

Object Recognition's Garden Path: Low Spatial Frequencies

John A. Rohrlich¹, Tsung-Ren Huang², Thomas E. Hazy³, and Randall C. O'Reilly⁴

¹Department of Psychology and Neuroscience, University of Colorado Boulder

²Department of Psychology, National Taiwan University, Taiwan

³eCortex, Boulder, CO

⁴Department of Psychology and Center for Neuroscience, University of California Davis

john.rohrlich@colorado.edu

March 23, 2020

Abstract:

Several experiments, notably one done by Bruner and Potter (1964), have demonstrated delayed object recognition when viewing a blurred image gradually come into focus. Bruner and Potter (1964) suggested that the wrong answer is held until there is an obvious contradiction. Others have hypothesized that “competitive activation” is responsible for delayed recognition. The results of the experiments reported in this paper are consistent with a third hypothesis, that delayed recognition is due to an initial organization of image elements that is incompatible with correct recognition and that the initial grouping and figure-ground perception, among other aspects of organization, drive subsequent perception via top-down cortical pathways. A total of 7 experiments using 3 forms of degradation supported this hypothesis. Images degraded by low-pass filtering produced significant delay in recognition, while degradation by fragmentation did not, and a third form of degradation similar to fragmentation mitigated the effect. The experiments also demonstrate that if images are low-pass filtered delayed recognition occurs with presentations of as little as 100 ms and early presentations lead to delayed recognition over long intertrial intervals, at least 105 seconds. Further, support for the hypothesis that top-down cortical influence is key to this phenomenon came from an experiment showing that masking eliminates delayed recognition for short presentations. Taken together these results support a hypothesis that delayed recognition is due to errors in perceptual organization that lead to incorrect responses and these errors are fostered by low spatial frequencies.

All authors were at the Department of Psychology and Neuroscience, University of Colorado Boulder when this research was done.

Supported by: ONR grants N00014-14-1-0670 / N00014-16-1-2128, ONR D00014-12-C-0638. Thanks to Dean Wyatt and Seth Herd for many helpful comments and suggestions. R. C. O'Reilly is Chief Scientist at eCortex, Inc., which may derive indirect benefit from the work presented here.

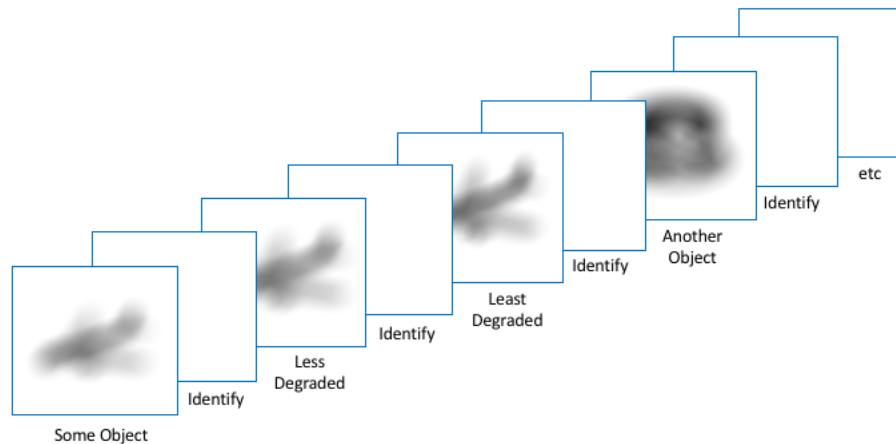


Figure 1: All experiments reported in this paper follow the same general pattern. Participants view a series of progressively less degraded versions of an object and make a guess about the identity of the object after every stimulus. The specific manner of degradation, timing of presentation and manner of responding is varied and detailed for each experiment.

Introduction

Perception, once thought to be a purely bottom-up process, is now largely viewed as an interactive process of bottom-up real-world input and top-down modulation (e.g., Eger, Henson, Driver, & Dolan, 2007; Fenske, Aminoff, Gronau, & Bar, 2006) and the belief that the top-down cortical connections are supporting continuous prediction via a generative model is gaining support (Clark, 2013; Bar, 2009). In perception the prediction is based on context, expectancy, and hypothesis testing (Esterman & Yantis, 2010; Zhang & Heydt, 2010; Gilbert & Sigman, 2007), and of course the current state, as what you see next is often well predicted by what you are seeing now. The prediction helps us to rapidly process sensory input by allowing us to predict *where* to look and *what* we will see.

The top-down influences of context and expectancy and the predictive process work so well that most of the time we do not question or test our interpretations of what we are seeing. However, when conditions produce a degraded, ambiguous visual input, we can make mistakes, and these errors have a lasting effect that interferes with later recognition. This was demonstrated in a widely cited experiment by Bruner and Potter (1964) in which participants viewed a blurred image gradually come into focus. They found that recognition was delayed when participants first viewed the pictures out of focus. Delayed recognition was also demonstrated by Galloway (1946) and Wyatt and Campbell (1951) with a somewhat different procedure. The phenomenon of delayed recognition is an interesting laboratory phenomenon, but it is also likely a real-world phenomenon when vision is hampered by poor weather, low lighting, distance or impaired eyesight. Delayed recognition is also interesting because it can help us understand visual processing under ideal viewing conditions, to the extent the process of object recognition is the same under good and poor viewing conditions. We explore this negative side of the top-down role in perception in seven experiments that feature sequences of progressively less degraded versions of common objects. All of the experiments have a similar form: Participants see one or more degraded versions of each object, in series from more to less degraded, and attempt to identify the object after each presentation (Figure 1).

In the Bruner and Potter (1964) experiment participants saw colored slides projected onto a screen. The images were blurred and gradually brought into more focus. The researchers varied both the initial degree of blur and the viewing time. In all conditions the projection was stopped before full focus and the participant was asked to report what they saw. They found that both viewing time and the initial degree of blur affected a participant's ability to identify what was shown in the image, such that the more blurred the initial image the less likely the participant was to identify the image at the final point of focus. Their conclusion was that participants form a hypothesis despite the degradation and because no obvious contradiction appears for a while the initial interpretation is maintained.

It is our hypothesis that the cause of delayed recognition is errors in perceptual organization which influence

subsequent perception via top-down visual processing mechanisms. The errors in organization make it impossible to correctly identify the object in the image. Correct recognition requires accurate perception of figure-ground, grouping, continuity, viewpoint and scale. Errors in any of these do not just lead to a wrong answer they lead you down the garden path to an entire category of answers that must be wrong.

The influence of prior viewing on basic perceptual organization has been demonstrated by Kimchi and Hadad (2002) as well as Razpurker-Apfeld and Kimchi (2007) who demonstrated priming of element grouping. Priming of figure-ground has been demonstrated by Peterson and Gibson (1994) and by Biederman and Cooper (1991), among others, using multiple types of stimuli and with multiple experimental paradigms. O'Herron and von der Heydt (2009) found that border ownership signals can persist for a second or more if the following stimulus is ambiguous. Thus, it follows from these studies that an error in organization, for example a grouping error, perceiving one object when there are two, will lead to a set of guesses which must be incorrect because the premise is incorrect.

The following examples of errors in organization, from a study by Davison (1964) concretely illustrate this process.

Example - grouping error:

Image: A small black puppy on grass, nibbling on a bone, with a dark shadow beneath its body

Participants erred by incorporating the dark shadow beneath the puppy into the animal's body making it appear more heavy-set and leading to guesses such as bull, hog and grand piano.

Example - viewpoint error:

Image: An aerial view of a cloverleaf interchange

Participants often assumed a frontal view rather than a top-down view and made the mistake of assuming the object was close to the observer. Together these errors led to seeing the surrounding town as various objects in the background.

Example - figure-ground error:

Image: Several glass ashtrays piled on a mahogany table

Participants made the assumption that the white objects are opaque, the result being that, as the color of the background emerges through the objects it is incorporated into them. Guesses include record album, magazines, etc.

Example - scale error:

Image: A set of fireplace implements on an Oriental rug

The picture was assumed to have been taken at either a very close distance (thus rendering it as small as pencils) or at a very great distance (thus rendering it as large as a building).

Since the Bruner and Potter (1964) publication, three additional studies of delayed recognition of objects have been published (Snodgrass & Hirshman, 1991, Luo & Snodgrass, 1994 and Schulkind, 2000). Snodgrass and Hirshman (1991) explored delayed recognition in five experiments using degradation by fragmentation rather than degradation by blurring. They hypothesized that if the participant simply held onto the guess until a contradiction appeared, as suggested by Bruner and Potter (1964), that right/wrong feedback after each presentation would eliminate delayed recognition. They found no significant diminishment of delay in the feedback group when compared with the no-feedback group. This result is consistent with our hypothesis that delayed recognition is not simply due to a wrong guess but that it is due to an error in the fundamental perceptual organization.

Two other experiments of Snodgrass and Hirshman's (1991) tested the hypothesis that "persisting activation" causes the delay, specifically that activation of "perceptual structures" persist increasing the activation of distractor items. They support this hypothesis with an experiment that shows less delayed recognition when stimuli are alternated with a simple arithmetic problem (sum two 1-digit numbers). Experiment 3, reported in this paper, found delayed recognition with a filled intertrial interval as long as 105 seconds which argues against any simple persisting activation hypothesis and is consistent with a hypothesis of influence from higher level cortical areas reconstituting basic grouping and figure-ground organization in lower level visual areas. We agree with Snodgrass and Hirshman that the influence of "perceptual structures" are at the root of this phenomenon but we suspect it is the reinstating of the basic perceptual organization through top-down cortical pathways rather than a simple short-lived activation that is key.

This paper reports on seven experiments using three types of degradation: low-pass filtering (i.e., blurring), fragmentation and a third type we call blocking to support the hypothesis that it is the initial perceptual organization that is responsible for delayed recognition. We support this by showing that different forms of degradation lead to significant differences in the degree of delayed recognition, specifically showing that a series of images that are degraded by low pass filtering leads to significantly more delay in recognition than a series degraded by fragmentation or blocking. Our low-pass filtering condition is similar to that used by Bruner and Potter (1964), Wyatt and Campbell (1951) and Galloway (1946) while the fragmentation condition is the same degradation used by Snodgrass and Hirshman (1991), Luo and Snodgrass (1994) and Schulkind (2000).

