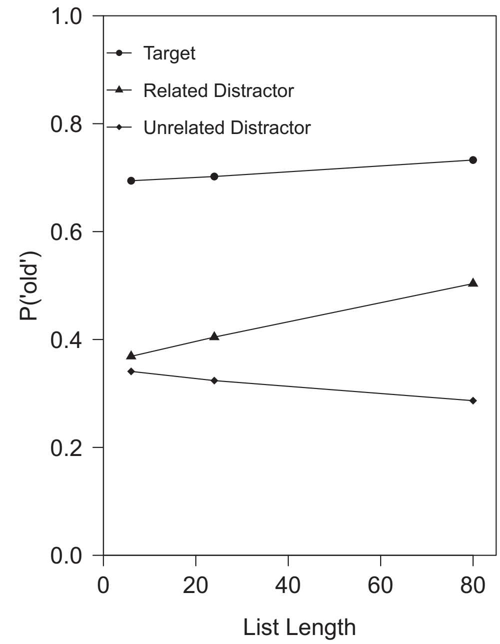

Figure 5. Model predictions of MINERVA 2 in the experimental paradigm of Dennis and Chapman (2010). Note that recognition performance increases with list length when targets are compared with related distractors, but decreases with list length when targets are compared with unrelated distractors. MINERVA 2 parameters: $M = 30$, $L = .35$, $h = .75$, 20.000 unrelated extra list items learned with $L_{extra} = .05$. Response criteria were set to 0.02, 0.024, and 0.036 for the short, medium, and long list, respectively.

Although we were quite successful in modeling both the list length effect and the inverse list length effect of Experiment 1 of Dennis and Chapman (2010), MINERVA 2 cannot explain the results of Experiment 2 in their paper. In this experiment, Dennis and Chapman used the same stimulus material and design as in their Experiment 1. But this time they randomized the positions of all the items within and between categories in the study phase. Thus, only the order of memory representations is different, compared to Experiment 1. As a result of this manipulation, the list length effect remained but the inverse list length effect disappeared. Because the order of memory representations is not relevant in MINERVA 2, it is clear that this results cannot be modeled by MINERVA 2 without any further assumptions, for example, concerning the learning strategy. Alternatively, one could eventually adjust the response criteria to fit the data. But we see no theoretical reason for response criteria to vary with the order of the study list. To summarize, MINERVA 2 can in principle model a list length effect and an inverse list length effect, but in its current form, it cannot account for the whole pattern of results found by Dennis and Chapman (2010).

Conclusions

According to item noise models such as MINERVA 2, adding items to a study list increases the variance of the familiarity distribution of both targets and distractors. As a consequence, intralist interference emerges and recognition performance should drop with increasing list lengths. But these models do not only predict intralist interference. They also predict interlist interference: Items from a previous study list will influence memory performance for items tested later. Interlist interference and additional interference from preexperimental items can mask intralist interference. With heterogeneous and familiar material, the list length effect is predicted to be masked by the influence of extralist items and no list length effect is expected in both blocks of a within-subjects design. This is exactly the pattern we found in our experiments (see Figure 3). In the between-subjects analyses of our experiments, however, we found a pronounced list length effect with homogeneous material in the first block. In the second block, there is no list length effect even for homogeneous material.

In this paper, we focused on the model MINERVA 2 but we think that this prediction also holds for other global memory models. In fact, the idea that interference effects predicted by these models might be masked by extralist items has already been discussed as *continuous memory* by, for example, Murdock and Kahana (1993). Moreover, even the idea that homogenization immunizes against the influence of extralist item is not new. Hintzman (1988) argues that identical context features in the study phase minimize the influence of previously learned items. According to him, homogenization is due to a common study context, whereas in our approach, identical item features additionally increase homogenization. This seems to be especially important if the context and the study material are familiar. Because reading words on a computer screen is a very familiar task, this homogenization by means of identical item features seems to be highly relevant.

In context noise models such as the BCDMEM (Dennis & Humphreys, 2001), the different contexts in which an item was studied are the only source of errors because items are represented orthogonally in memory. As a consequence, adding different items to a study list should not add noise to a single item and no list length effect should occur. Allowing only for context noise, the model neither can explain the results of our experiments, nor the results reported by Kinnell and Dennis (2012). Moreover, it is hard to explain why our homogeneity manipulation (i.e., applying image-processing filters) has such drastic effects on item presentations. Moreover, pure context noise models are inherently weak in explaining the typical rise in false alarm rates for similar distractors, which is routinely found for word material (e.g., Anisfeld & Knapp, 1968; Gillund & Shiffrin, 1984) as well as abstract material (e.g., Franks & Bransford, 1971; Posner & Keele, 1970). Although for verbal material an implicit associative response process (Underwood, 1965) might be possible, this process seems to be hardly plausible for abstract visual material. At last, there is no logical reason for the strict representational distinction of context and stimulus information. If the learning episode is a bit more complex than the presentation of a single word, participants will not always know whether a single feature is part of the item itself or the context. For example, studying the revelation effect, Cameron and Hockley (2000, Exp. 5) presented their participants with pairs of words in the study phase. However, in the following recognition test they tested only a single word. Thus, one of the words constitutes the item information, the other one is part of the context information. How could item information in this case be represented fully orthogonally, but context information with representational overlap?