Overall, the experiments that have used degradation by fragmentation appear to show a weaker effect and one that is easily overcome. Luo and Snodgrass (1994) saw the delay disappear by interposing a very simple task and Schulkind (2000) saw the effect disappear after a mere 300-500 ms exposure to a blank screen. Because fragmentation interferes with continuity and grouping, the stimuli used in the experiments of Snodgrass and Hirshman (1991), Luo and Snodgrass (1994) and Schulkind (2000) have poor global form. The strength and stability of an organization is determined by the cohesiveness of the parts forming the whole. When the elements follow the Gestalt principles of similarity, proximity, good continuation and closure the global structure is easily seen. Filtering out the high frequencies but keeping the low (i.e., blurring) maintains continuity and grouping. Good form requires low spatial frequencies (Lagasse, 1993) and these are absent in degradation by fragmentation. Images with less good form, those lacking continuity and low spatial frequencies, will elicit weaker organization which is slower to develop.

In summary, we propose that delayed recognition is due to an initial organization of elements that is incompatible with correct recognition and that the grouping and figure-ground perception from the first view influences the next view via the normal top-down cortical processes. We provide evidence for the perceptual error hypothesis by demonstrating that degradation leading to strong organization leads to significantly more delay than degradation of two other forms that induce weak organization. The hypothesis is further supported by experiments showing that degradation by low-pass filtering occurs with presentations of only 100 ms and that the effect can persist through a significant interval if the original image has the low spatial frequencies needed for good form.

Experiment 1

How general is delayed recognition? The stimuli in the experiments of Galloway (1946), Wyatt and Campbell (1951), and Bruner and Potter (1964) were projected colored slides. Will we see significant delayed recognition with out of focus black and white line drawings?

Experiment 1 is similar to those of Galloway (1946), Wyatt and Campbell (1951), and Bruner and Potter (1964) in that viewers see progressively more focused versions of images and are asked to identify the object in the image, but Experiment 1 used black and white line drawings rather than colored projected slides, a different method of blurring, convolutional rather than optical, and a different manner of presentation, printed rather than projected. Moving to digital sampling of the images gave us more precise control and an opportunity to use alternative types of degradation with the same base images.

The experiment of Bruner and Potter (1964) was done by projecting 35-mm Kodachrome slides onto a projection screen. For each of the eight different images the picture was gradually brought from blurred to less blurred, stopping at a predetermined point where most participants would still not be able to recognize the image. There were three degrees of initial blur and three time periods (13 seconds, 35 seconds and 122 seconds) over which the focus was improved. Participants sat in two semicircular rows at an average distance of 3.5 m from the screen. When the image reached the final degree of focus participants were asked to report what they saw, presumably in writing. The experiment by Wyatt and Campbell (1951), also done in a group setting using projected images, displayed the images at discrete degrees of focus rather than continuously varying the focus. The image was shown for five seconds at each degree of blur with 25 seconds allotted after each five second presentation for describing in writing what they saw.

Experiment 1 using black and white line drawings is a generalization of the earlier experiments (Bruner & Potter, 1964; Galloway, 1946; Wyatt & Campbell, 1951) and serves as a baseline for Experiments 2 and 3. Experiment 2 used an alternative type of degradation we call blocking to create images that would lead to weak perceptual organization and Experiment 3 interspersed images of other objects to see if that would be enough to eliminate the interference.

Method

Participants

The 40 participants in Experiment 1 were students in Introductory Psychology at the University of Delaware. They participated as part of their course requirements. There was no test of visual acuity, but all participants who normally wore glasses or contact lenses did so for the experiment. A sample size of 30 would provide an estimated power of 0.95 based on the results of the Bruner and Potter (1964) experiment, $\eta^2 = .78$. Sample size was calculated using G*Power software (Faul, Erdfelder, Buchner, & Lang, 2009) using $\alpha < .05$ and power .95.

Stimuli

Forty-two pictures, line drawings, were selected from the Peabody Picture Vocabulary Test (Dunn, 1965). Most of the selected pictures were of single objects. No animals, humans or body parts were included. From this set of 42 pictures 30 were chosen for the experiment as detailed below. Each of the 42 pictures was photographed using 35 mm high contrast copy film and each transparency was scanned in 256 levels of grayscale. The resulting digital images were blurred by low pass filtering which in the digital domain is done by convolving the image value with a square filter that averages pixel values, in this case grayscale values. Using this technique, the degree of blur can be increased by increasing the filter size (i.e., the width and height of the "neighborhood" of pixels that are averaged to get the new grayscale value for each pixel). Each resulting digital image was printed on 8.5" x 11" white paper in half-tone form on a Tektronix 4013.

For each of the 42 original images two versions, one severely blurred and one minimally blurred, were created. The intent was to create a range of degradation where the most degraded version of each object would be recognizable by 10% of participants when viewed as a single presentation, and the least degraded recognized by 90% of participants when viewed as a single presentation. Forty volunteers from a child psychology class at The University of Delaware, different participants than in the experiment, tried to recognize the stimuli. Each volunteer saw the more degraded (i.e., more blurred) version of 21 objects and the less degraded (i.e., more focused) version of the other 21 objects. The volunteers were seated in a large lecture room and each had their own booklet of 42 images and an answer sheet.

From the 42 pictures the 30 which came closest to meeting the goal of 10% and 90% recognition were chosen. Four additional pictures were retained for practice trials. Using the results of the single presentation viewing as a guide, five degraded stimuli were created from each original fully focused digital image. The levels of degradation (i.e., the filter sizes) were chosen so that recognition rate of the most degraded version of each object (i.e., D5) would average 10% while the average recognition rate of each object at degradation level D4 would average 30% when seen without previously viewing the D5 version of the object. For D3 we aimed for a recognition rate of 50%, D2 a recognition rate of 70% and for the most focused but still blurry version (i.e., D1) a recognition rate of 90% when viewed without previously viewing any of the more blurred versions of the object.

The process of setting the filter sizes to achieve the goal set forth above was done in two steps using the results of the recognition data collected as previously described. First, the convolutional filter sizes for the most degraded (i.e., D5) and least degraded (i.e., D1) versions were adjusted. If the D5 version of the picture had a recognition rate of less than 10% the filter size was decreased (less blurring) and if it was recognized by more than 10% of participants the filter size was increased (more blurring). The same procedure was followed for the D1 versions. These adjustments were judgement calls. The next step was to choose the filter sizes for the three middle levels of degradation (i.e., D4, D3 and D2). These were calculated so as to have a relatively consistent ratio of pixels in the filter window from one degree of degradation to the next. For example, if the D1 and D5 stimuli were created with filter sizes of 10 x 10 (D1) and 40 x 40 (D5) the three middle stimuli would be degraded using kernels of 14 x 14 pixels, 20 x 20 pixels, and 28 x 28 pixels. In this example the grayscale value of each pixel in the D1 version of the stimulus, the most focused version, would be the result of averaging the grayscale values of 100 (10 x 10) neighboring pixels while each pixel of the D2 version would be calculated by averaging 196 (14 x 14) pixels, D3 400 pixels, D4 784 pixels and D5, the most degraded version, 1600 (40 x 40) pixels. In this case the number of pixels averaged is approximately double as you move from any level of degradation to the next more degraded version. Actual ratios were in the 1.3 to 2.0 range for the full set of 30 stimuli.

The 150 stimuli, 30 line drawings each degraded with five progressively larger filters, were printed on 8.5" x 11" white paper.

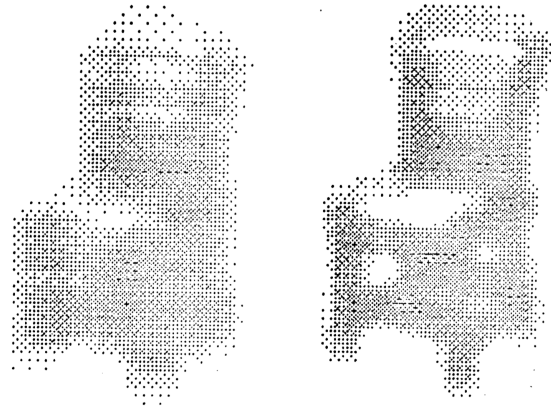


Figure 2: Line drawing of chair degraded with an averaging window of 28x28 (degradation D4) and of 18x18 (degradation D2).

Procedure

Experiment 1 used a within-subject design with five conditions. In condition C1 participants saw only the least degraded version (i.e., D1) of the original image. For condition C2 participants saw the D2 version immediately before seeing the D1 version (i.e., more degraded before less degraded). For C3 participants saw the D3 version of the original image immediately before seeing the D2 version and then they saw the D1 version. The C4 condition started with the D4 version of the original image and the C5 condition began with the D5 image. Thus, the conditions were different in both the degree of degradation of the first stimulus in a series and in the length of the series (we use the term series here even for the C1 condition which is a single presentation). The final image in each series was always the D1 version of the image.

The 30 pictures were divided into five groups of six so that each participant would attempt to recognize six different objects in each of the five conditions. A Latin square design was used to assign groups of stimuli to the five experimental conditions. For any one participant the pictures in each condition were different but across participants each group of pictures appeared equally in all conditions.

The experiment was conducted in a large room with four tables arranged in a square with participants sitting on the inside facing out so that they could not see another participant's stimuli. Each participant was given a three-ring notebook of stimuli and a separate sheet of paper to record their best guess after seeing each stimulus. The 90 stimuli in each notebook were separated by larger (9" x 12") and heavier sheets of blue and orange paper with the color of the paper alternated between stimuli to reduce the chance of exposing a stimulus out of sequence. Participants were given 15 seconds to both view and identify each stimulus with a beep sounding every 15 seconds to signal participants to turn to the next stimulus.

Before starting participants were told that they would be trying to identify pictures that had been altered to make them more difficult to recognize and that they might see an object once or as many as five times and that each time they saw the object it should be more easily recognized but might not seem so. They were also told that the original fully focused pictures were line drawings and a simple line drawing of a house was shown as an example. After seeing the example, the participants tried the practice stimuli to get them familiar with the procedure. The importance of guessing was strongly emphasized but participants were not forced to answer.

Statistical Analysis

Experiment 1 and the other six experiments reported in this paper are unusual in that the number of repeated measures varies by condition. Linear mixed-effects models are good in this situation because they are not only robust to missing data but also to designs with a different number of measures per condition (see Kliegl, Wei, Dambacher, Yan, & Zhou, 2011). For Experiment 1 there were 15 data points per subject, five for condition C5 where the participant saw five stimuli per series, four for condition C4 where the subject saw four stimuli per series and so on. The maximum score for each data point was six as there were six different objects shown for each condition. The factors of "condition", how degraded was the initial stimulus of each series, and "degradation", how degraded was each

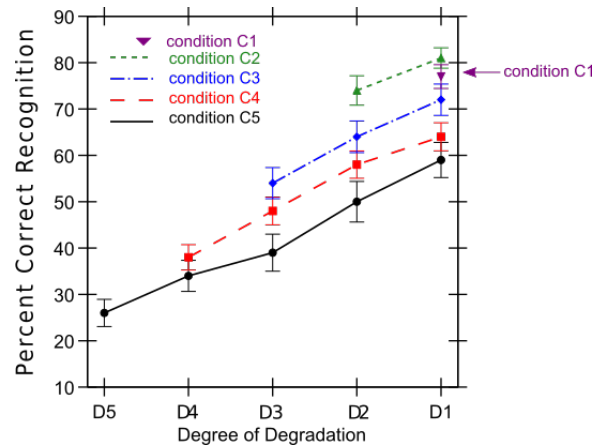


Figure 3: Mean percent correct responses, Experiment 1. Each plot line represents a different within-subject condition with participants seeing 6 different objects in each condition. In the C1 condition the participant saw the object only in its least degraded form. In the C5 condition the participant viewed the object 5 times in a sequence from more to less degraded, guessing after each presentation. Error bars represent the standard error for each data point.

particular stimulus, were considered fixed effects. Subject was treated as a random effect. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question.

Results

Each participant's responses were scored by the experimenter using the criteria established for the preliminary selection of stimuli. Non-responses were treated as wrong answers. If a participant correctly identified a stimulus and then misidentified the stimulus later in the series (i.e., when the image was more focused) the response was marked as correct when misidentified. This occurred only 6 times in 479 possible instances. Scoring this way favors the null hypothesis.

Degradation The first presentation of each object was an unbiased opportunity for recognition. The results associated with these initial presentations represent the simple effects of degradation. Two measures are pertinent: the percentages of responses across levels of degradation (participants were strongly encouraged but not required to respond) and the percentages of recognition (i.e., correct responses). Both measures increased with increased focus. The percentages of responding averaged across all subjects were 69%, 75%, 86%, 95%, and 98% for degradations D5, D4, D3, D2 and D1 respectively. The mean percentages for correct recognition for the first stimulus in each condition, C5, C4, C3, C2 and C1 respectively were 26%, 39%, 55%, 75%, and 79%. These percentages span a narrower range than the expected range of 10% to 90% but still show that the “blurring” technique did produce the expected monotonic function.

Prior Exposure The mean percentage of recognition for each stage of degradation and condition, summed across stimuli and participants, are shown in Figure 3. These percentages are for all trials. Non-responses were considered incorrect responses. Notice that while recognition increases with increased focus, there is a clear detrimental effect associated with beginning with a more degraded stimulus. With only one exception, the mean number of recognitions, across all levels of degradation is greater for participants starting with more focused stimuli. In general, there appears to be a rather constant and consistent difference between conditions which persists through all levels of degradation. This impression is substantiated by comparing the linear mixed-model with and without condition (i.e., degree of initial degradation) as a fixed effect. Removing condition from the model decreased the goodness of fit, as indicated by the likelihood ratio test — effect of condition: $\chi^2(1) = 93.70, p < .001$. The data for analysis was the number of correct responses for each combination of condition and level of degradation, a total of 15 data points per participant.

Stimulus Item Effects The 30 pictures used for this experiment were all black and white line drawings, nevertheless, there was variability among them. Most of the pictures were of single objects but four had multiple objects. Many could be identified by their shape alone, but some required the participant to attend to details. Some were symmetrical while others were quite asymmetrical. These are some of the more obvious qualitative differences.

There also were quantitative differences. In their most degraded state the majority of pictures were identified 25% of the time or less but a few were identified considerably more often. Four were identified 50% of the time and three

in 63% of the presentations. In their most focused state most of the pictures were identified at least 75% of the time; seven were identified less frequently.

It is reasonable to wonder if simply offering a response on the initial view influences later recognition. If, as we hypothesize, it is the initially formed organization that is responsible for delayed recognition, and not the simple offering of a response, we would not see a difference in recognition rates for trials on which the participant guessed and trials where participants did not guess. In this experiment the mean probability of a correct response following an incorrect response was .44 and following no response it was .43. So, to the degree that not offering a response is the same as not having some guess about the object's identity, the data are supportive of some *non-naming* cause for the interference in recognition.

What about the grouping and figure-ground errors that we believe underlie the delayed recognition? In Experiment 1 participants erred in perceiving the number of objects in a picture; they erred by taking the wrong perspective; and they erred by reversing figure and ground. Mistaking the number of objects in a picture was common when more than one object was present. The picture of two suitcases was frequently seen as a camera, presumably with the larger suitcase in the background seen as the body and the smaller suitcase as the lens, the picture of two cars was often seen as some type of insect, and the picture of flowers was sometimes seen as an animal. When participants made errors in perspective they tended to see objects as if they were looking down on them while in fact they were looking at a front view. Examples include seeing the saddle as an airplane and the coffee cup as a piano. Sometimes the organizational error is not easily described, as in the case of the smoking chimney being seen as a clown's face.

Discussion

The results of Experiment 1, in line with the results of Bruner and Potter (1964), show that recognition of a "blurred" object is less likely if viewing the blurred object is preceded by viewing an even more blurry version of the object. The results presented in Figure 3 are similar to those presented by Galloway (1946) and Wyatt and Campbell (1951) who also used discrete degrees of degradation. We also found that participants in this experiment, like those in Davison's (1964), make organizational errors in grouping, figure-ground and perspective. If delayed recognition is caused by the inaccurate organization of elements it follows that we can eliminate the delay in recognition by eliminating the organizational errors. Experiment 2 attempted to do that.

Experiment 2

The second experiment was undertaken to test the hypothesis that participants will be less influenced by prior presentations when the stimuli elicit weaker, less stable organizations. The strength and stability of an organization is determined by the cohesiveness of the parts forming the whole. When the elements follow the Gestalt principles of similarity, proximity, good continuation and closure the global structure is easily seen and this is termed "good form". Images with less than good form, those lacking continuity and low spatial frequencies, will elicit weaker organization which is slower to develop. Good form requires low spatial frequencies (Lagasse, 1993) so for Experiment 2, an alternative manner of degradation, one that lacked low spatial frequencies and had reduced continuity, was employed to test the hypothesis that good form in a degraded image was the primary factor in delayed recognition. Compare the "block" and "blur" images of Figure 4 to see how even with relatively less degraded images there is less confusion about continuity in the blurred image. Also, notice how much quicker the global shape is seen in the blurred image as compared to the block image.

The alternative form of image degradation used in Experiment 2 reduced low spatial frequencies by creating images from solid black squares. In the more degraded states the images were composed of relatively few elements, reducing continuity and lessening the likelihood of good form (Lagasse, 1993; Kimchi, 1998). Less degraded images had decreased element size and more elements (Figure 5), thus increasing continuity and making the object more easily recognized. Squares were used, rather than circles, as they interfere more with global organization (Palmer, 1980).

Method

Participants

Thirty participants, 15 female and 15 male students in Introductory Psychology at the University of Delaware participated as part of their course requirements. A power analysis using the results of Experiment 1, (Cohen's $d = 1.07$, $\alpha < .05$, power .95 and a two tailed t-test) showed that a sample size of 14 would be sufficient. This is a smaller

sample size than the estimate based on the Bruner and Potter (1964) data which used fewer stimuli and different procedure but a similar degradation. However, in all likelihood, 14 participants would not yield sufficient power with the alternative degradation used in Experiment 2. Power was calculated using G*Power software (Faul et al., 2009).

Stimuli

The stimuli for Experiment 2 were generated from the same 30 line drawings as for Experiment 1 but the stimuli were degraded by increasing pixel size rather than blurring. Figure 5 shows chair stimuli as they appeared in Experiment 2 at degradation levels D5 and D2 (cf. Figure 2). Degradation was increased by increasing the size and reducing the number of pixels composing a picture. Fewer pixels of larger size decrease continuity and reduce detail and thus a more degraded stimulus. Reducing the size of pixels and increasing the number of pixels has the opposite effect, increased continuity and increased detail. A rendering with pixels of size 1 x 1 would essentially give us the original line drawing.

The stimuli were created from the same grayscale data generated for Experiment 1. To create a stimulus we started by taking a grayscale matrix dividing it into blocks of N x N size. The grayscale values of all of the pixels in each block were averaged and if the average was above a predetermined threshold all of the pixels in that block were set to black, effectively creating a larger pixel. If the average was below threshold all of the pixels in the block were set to white.

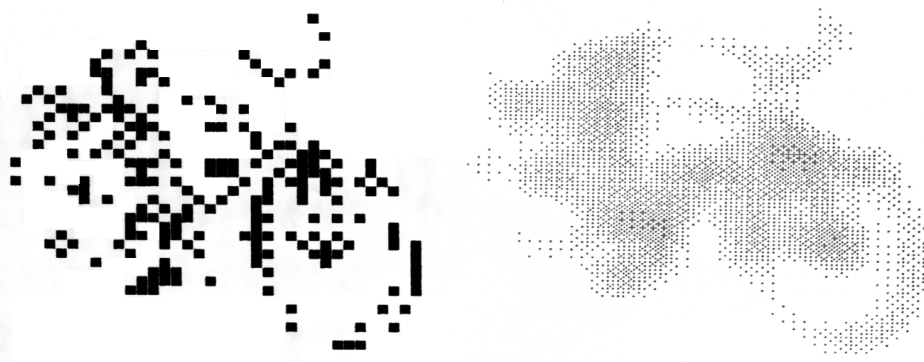


Figure 4: Comparison of bicycle rendered with the block and blur degradation.

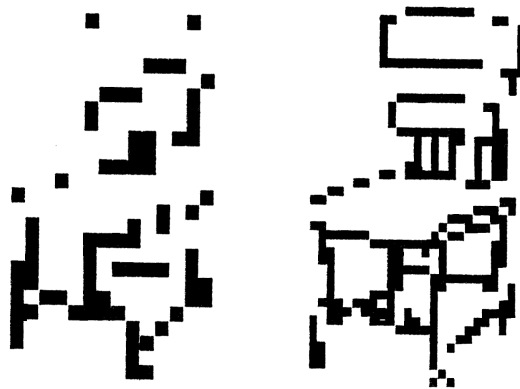


Figure 5: Line drawing of chair degraded by averaging over 7x7 and 4x4 pixel blocks, stimuli D5 and D2 respectively.