Although context noise models cannot explain intralist-interference effects, they offer a promising account for word-frequency effects. Because high-frequency words have been encountered a lot in different contexts, the retrieved context of a test word is noisier compared to low-frequency words. As a result, recognition performance is better for low-frequency words (Schulman, 1967). Moreover, the typical mirror effect (Glanzer & Adams, 1985) found in recognition memory is a consequence of the Bayesian decision rule of the model (but see Hemmer & Criss, 2013, for a more complex pattern). Both the word-frequency effect and the mirror effect cannot be explained by global matching without additional assumptions. One exception might be the REM model (Shiffrin & Steyvers, 1997) that also has a Bayesian decision rule, but a slightly different approach for the frequency effects. Nevertheless, when context information is taken into account in a global matching approach, items tested in the same context should elicit a stronger memory signal than test items from different contexts. This prediction has been confirmed empirically (e.g., Arndt, 2010; Murnane & Phelps, 1993).

Dual-process models of recognition memory (Jacoby, 1991; Mandler, 1980) explain an observed list length effect by a disrupted recollection process (Yonelinas, 1994), whereas the familiarity process is unaffected by additional items in the study list. In principle, this approach is in line with the results of our experiments. In fact, there is some evidence that participants rely more on recollection if the similarity of test items increases (e.g., Norman, 2002). Thus, for more homogeneous material a stronger list length effect might be expected. However, it is at least astonishing from that point of view that the relative contribution of recollection is higher for abstract visual material such as fractals (Kinnell & Dennis, 2012) or checkerboard patterns (Brandt, 2007) compared with word material. Concerning the between-subjects analysis, we see no obvious reason why the recollection process is selectively impaired in the first blocks only.

The goal of this paper was to shed light on some seemingly contradicting results concerning the list length effect. Kinnell and Dennis (2012) argued that homogeneity might be a necessary condition for creating intralist interference in recognition memory. Based on this assumption, we presented a profound theoretical analysis of this issue based on the MINERVA 2 model. In short, according to our approach, intralist interference is always present but its effect is masked by the influence of preexperimental memory representations when familiar heterogeneous stimulus material is used. Although we focused on MINERVA 2, we are convinced that this argument also holds for other global matching models. Moreover, we do not think a definite decision has to be made between item noise or context noise models. In a more recent paper, Osth and Dennis (2015) reanalyzed several experiments with a model incorporating item noise, context noise and background noise (i.e., the effect of extralist items) and estimated the relative amount of these sources of interference in several paradigms. Whereas global matching models used to focus on item noise mainly and the BCDMEM model focused on context noise, it is definitively promising to consider all sources of interference within a single model. Analyzing the experiments of Kinnell and Dennis (2012), Osth and Dennis concluded that item noise is relatively strong for fractals and scenes. But like Kinnell and

Dennis (2012), they did not offer a sound theoretical explanation for this finding. With the theoretical and empirical work in this paper, on the other hand, we do offer an explanation. Within our approach, intralist interference does not depend on certain stimulus material, but highly frequent heterogeneous item material opens the door for background noise, which minimizes the relative amount of intralist interference.

The current paper provides some evidence for item-specific interference, which it is at the heart of item noise models. Nevertheless, the role of context-specific interference as well as background noise might also play an important role in these models. For example, having learned an item many times in very different contexts can completely disturb the process of contextual reinstatement in item noise models and therefore constitute the basis of semantic, that is, context-free, memories. Overall, we share the belief that item noise as well as context noise play an important role in recognition memory processes (Criss, Malmberg, & Shiffrin, 2011; Criss & Shiffrin, 2004; Osth & Dennis, 2015).