For each of the 30 objects three stimuli with different pixel sizes were created for viewing by 30 participants, different participants than in the experiment. Each participant viewed 10 different objects at each of the three degrees of degradation. The frequency with which each picture was recognized at each of the three degradations was used as

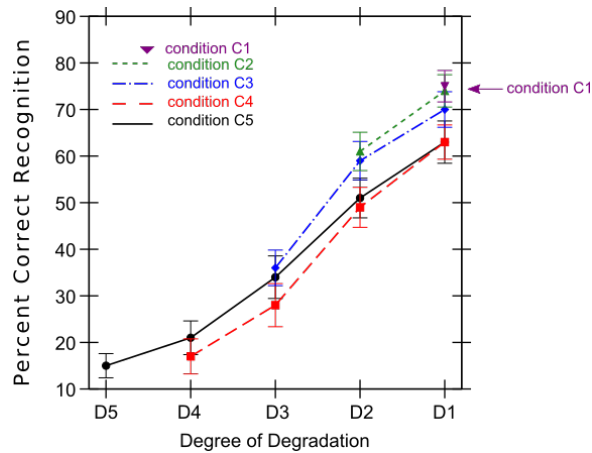


Figure 6: Mean percent correct responses, Experiment 2. Each plot line represents a different within-subject condition with participants seeing 6 different objects in each condition. In the C1 condition the participant saw the object only in its least degraded form. In the C5 condition the participant saw the object 5 times from more to less degraded, guessing after each presentation. Error bars represent the standard error for each data point.

a guide for the final determination of block size for each degree of degradation. As in Experiment 1 a series of five stimuli were created for each object, D5 being the most degraded and D1 being the least degraded. The degrees of degradation were chosen so that the mean frequency of recognition at each degree of degradation would approximate the mean frequency of recognition found in Experiment 1.

Procedure

Both the design and procedure were identical to those of Experiment 1.

Results

Degradation

The percentages of responding for degradation stimuli D5 through D1 were 52%, 63%, 82%, 86%, and 92% respectively. Mean percentages of recognition were 16%, 18%, 37%, 63%, and 77% (see Figure 6). Both measures are slightly lower at all degradation levels than those from Experiment 1.

Prior Exposure

The data were analyzed in the same manner as in Experiment 1, by comparing the linear mixed-model with and without condition (i.e., initial degree of degradation) as a fixed effect. Removing condition from the model decreased the goodness of fit, as indicated by the likelihood ratio test — effect of condition: $\chi^2(1) = 9.93, p = .002$, showing some delayed recognition when viewing more degraded images before less degraded with our block degradation.

With both experiments showing a significant effect of initial degradation, though looking to be a much weaker effect with the block stimuli, another analysis was done to directly compare the results of the two experiments. This analysis looked at the difference in a participant's performance at each level of degradation and condition. Consider the simple case where there are only two levels of degradation. The measure of interest is the difference in recognition rate at the least degraded level when the image is seen only at that level and when previously seen in a more degraded state. For the analysis we generated difference scores for each participant at each level of degradation by taking the score for each condition and subtracting the score for the same level of degradation for the next longer condition (see Figure 7). This results in 10 difference values in total for each participant. We compared linear mixed-models with and without degradation type (blurred images vs block images) and found them to be statistically significantly different: $\chi^2(1) = 4.6219, p = .03$. So, while the block stimuli did produce some interference in recognition it was significantly less interference than found in Experiment 1 using blurred images.

Stimuli

The probability that a stimulus would elicit a response at degradation D5 ranged from 0% to 100%. At degradation D1 21 of the 30 stimuli elicited responses 100% of the time. In general there is slightly more variability than found

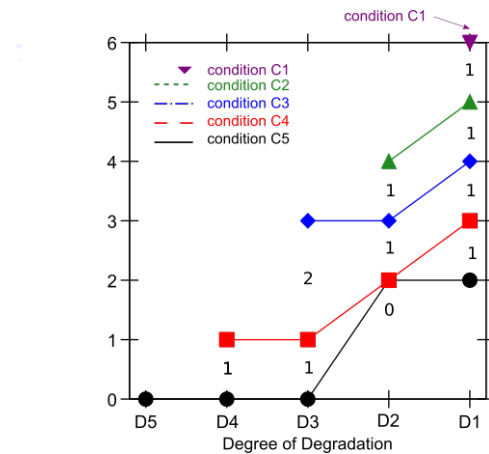


Figure 7: Hypothetical difference score calculation for one participant. In this figure the numbers between data points are the difference scores, 10 per participant. Each difference score is calculated by comparing the number of correct responses between within-subject conditions at each degree of degradation. This means that a participant's number of correct responses for condition C2 are subtracted from the number of correct responses in condition C1, condition C3 scores are subtracted from condition C2 scores and so on. More specifically the number of correct responses the participant had for condition 2 at degradation degree 1 (i.e., C2D1) is subtracted from their score for C1D1, a difference of 1 in this example. The score for condition C3 at degradation D1 (i.e., C3D1) is subtracted from the score for C2D1 and C3D2 is subtracted from C2D2 and this continues with the final difference scores being the differences between conditions C4 and C5 at D4, D3, D2 and D1.

with the blurred stimuli. For the 30 objects the frequencies of responding averaged across degradation levels correlated significantly with those from Experiment 1, $r = .36, p < .05$. This correlation shows that while the manner of degradation is different certain inherent characteristics of the pictures are not lost. As with the probability of responding, the probability of recognition across the 30 sets of stimuli showed greater range than in Experiment 1.

Discussion

The results of this experiment show that the lower rate of recognition found with serial presentation of blurred pictures may be associated with low spatial frequency information and not degradation more generally. Low spatial frequencies lead to good continuation, good form (Lagasse, 1993) and stronger global shape-level perception, but the images can be ambiguous enough to lead to relatively strong misinterpretations. With the block degradation you see a jumble of minimally continuous blocks. The block degradation leads to weak global organization and is less influential in later perception than stimuli that lead to strong organization.

Experiment 3

The participants in the Bruner and Potter (1964) experiment saw images gradually coming into focus without interruption and they were only asked to offer a guess at the final state of focus. The participants in the Wyatt and Campbell (1951) experiment saw images for 5 seconds at each discrete step of degradation and had 25 seconds to answer after each trial. Both experiments showed significant delay in recognition as did our own Experiment 1, but would we see this delay if other stimuli were interspersed between trials? Non-verbal stimulus priming can last quite a long time, see for example a study by Biederman and Cooper (1991) who showed non-verbal priming with approximately 7 minutes between first and second presentations. Experiment 3 sought to see if the top-down processes that are responsible for delayed recognition continue to influence over a delay of up to 105 seconds and the interspersing of up to seven stimuli of other objects.

Method

Participants

Twenty participants, 10 female and 10 male students in Introductory Psychology at the University of Delaware participated as part of their course requirements. A sample size of 14 should be sufficient based on power analysis of

results from Experiment 1 (Cohen's $d = 1.07$, $\alpha < .05$, power .95 and a two tailed t-test) using the same type of stimulus degradation

Stimuli

The stimuli were the same as those used in Experiment 1 (i.e., low-pass filtered, see Figure 2).

Procedure

The design was the same as in Experiment 1 with the exception of stimulus order. Stimuli were interspersed so that a participant never saw two versions of the same object in succession. For 15 of the 30 objects the stimuli in an object series were each separated by a single stimulus of some other object, for example, **Windmill D3**, Bicycle D3, **Windmill D2**, Broom D3, **Windmill D1**. The stimuli of the other 15 object series were each separated by 7 other object stimuli (6 when 7 was not possible), for example, **Chair D5**, Lamp D3, Mug D4, Lamp D2, Mug D3, Chimney D1, Apple D2, **Chair D4**. The stimuli used in each condition were counterbalanced across participants.

The procedure was also modified so that though participants did not see the same object successively they would know their last response to the upcoming stimulus. For Experiment 3 the colored paper that separated stimuli was cut so that participants would see the identifying letter of the next stimulus object before turning to that stimulus. Participants were instructed to check this identification before turning the page. Participants were also required to check their answer sheet for their last response to the picture before turning to the new stimulus. In all other ways the procedure was identical to that of Experiment 1.

Results

Degradation

The stimuli were the same as in Experiment 1 so the statistics associated with initial presentations (pure effect of degree of blur) should be quite similar to those from Experiment 1. The percentages of responding for degradation levels 5 through 1 were 68%, 85%, 88%, 95%, and 98%, nearly identical to the percentages for Experiment 1. Mean percentages for correct identification were 15%, 42%, 60%, 68%, and 83%, some a bit lower and some a bit higher than in Experiment 1.

Prior Exposure

The mean percentages of correct responses for each stage of degradation and condition, summed across stimuli and participants, are shown in Figure 8 (panel a). Visual inspection suggests substantial interference though less than found in Experiment 1. The linear mixed-model analysis included condition, degree of degradation and interval duration (short - one intervening stimulus; or long - 7 intervening stimuli) revealed a statistically significant effect of condition but no difference between the short and long interval conditions. Dropping condition from the model decreased the goodness of fit, as indicated by the likelihood ratio test — effect of condition: $\chi^2(1) = 35.70$, $p < .001$. Dropping interval length from the model resulted in a likelihood ratio of $\chi^2(1) = 1.0365$, $p = .31$. Figure 8 (panels b & c) shows average percent correct, by condition and degree, for each of the two interval conditions.

We also compared the results of Experiment 1 and Experiment 3 directly using each participant's difference scores (see Experiment 2 Results for a description of difference scores) and found no statistically significant difference, $\chi^2(1) = 0.269$, $p = .60$

Discussion

The results of Experiment 3 demonstrate that the decreased rate of recognition associated with the serial presentation of blurred pictures is not diminished by interspersing the presentations of stimuli from one object series with the stimuli from another object series. This is contrary to what was found by Snodgrass and Hirshman (1991) and Schulkind (2000). We attribute the difference in results to the difference in the manner of degradation. Degradation by low-pass filtering leads to strong organization while fragmentation leads to weak organization and therefore degradation by fragmentation is more easily disrupted by adding an interval of time between stimuli or adding intervening viewing of other objects. In Snodgrass and Hirshman's (1991) experiment using fragmentation the interval between trials was filled with a simple arithmetic problem and in Schulkind's (2000) experiment the interval was extremely brief, 300 or 1500 ms, and yet neither experiment showed delayed recognition. On the other hand, Experiment 3 using low-pass filtering showed delayed recognition even with an interval of nearly 2 minutes with the participant viewing other objects during the interval.

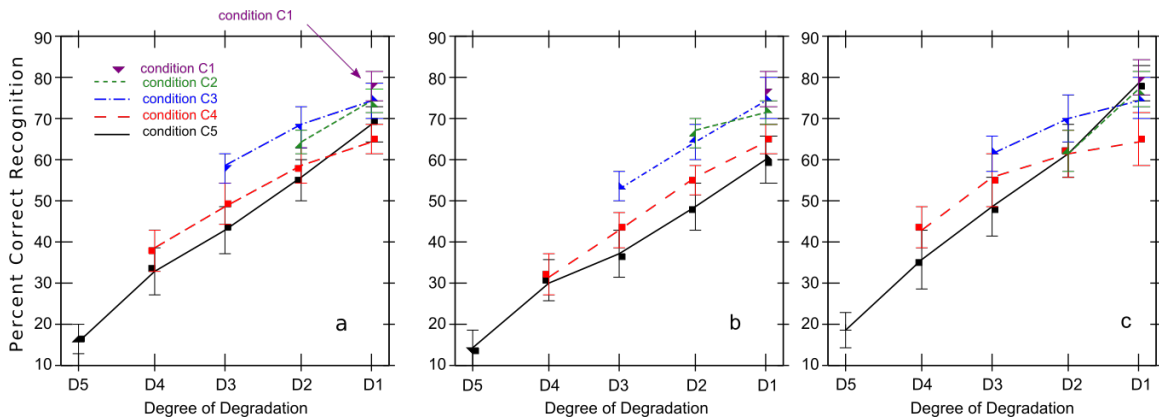


Figure 8: Mean percent correct responses, Experiment 3, a) collapsed across interval length, b) short interval, c) long interval. Each plot line represents a different within-subject condition with participants seeing 6 different objects in each condition. In condition C1 the participant saw the object only in its least degraded form. In condition C5 the participant saw the object 5 times from more to less degraded, guessing after each presentation. Error bars represent the standard error for each data point.

Experiment 4: Recognition of low-pass filtered images varying in complexity and number

Experiment 4 was the first in a second series of experiments undertaken largely to directly test the difference between low-pass filtering and fragmentation. Experiment 2 showed that an alternative degradation to low-pass filtering, we call block degradation (see Figure 5), significantly diminished the negative effect of seeing degraded images. It is similar to fragmentation in that continuity and good form are diminished when compared to low-pass filtered images, but the blocking degradation lacks the high spatial frequencies of simple fragmentation. The experiments by Snodgrass and Hirshman (1991), Luo and Snodgrass (1994) and Schulkind (2000), all using fragmentation, had mixed results, sometimes showing delay and sometimes not. These results are not surprising as fragmentation will lead to weaker grouping and figure-ground organization than low-pass filtering. However, because the stimuli used by those researchers and the procedure are somewhat different from those of the first three experiments, we do not have a direct test. Thus, it made sense to run a second series of experiments that compared low-pass filtering and fragmentation using the same base stimuli and procedure, only varying the manner of degradation. Experiment 4 uses low-pass filtering and Experiment 5 uses fragmentation.

Experiment 4 also looked at the factors of object complexity and the composition of the image. If errors in grouping and figure-ground are responsible for delayed recognition then objects and composition which promote organizational errors should lead to more delay. We hypothesized that “complex” objects would lead to more delay than “simple” objects because they would increase the likelihood of organizational errors. We also predicted, based on the protocols of Davison (1964) and our own protocols, that having multiple objects in an image would lead to more delay than having a single object, again because it would likely increase grouping and figure-ground errors.

To compare low-pass filtering with fragmentation we needed a new baseline experiment using the materials and procedure just described. We also wanted to try and induce grouping and other perceptual errors as a more direct test of the hypothesis that these errors lead to delayed recognition. The error analysis by Davison (1964) and our own informal analysis led us to think that manipulating image complexity and the number of items in an image might give us a knob to turn up or down the frequency of perceptual errors. We chose a set of undegraded stimuli that had been rated on complexity and we generated additional stimuli to manipulate multiplicity.

Method

Participants

Twenty-four undergraduate students at the University of Colorado who received course credit participated in the experiment.

Stimuli

Pictures were selected from the Rossion and Pourtois (2004) picture set which contains enhanced color versions of the pictures in the Snodgrass and Vanderwart (1980) set. Rossion and Pourtois had French speaking psychology

students name each object and had two additional groups of students rate each object on familiarity and complexity. For the series of experiments reported in this paper the set of 260 was first pared to 176, by eliminating any people or body parts and then by applying the following criteria: Name agreement $\geq 85\%$, complexity standard deviation ≤ 1.0 . From the reduced set of 176, two sets of 32 were chosen, a low-complexity set, mean rating 1.9, and a high-complexity set, mean rating 3.8 (scale 1 to 5, very simple to very complex). Ideally the mean familiarity ratings for the two sets would be the same, in fact they were 3.72 and 3.13 respectively (scale 1 to 5, very unfamiliar to very familiar). Three additional pictures were chosen for practice items.

Every picture in the Rossion and Pourtois (2004) set was of a single object, for example, one stalk of asparagus (simple), one toaster (complex). See Figure 9.



Figure 9: Example single object pictures - simple left, complex right.

A parallel set of stimuli was created with each picture containing two or more of the same object, for example, four toothbrushes, three elephants, two pianos. Figure 10 shows examples of stimuli with multiple objects. The positioning of the multiple objects varied in an unsystematic way but was generally done so as to enhance the likelihood of confusion in number. This gave us four sets of objects, 32 single-simple, 32 single-complex, 32 multiple-simple and 32 multiple-complex.

Multiple object images naturally subtend a larger visual angle than single object images and this might be a substantial cue to participants. To compensate for the image size differences each of the single object stimuli were scaled up by a factor of 1.22 in width and height and multiple object images were scaled down by a factor of .87. These scaling factors were based on the average bounding box ratio of 2:1 when comparing the multiple object images with single object images.

The seven degrees of degraded stimuli for each fully focused single or multiple object original were generated by first converting each to grayscale and then low-pass filtering by convolution. The convolutional kernel weighted each neighboring pixel value equally, that is, using a box blur, also called a box linear filter. The degree of image blurring was controlled by varying the size of the convolutional kernel. A larger kernel, one averaging more pixels, produces more blur. For each picture a maximum and minimum kernel size was chosen based on pilot data that sought to find kernel sizes that produced 20% and 80% recognition when the object was seen as the only or first example of the object in a series. Intermediate kernel sizes, for intermediate degrees of blurring, were calculated by dividing the range between minimum and maximum into divisions that were equivalent percentage changes. Figure 11 shows the D7, D4 and D1 stimuli of the clothespin series from the single-simple set.

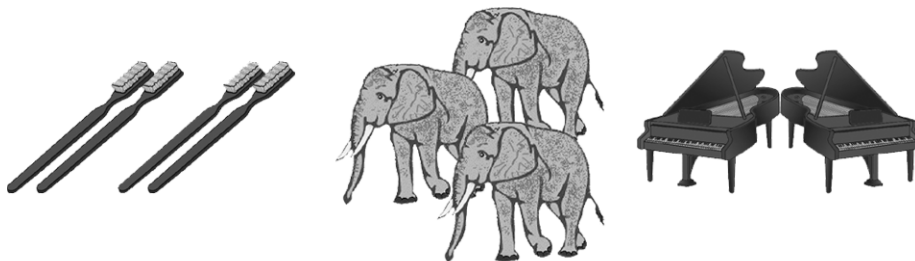


Figure 10: Example multiple object pictures -simple left, complex center and right.



Figure 11: A series of 3 trials showing low-pass filtered clothespin.

Procedure

Experiment 4 used a within-subject design with four conditions. In condition C1 participants saw only the least blurred version (i.e., D1) of the original image. For condition C3 participants saw the D3 stimulus before seeing the D2 stimulus, before seeing the D1 stimulus (i.e., more degraded before less degraded). The C5 condition started with the D5 version of the original image and was followed by D4, D3, D2 and D1. The C7 condition began with the D7 image. Thus, the conditions were different in both the degree of degradation of the first stimulus in a series and in the length of the series (we use the term series here even for the C1 condition which is a single presentation). The final image in each series was always the D1 version of the object. Each stimulus was shown for 3 seconds and participants were asked to make a best guess after each stimulus was erased by using the computer keyboard to enter an answer in the text field on the screen. There was no time limit on answering.

In total each participant was shown 64 objects over the course of 256 trials. A trial being the viewing of a single stimulus. Stimuli from the same series (i.e., levels of degradation of the same object) were shown consecutively. The full set of 64 objects were divided into two sets and 12 of the participants saw one set of 32 objects in single form (e.g., one toothbrush, one shoe, etc.) and one set in multiple form (e.g., two pianos, three elephants, etc.) The other 12 participants saw the opposite sets so that no saw the same object in both single and multiple form. Within each set half of the objects were from the simple category and half from the complex and 25% of objects were presented in each of the four conditions, C1, C3, C5 and C7. The ordering of objects was random for each participant.

Stimuli were displayed on a 17" Macintosh iMac with LCD display of 1440 x 900 pixels. The images after scaling were approximately 345 x 345 pixels including the white space around each object or objects. Each participant was instructed to find a comfortable distance from the screen which generally was about 24". At this distance stimuli would be within a maximum visual angle of 8 degrees but for most objects was within 5 degrees of visual angle. Participants were further instructed to refrain from moving their head side-to-side or nearer or farther from the screen. They read a set of instructions explaining that they would be seeing images of objects that they would normally be able to recognize but that the pictures were modified to make recognition more difficult. The instructions also explained that they would most often see an object more than once but that they would not know how many times and they should pay close attention on each presentation as it might be the last one for that object. The task was simply to look at each picture and try to identify the object or objects and to type their answer in the box that appeared after the image was erased. The experimenter also paraphrased the instructions and added that they were not being timed and that it was self-paced, the next image would not appear until they pressed the return key after entering their answer.

The practice trials included three objects, one shown in a series of five degradations, one in a series of three degradations and one at a single level of degradation. The practice items were a single guitar, a single tiger and a pair of shoes. The experimenter stayed with the participant during the practice to answer questions and ensure that the procedure was understood. Participants were not given feedback on their performance on the practice trials or on the experimental trials, but the goal of the research was explained after the experiment.

Picture names were considered correct if they were a variant or an obvious misspelling. Examples of variants include: bicycle and bike or traffic light and stoplight.