References

Anisfeld, M., & Knapp, M. (1968). Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology, 77,* 171–179. http://dx.doi.org/10.1037/h0025782

Annis, J., Lenes, J. G., Westfall, H. A., Criss, A. H., & Malmberg, K. J. (2015). The list-length effect does not discriminate between models of recognition memory. *Journal of Memory and Language, 85,* 27–41. http://dx.doi.org/10.1016/j.jml.2015.06.001

Arndt, J. (2010). The role of memory activation in creating false memories of encoding context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 66–79. http://dx.doi.org/10.1037/a0017394

Brandt, M. (2007). Bridging the gap between measurement models and theories of human memory. *Zeitschrift für Psychologie/Journal of Psychology, 215,* 72–85. http://dx.doi.org/10.1027/0044-3409.215.1.72

Buratto, L. G., & Lamberts, K. (2008). List strength effect without list length effect in recognition memory. *The Quarterly Journal of Experimental Psychology, 61,* 218–226. http://dx.doi.org/10.1080/17470210701566713

Cameron, T. E., & Hockley, W. E. (2000). The revelation effect for item and associative recognition: Familiarity versus recollection. *Memory & Cognition, 28,* 176–183. http://dx.doi.org/10.3758/BF03213797

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. http://dx.doi.org/10.1016/S0749-596X(03)00061-5

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. http://dx.doi.org/10.3758/BF03210740

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language, 64,* 316–326. http://dx.doi.org/10.1016/j.jml.2011.02.003

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

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. http://dx.doi.org/10.1016/j.jml.2010.06.001

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* 452–478. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.jml.2008.06.007

Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. *Journal of Experimental Psychology, 90,* 65–74. http://dx.doi.org/10.1037/h0031349

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67. http://dx.doi.org/10.1037/0033-295X.91.1.1

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

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66,* 325–331. http://dx.doi.org/10.1111/j.2044-8295.1975.tb01468.x

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. http://dx.doi.org/10.1037/0278-7393.20.6.1355

Hemmer, P., & Criss, A. H. (2013). The shape of things to come: Evaluating word frequency as a continuous variable in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1947–1952. http://dx.doi.org/10.1037/a0033744

Hintzman, D. L. (1986). 'Schema abstraction' in a multiple-trace memory model. *Psychological Review, 93,* 411–428. http://dx.doi.org/10.1037/0033-295X.93.4.411

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551. http://dx.doi.org/10.1037/0033-295X.95.4.528

Hintzman, D. L., & Waters, R. M. (1970). Recency and frequency as factors in list discrimination. *Journal of Verbal Learning and Verbal Behavior, 9,* 218–221. http://dx.doi.org/10.1016/S0022-5371(70)80054-8

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. http://dx.doi.org/10.1037/0278-7393.21.2.302

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

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541. http://dx.doi.org/10.1016/0749-596X(91)90025-F

Kinnell, A., & Dennis, S. (2011). The list length effect in recognition memory: An analysis of potential confounds. *Memory & Cognition, 39,* 348–363. http://dx.doi.org/10.3758/s13421-010-0007-6

Kinnell, A., & Dennis, S. (2012). The role of stimulus type in list length effects in recognition memory. *Memory & Cognition, 40,* 311–325. http://dx.doi.org/10.3758/s13421-011-0164-2

Macmillan, N. A., & Creelman, C. D. (2008). *Detection theory: A user's guide*. New York, NY: Psychology Press.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271. http://dx.doi.org/10.1037/0033-295X.87.3.252

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review, 100,* 183–203. http://dx.doi.org/10.1037/0033-295X.100.2.183

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104,* 839–862. http://dx.doi.org/10.1037/0033-295X.104.4.839

Murdock, B. B., & Kahana, M. J. (1993). Analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 689–697. http://dx.doi.org/10.1037/0278-7393.19.3.689

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 882–894. http://dx.doi.org/10.1037/0278-7393.19.4.882

Norman, K. A. (2002). Differential effects of list strength on recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1083–1094. http://dx.doi.org/10.1037/0278-7393.28.6.1083

Osth, A. F., & Dennis, S. (2015). Sources of interference in item and associative recognition memory. *Psychological Review, 122,* 260–311. http://dx.doi.org/10.1037/a0038692

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308. http://dx.doi.org/10.1037/h0028558

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814. http://dx.doi.org/10.1037/0278-7393.21.4.803

Schulman, A. I. (1967). Word length and rarity in recognition memory. *Psychonomic Science, 9,* 211–212. http://dx.doi.org/10.3758/BF03330834

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

Strong, E. K. J. (1912). The effect of length of series upon recognition memory. *Psychological Review, 19,* 447–462. http://dx.doi.org/10.1037/h0069812

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70,* 122–129. http://dx.doi.org/10.1037/h0022014

Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of d', az, and a'. *Perception & Psychophysics,* (*4*), 643–654. http://dx.doi.org/10.3758/BF03208765

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354. http://dx.doi.org/10.1037/0278-7393.20.6.1341

Received February 20, 2017
Revision received February 19, 2018
Accepted February 22, 2018 ■