Statistical Analysis

Analysis of data for Experiment 4 was done in the same manner as for Experiments 1 to 3, using a mixed-model. For the analysis we used the participants scores from the 1st, 3rd, 5th and 7th stimuli of each series. This resulted in 10 data points per subject, four data points for condition C7, three for condition C5, two for condition C3 and one

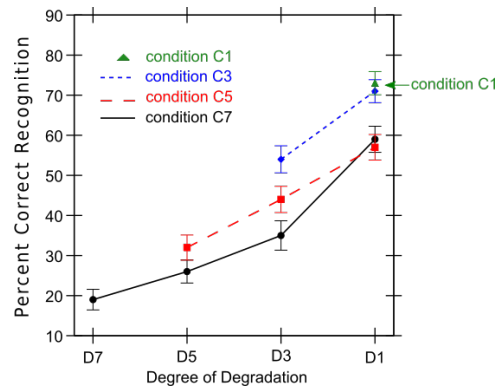


Figure 12: Mean percent correct responses, Experiment 4. Each plot line represents a different within-subject condition with participants seeing 16 different objects in each condition. In condition C1 the participant saw the object only in its least degraded form. In condition C7 the participant saw the object 7 times from more to less blurred, guessing after the odd number presentations. Error bars represent the standard error for each data point.

for condition C1. The 2nd, 4th and 6th stimulus of each series provided a smoother transition between degrees of degradation but were not valuable in the analysis because they were never the first in a series.

Results

Figure 12 shows the mean percentage of correct responses averaged across participants. The average recognition rates on first exposures for the four conditions were 19%, 32%, 54% and 73%, close to the target range of 20% to 80% but Figure 13 reveals that there were large differences by stimulus type. Average correct response rates for the most degraded versions (i.e., D7) ranged from 5% to 41% and for the least degraded (i.e., D1) the range was 51% to 78%. However, the difference in correct response rate between most degraded and least degraded within stimulus type was nearly the same across stimulus types, values of 38%, 38%, 46% and 40%.

A linear mixed-model with fixed effects for condition, object multiplicity, object complexity and degree of degradation showed all to be statistically significant factors. Starting with a more degraded image reduced the user's likelihood of correct recognition when the image became more focused, the same phenomenon seen in all of the previous experiments using low pass filtered images, $\chi^2(3) = 63.33, p < .001$. Having multiple elements in the image made recognition more difficult but did not interact with condition (i.e., did not lead to further delayed recognition), multiplicity: $\chi^2(2) = 238.10, p < .001$, multiplicity x condition: $\chi^2(1) = 0.643, p = .42$. Complexity made recognition easier and like multiplicity did not interact with condition: complexity: $\chi^2(2) = 66.309, p < .001$, complexity x condition: $\chi^2(1) = 0.1915, p = .66$. Figure 13 is a graph of multiplicity x complexity x condition and it shows that regardless of condition complexity made recognition easier and multiplicity made recognition more difficult. We had expected both multiplicity and complexity to interact with condition to increase the degree of interference found when starting with a more degraded image but neither multiplicity nor complexity showed a statistically significant interaction with condition.

Discussion

The results of Experiment 4 are consistent with the results of Experiments 1 and 3, showing that images of low-pass filtered objects are more difficult to identify if preceded by viewing even more degraded versions of the same objects and shows that the phenomenon is robust to manner of presentation, whether projected, printed or displayed on a liquid crystal display. The hypothesis that more complex stimuli would amplify delayed recognition in comparison to stimuli rated as simple was not confirmed. The absolute percentages of correct responses were different but there was no interaction of factors. Likewise, having multiple objects in the images did not add to delayed recognition beyond what was found with the stimuli of single objects. Looking at figure 13 it is clear that while the ranges are different the slopes of the lines are roughly parallel showing that delayed recognition occurs over a wide range of degradation.

In all likelihood the manipulation of stimuli, along the dimensions of complexity and multiplicity, did not create the expected organizational errors. The multiple object stimuli were designed to induce grouping errors and lead to errors in object count but only 11% of incorrect first exposure responses revealed a mistake in object count which

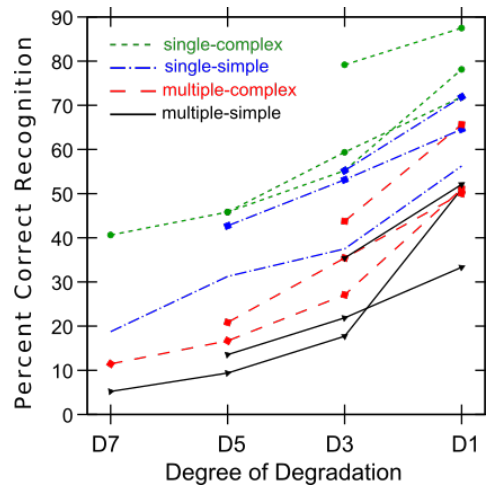


Figure 13: Mean percent correct responses, Experiment 4. Each of the 12 lines shows the mean correct response across subjects in a for one of 3 within-subject condition and one of the 4 combinations of stimulus type (single-simple, single-complex, multiple-simple and multiple-complex). Each participant saw 4 series of objects for each of these 12 condition-stimulus combinations. In condition C7 the participant saw the object 7 times from more to less blurred, guessing after the odd number presentations. In condition C5 the series was of length 5 and for C3 length 3.

suggests that the manipulation was not successful. The complexity manipulation may have failed because even the “complex” stimuli were quite simple compared to those used by Bruner and Potter (1964). A toaster or a tree is not complex compared to something like an aerial view of a town and cloverleaf interchange.

Experiment 5: Recognition of fragmented images

Experiment 4 demonstrated strong recognition interference when participants saw a series of images where each successive image was less blurry than the previous. Degradation by fragmentation retains all of the spatial frequencies for those areas not deleted from the image in contrast to blurring which keeps all of the image but removes the high-spatial frequency information. Thus, to test the hypothesis that it is the low spatial frequencies that are responsible for the recognition interference, degradation by fragmentation was used in Experiment 5 following the same procedure as Experiment 4.

The fragmentation for Experiment 5 was done by dividing the image into blocks and displaying only a portion of the blocks. An image degraded by fragmentation looks very different from one degraded by low-pass filtering and it is not conducive to forming a strong gestalt due to breaks in continuity. Fragmentation is similar to the degradation used in Experiment 2 which also broke the image into blocks, and which significantly diminished the interference generated from seeing objects in a more to less degraded sequence.

The strength and stability of an organization is determined by the cohesiveness of the parts forming the whole and strong organizations, ones with good form, are composed of elements which are weak by themselves but meaningful when grouped. A picture composed of elements that exert strong local influences will not have good form (Palmer 1980). Fragmented images retain the strong local influence of the high frequency spatial information and the fragmentation technique breaks the image into large noncontiguous squares that break global continuity. Low-pass filtering, on the other hand, creates an image that allows for a strong but possibly incorrect grouping of the elements. Thus, it was hypothesized that degrading images by fragmentation would eliminate or significantly reduce visual interference in this serial presentation paradigm.

Method

Participants

Twenty-four students from the University of Colorado participated for course credit. None of the participants were in Experiment 4.

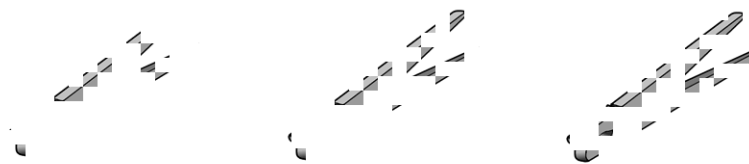


Figure 14: A series of 3 trials showing a fragmented clothespin.

Stimuli

The base images from which the stimuli were generated were 32 of the 64 from Experiment 4, half of the simple objects and half of the complex objects with half of each of these sets having single objects and half having multiple objects. The stimuli were degraded by dividing each original image of approximately 345 x 345 pixels into blocks of 15 x 15 pixels. In the most degraded state (i.e., D7) only 10% of blocks were shown and in the least degraded state 30% (i.e., D1). Only blocks that included object pixels were included in the percentage calculation. On each successive trial after the first presentation of an object series blocks were added to those already visible. The selection of initial blocks and “add-in” blocks was random across participants. Figure 14 shows the D7, D4 and D1 stimuli of the single clothespin series.

Procedure

The procedure, software and apparatus for Experiment 5 were identical to the procedure, software and apparatus of Experiment 4.

Results

Figure 15 shows the mean percentage of correct responses averaged across participants. The average recognition rates on first exposures for the four conditions were 11%, 31%, 52% and 65%. This range of 11% to 65% is lower than the range of 19% to 73% of Experiment 4. The null hypothesis was that there would be no benefit or detriment to seeing the stimuli in series regardless of the degree of fragmentation of the initial stimulus in the series. The null hypothesis was not disproved. The likelihood ratio test comparing a model with fixed effects for condition and degree of degradation and a model without condition, $\chi^2(1) = 1.6575, p = .20$, showed no effect of condition.

The data of Experiments 4 and 5, low-pass filtering and fragmentation, were compared directly using a mixed-model of difference scores as done for previous between experiment comparisons. The result confirms a statistically significant difference in recognition interference based on degradation type, $\chi^2(1) = 39.317, p < .001$. The analysis was done by comparing each subject’s difference scores at each level of degradation and condition. Difference scores are calculated for each participant by comparing their performance at each level of degradation by taking the score for each condition and subtracting the score for the same level of degradation for the next longer condition (Experiment 2 includes a more detailed explanation of difference score calculation, also see Figure 7). This resulted in 6 difference values for each participant.

Discussion

Comparing the results of Experiment 4 with those of Experiment 5 we have the first direct test of degradation using low spatial frequency filtering and degradation using fragmentation, and we see that the results are significantly different, and critically, that the fragmentation condition produced no detriment in recognition when seeing the more fragmented images before the less fragmented. This is not to say that fragmentation cannot produce delayed recognition, the results were in that direction, but rather that fragmentation is less likely to lead to delayed recognition because fragmented images have strong local elements that do not create good global form and that the most interfering images are those that have a strong misleading global form which is facilitated by low spatial frequencies. We must also consider that the failure to find delayed recognition could be due to a difference in absolute recognition rates which were lower in Experiment 5 than Experiment 4, though this seems unlikely. The four different stimulus types

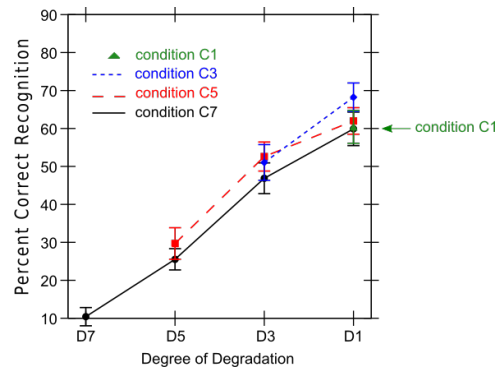


Figure 15: Mean percent correct responses, Experiment 5. Each plot line represents a different within-subject condition with participants seeing 8 different objects in each condition. In condition C1 the participant saw the object only in its least degraded form. In condition C7 the participant saw the object 7 times from more to less fragmented, guessing after the odd number presentations. Error bars represent the standard error for each data point.

of Experiment 4 span a wide range of absolute recognition but similar between condition differences suggesting that delayed recognition is a robust phenomenon that occurs over a wide range of recognition rates (see Figure 13).

Like Experiment 5, Experiment 2 also attempted to eliminate recognition delay by altering the manner of degradation. Why was the fragmentation degradation of this experiment more successful than the “block” degradation of Experiment 2 which showed a significantly diminished effect but still a significant delay? The main difference is the maintenance of high spatial frequency information in the fragmented images used in this experiment as compared to the solid black squares which have no high frequency information other than their edges. Dropping the high spatial frequency information in the “block” degradation manipulation meant that the blocks needed to be smaller to maintain reasonable levels of recognition and having a larger number of smaller elements leads to stronger organization which in turn leads to more interference. These two factors, lack of high frequency information and smaller elements, would place the stimuli of Experiment 2 somewhere between low-pass filtering and fragmentation in producing strong global perceptual organization, and thus block degradation would more likely to lead to delayed recognition than full fragmentation, but less so than low-pass filtering.

Experiment 6a: Recognition of low-pass filtered images, 100 ms presentation

Errors in grouping and figure-ground should not be a factor in delayed recognition if top-down processes do not influence these low level aspects of visual processing but there is good evidence that both factors can be primed by previous experience (Kimchi & Hadad, 2002; Razpurker-Apfeld & Kimchi, 2007; Peterson & Gibson, 1994; Biederman & Cooper, 1991). This hypothesis was tested in Experiments 6a and 6b by using a 100 ms stimulus presentation, comparing unmasked and masked presentation, to see if limiting top-down influence via masking would limit delayed recognition.

In Experiments 4 and 5 each stimulus was presented for 3 seconds and the prior research used presentation times of at least 1 second, but usually much longer. Would a presentation time of 100 ms lead to delayed recognition?

The perception of figure-ground follows a reliable time course. The sequence of activity moves from local feature response to boundary detection to figure-ground where neurons within the entire area of the figure show enhanced activity compared to the background area (Lamme, 1995; Lamme, Rodriguez-Rodriguez, & Spekreijse, 1999; Supèr, Spekreijse, & Lamme, 2001; Zipsper, Lamme, & Schiller, 1996). Lamme et al. (1999) found this process takes about 150 ms with relatively flat activity before 80 ms after which there is a strong response from the neurons with receptive fields covering the border and then neural response moving “inward” until the activation is relatively flat across the entire area of the figure. The boundary enhancement can be explained by lateral inhibition but lateral inhibition on its own would lead to depressed activation on either side of the boundary, whether figure or background. The delay in the figure-ground enhancement suggests that top-down cortical connections might be responsible. Support for the excitation in the full area of the figure coming from recurrent connections was found by Supèr et al. (2001) who compared neural activity in V1 in monkeys doing a figure detection task. This excitation of figure is not due to attention (Poort, Raudies, Wannig, Lamme, Neumann, & Roelfsema, 2012) and not to the decision process (Supèr

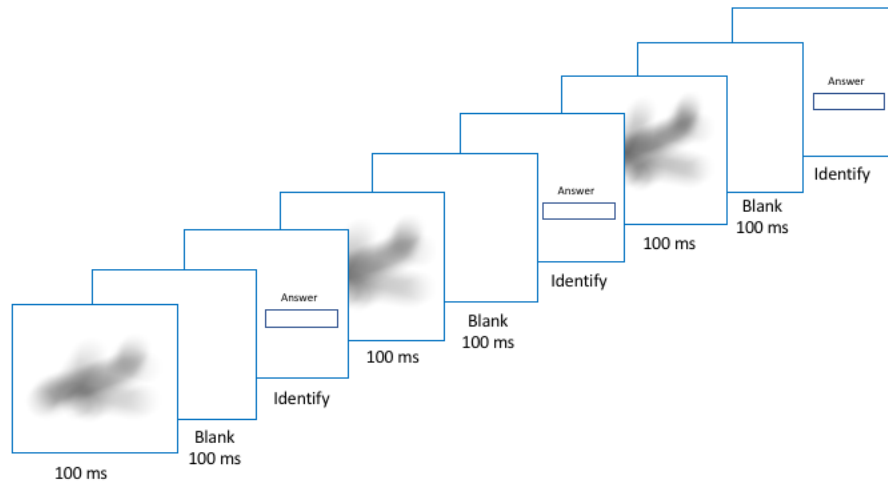


Figure 16: After each stimulus the screen was blank for 100 ms before the answer box and prompt were displayed.

et al., 2001). Additional support for the development of figure-ground involving extrastriate areas of the visual system comes from Zhou, Friedman, and Heydt (2000) who found that V1, V2 and V4 all show border ownership starting at about 100 ms.

Without masking, cortical activity initiated in the 100 ms presentation can continue uninterrupted and top down cortical influences can proceed despite the short exposure to the stimulus. This suggests that we will continue to see delayed recognition even with a brief stimulus exposure, a presentation which is much shorter than anyone has used with this paradigm. Experiment 6b, in contrast to 6a, used masking to block recurrent processing and thus we expected it to eliminate or reduce delayed recognition even with low-pass filtered images.

Method

Participants

Twenty undergraduate students who received course credit participated in the experiment.

Stimuli

The objects chosen for Experiment 6a were a random sample of the those used for Experiment 4 taking half from each category of single-simple, multiple-simple, single-complex and multiple-complex. The degrees of degradation (i.e., blurring) were adjusted based on the data from Experiment 4 to try and more closely approximate recognition levels of 20% and 80% for the most blurred and least blurred stimuli of each object series. In addition, because presentation time would be only 100 ms, all kernel sizes were reduced by 20% to reduce the degree of degradation for all stimuli.

Procedure

The procedure and apparatus for Experiment 6a was the same as for Experiments 4 and 5 with the following differences: each stimulus was preceded by a 300 ms fixation cross and each stimulus was presented for 100 ms rather than 3,000 ms (see Figure 16). In addition, there was a 100 ms interval after the stimulus was removed before the answer field was displayed whereas the answer field appeared immediately following the stimulus in Experiments 4 and 5.

Results

Figure 17 shows the mean percentage of correct responses averaged across participants. The average percent correct on first exposures for the four conditions was 14%, 32%, 51% and 63%. These recognition rates are lower than the rates in Experiment 4 (19%, 32%, 54%, 73%) but still a sufficient range to reveal delayed recognition.

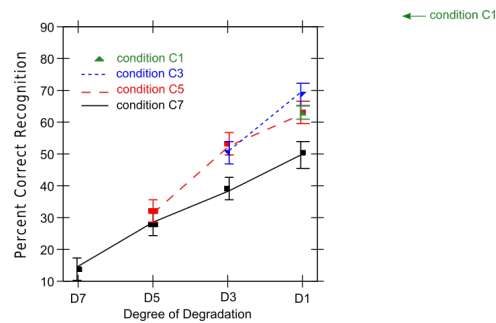


Figure 17: Mean percent correct responses, Experiment 6a. Each plot line represents a different within-subject condition with participants seeing 8 different objects in each condition. In the C1 condition the participant saw the object only in its least degraded form. In condition C7 the participant saw the object 7 times from more to less degraded, guessing after each stimulus in the series. Error bars represent the standard error for each data point.

A comparison of a linear mixed-model with fixed effects of condition and degree of degradation and a model without condition revealed a statistically significant effect as indicated by the likelihood ratio test — effect of condition: $\chi^2(1) = 13.44, p < .001$.

Discussion

Reducing the presentation time to 100 ms did not diminish the interference created by viewing more degraded stimuli. A presentation of 100 ms is considerably less than the 3 second viewing time of Experiment 4 or the viewing times of the Bruner and Potter (1964) experiment where the average time in the shortest condition was 13 seconds and in the longest condition 122 seconds, and again shows that the phenomenon of delayed recognition is quite robust when images are low-pass filtered. Would we still see interference if the image was masked after the 100 ms presentation?

Experiment 6b: Recognition of low-pass filtered images, 100 ms masked presentation

Experiment 6a showed delayed recognition with a 100 ms unmasked presentation. Figure-ground formation, as reviewed earlier, takes more than 100 ms. Experiment 6b sought to see if knocking out early top-down neural activity using a visual post-stimulus masking paradigm would eliminate the delay in recognition we see with the unmasked presentation. Backward masking has been suggested to selectively disrupt feedback (see Di Lollo, Enns, & Rensink, 2000; Lamme & Roelfsema, 2000, for reviews) and has been used in a variety of experiments to dissociate the effects of feedforward and recurrent processing (e.g., Lamme, Zipsler, & Spekreijse, 2002; Serre, Oliva, & Poggio, 2007). In Experiment 6b a mask with stimulus onset asynchrony of 100 ms was used to block the influence of past experience and prediction based on previously seen images and thereby eliminate delayed recognition due to previous views of more degraded versions of the object.

Method

Participants

Twenty undergraduate students who received course credit participated in the experiment. The participants were different than those in Experiment 6a.

Stimuli

The stimuli were the same as those used in Experiment 6a.

Sixteen masks were created by using pieces of the original undegraded images. Each mask was 400 by 400 pixels, as were the stimuli, and constructed in 2 steps, first by taking the center 100 x 100 pixels from 16 random stimuli and combining those into a new 400 x 400 pixel image and then by dividing that image into 10 x 10 pixel blocks and randomly shuffling them. The mask used after each stimulus was chosen randomly from the set of 16 masks.

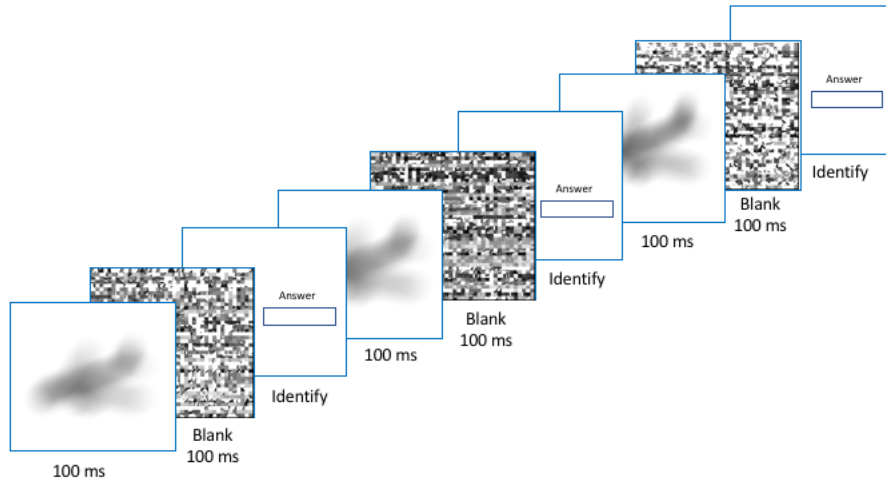


Figure 18: After each stimulus a mask appeared for 100 ms before the answer box and prompt were displayed.

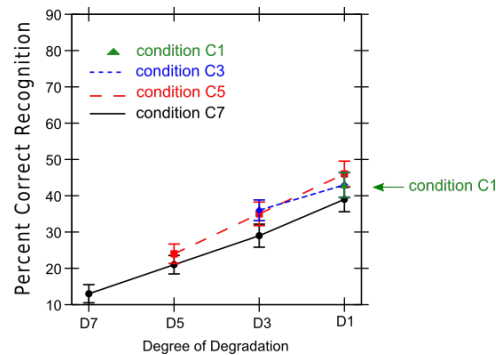


Figure 19: Mean percent correct responses, Experiment 6b. Each plot line represents a different within-subject condition with participants seeing 8 different objects in each condition. In the C1 condition the participant saw the object only in its least degraded form. In condition C7 the participant saw the object 7 times from more to less degraded, guessing after each stimulus in the series. Error bars represent the standard error for each data point.

Procedure

The procedure and apparatus for Experiment 6b were the same as for Experiment 6a with the following difference: the blank interval of 100 ms between the stimulus and the answer field was replaced with a mask (see Figure 18).

Results

The average percent correct on first exposures for the four conditions were 13%, 24%, 35% and 43% as shown in Figure 19. These recognition rates are all lower than those for the unmasked condition. Using the same linear mixed-model analysis used in Experiment 6a we found no effect for condition. The likelihood ratio test was not statistically significant $\chi^2(1) = 1.9569, p = .16$. A between experiment analysis using difference scores to test for a statistically significant effect of mask vs no mask did not show a statistically significant difference, $\chi^2(1) = 0.9492, p = .33$, despite the results seen when analyzing each experiment on its own.

Discussion

An error in grouping or figure-ground perception would not be a problem on later stimulus views if these processes were not part of bi-directional cortical processing but as referenced earlier these aspects of vision are indeed influenced by higher level cortical areas (Zipser et al., 1996; Supèr et al., 2001). We believe that it was the blocking of the top

down cortical signal via masking that is responsible for the lack of delayed recognition in Experiment 6b, though this result would be more convincing if the between experiment analysis had shown a statistically significant difference of mask vs. no mask and the absolute recognition rates had spanned an equal range.

General Discussion

In two series of seven total experiments we explored the phenomenon of delayed recognition of degraded visual images and demonstrated that the “interference” created by the more degraded images in the series can be diminished or eliminated by the manner of degradation. Specifically showing that degradation by low-pass filtering (Experiments 1, 3, 4 and 6) consistently causes delay and that degradation by fragmentation (Experiment 5) does not cause delayed recognition. We also found that another type of degradation, blocking (Experiment 2), which like fragmentation reduces continuity and generates images of less good form than low-pass filtering, does not fully eliminate but does decrease the strong recognition interference we see with the blurred image sequences. Further, we found that even with presentations of only 100 ms, images that are low-pass filtered will lead to delayed recognition and that masking these presentations reduces the effect of seeing the more degraded stimuli.

All of these results are consistent with the hypothesis that errors in the basic organization of the image elements leads to errors on subsequent viewing of a less degraded version of the image through top-down cortical influence. The errors are fostered when the degradation is done by low-pass filtering creating an image of good form but one which is ambiguous, an image that can lead to errors in figure-ground, grouping or other error types. Images that are degraded by fragmentation or in another manner that does not lead to good form are less likely to lead to strong top-down influence of figure-ground or grouping and thus is less likely to lead to recognition interference. Snodgrass and Hirshman (1991) put the cause of delay on persisting activations but neural activations do not last very long. O’Herron and von der Heydt (2009) found that border ownership neurons in area V2 could last a second or more when the border stimulus was ambiguous, but our experiments show that delayed recognition occurs with much longer interstimulus intervals. Experiment 3 demonstrates delayed recognition with an intertrial interval of 105 seconds, consistent with a hypothesis of influence from higher level cortical areas reconstituting basic grouping and figure-ground organization in lower level visual areas.

Though we believe we have provided strong evidence for the hypothesis that organizational errors are the cause of delayed recognition an experiment that was able to manipulate the absolute degree of organizational error would provide more direct evidence. We attempted in Experiment 4 to induce grouping errors by manipulating the number of objects in the image and though this made recognition more difficult it did not increase or decrease the degree of recognition delay. We also used a set of stimuli that varied in “complexity”. The more complex stimuli were easier to recognize than the less complex but complexity, like multiplicity, did not alter the degree of recognition delay. A more direct test of organizational errors leading to delayed recognition will require study of the induction of such errors.

Putting our results together with those of Bruner and Potter (1964), Wyatt and Campbell (1951) and Galloway (1946), when you remove high frequency spatial information you will see delayed recognition across various presentation modes, at least when images are projected, printed or seen on a computer display. We believe that this phenomenon also occurs outside the laboratory. If delayed recognition does occur outside the laboratory, as it might under low light levels, fog and other conditions that remove high frequency information, understanding its cause could have practical implications.

Another Possible Explanation: Information Integration

We have demonstrated that trying to identify a severely fragmented object as the fragmentation is gradually reduced does not hinder recognition whereas you are hindered by earlier exposures when the images are low-pass filtered. We speculate that this negative effect is due to persistence of organizational errors in the initial view but we might also consider that the difference lies in the integration of information from view to view. With a sequence of low-pass filtered images the changes from one stimulus to the next are small and wide spread, every part of the image is slightly different from view to view. With fragmentation the changes from stimulus to stimulus are restricted to significant changes to small portions of the image. Perhaps it is this difference that accounts for or contributes to the different results we observed with low-pass filtered images and fragmented images. The validity of this alternate explanation might be tested by modifying the fragmentation procedure to spread out the new additional information across the image, adding a bit to each of the visible portions rather than adding a few large blocks of new information on successive stimuli.

Conclusion

A series of seven experiments support a hypothesis that delayed recognition is fostered by low spatial frequencies which more than high spatial frequencies lead to perceptual errors that lead the observer down the garden path to misidentification. We compared two alternative types of degradation to low-pass filtering, both of which we predicted would reduce the likelihood of organizational errors, and found that one significantly reduced delayed recognition and the other eliminated it. The results also demonstrate that low-pass filtering creates a significant and long-lasting delay in recognition when the normal top-down visual processing mechanisms are allowed to proceed. Degradation by fragmentation produced no significant delay in recognition. We attribute the cause of the delay in recognition to perceptual errors and the normal top-down visual processes that are normally so helpful in maintaining a stable and predictable world, but which likely work against us when images are blurred or circumstances reduce high frequency spatial information.

*

References

- Bar, M. (2009). Predictions: A universal principle in the operation of the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364.
- Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology*, 23(3), 393–419.
- Bruner, J., & Potter, M. (1964). Interference in visual recognition. *Science*, 144(3617), 424–425.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.
- Davison, G. C. (1964). The negative effects of early exposure to suboptimal visual stimuli. *Journal of personality*, 32(2), 278–295.
- Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology*, 129(4), 481–507.
- Eger, E., Henson, R. N., Driver, J., & Dolan, R. J. (2007). Mechanisms of top-down facilitation in perception of visual objects studied by fMRI. *Cerebral Cortex*, 17(9), 2123–2133.
- Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, 20(5), 1245–1253.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4), 1149–1160.
- Fenske, M., Aminoff, E., Gronau, N., & Bar, M. (2006). Top-down facilitation of visual object recognition: Object-based and context-based contributions. *Progress in Brain Research*, 155, 3–21.
- Galloway, D. W. (1946). An experimental investigation of structural lag in perception. *American Psychologist*, 1, 450.
- Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, 54(5), 677–696.
- Kimchi, R. (1998). Uniform connectedness and grouping in the perceptual organization of hierarchical patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4), 1105.
- Kimchi, R., & Hadad, B.-S. (2002). Influence of past experience on perceptual grouping. *Psychological Science*, 13(1), 41–47.
- Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, 1.
- Lagasse, L. L. (1993). Effects of good form and spatial frequency on global precedence. *Perception & Psychophysics*, 53(1), 89–105.

- Lamme, V., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience*, *14*(7), 1044–1053.
- Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *The Journal of Neuroscience*, *15*, 1605.
- Lamme, V. A., Rodriguez-Rodriguez, V., & Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex (New York, N.Y.: 1991)*, *9*(4), 406–413.
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579.
- Luo, C. R., & Snodgrass, J. G. (1994). Competitive activation model of perceptual interference in picture and word identification. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(1), 50–60.
- O’Herron, P., & von der Heydt, R. (2009). Short-Term Memory for Figure-Ground Organization in the Visual Cortex. *Neuron*, *61*(5), 801–809.
- Palmer, S. E. (1980). What makes triangles point: Local and global effects in configurations of ambiguous triangles. *Cognitive Psychology*, *12*(3), 285–305.
- Peterson, M. A., & Gibson, B. S. (1994). Must figure-ground organization precede object recognition? An assumption in peril. *Psychological Science*, *5*(5), 253–259.
- Poort, J., Raudies, F., Wannig, A., Lamme, V. A. F., Neumann, H., & Roelfsema, P. R. (2012). The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron*, *75*(1), 143–156.
- Razpurker-Apfeld, I., & Kimchi, R. (2007). The time course of perceptual grouping: The role of segregation and shape formation. *Attention, Perception, & Psychophysics*, *69*(5), 732–743.
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart’s object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217–236.
- Schulkind, M. D. (2000). Perceptual interference decays over short unfilled intervals. *Memory & Cognition*, *28*(6), 949–956.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(15), 6424–6429.
- Snodgrass, J. G., & Hirshman, E. (1991). Theoretical explorations of the Bruner-Potter 1964 interference effect. *Journal of Memory and Language*, *30*(3), 273–293.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174.
- Supèr, H., Spekreijse, H., & Lamme, V. A. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience*, *4*(3), 304–310.
- Wyatt, D. F., & Campbell, D. T. (1951). On the liability of stereotype or hypothesis. *The Journal of Abnormal and Social Psychology*, *46*(4), 496.
- Zhang, N. R., & Heydt, R. (2010). Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex. *The Journal of Neuroscience*, *30*.
- Zhou, H., Friedman, H., & Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience*, *20*(17), 6594–6611.
- Zipser, K., Lamme, V. A. F., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience*, *16*, 7376–7389.