Visual Cognition, 2013
http://dx.doi.org/10.1080/13506285.2013.865694

Routledge
Taylor & Francis Group

Quantifying the role of context in visual object recognition

Elan Barenholtz

Department of Psychology/Center for Complex Systems, Florida Atlantic University, Boca Raton, FL, USA

(*Received 13 October 2012; accepted 10 November 2013*)

An object's context may serve as a source of information for recognition when the object's image is degraded. The current study aimed to quantify this source of information. Stimuli were photographs of objects divided into quantized blocks. Participants decreased block size (increasing resolution) until identification. Critical resolution was compared across three conditions: (1) when the picture of the target object was shown in isolation, (2) in the object's contextual setting where that context was unfamiliar to the participant, and (3) where that context was familiar to the participant. A second experiment assessed the role of object familiarity *without* context. Results showed a profound effect of context: Participants identified objects in familiar contexts with minimal resolution. Unfamiliar contexts required higher-resolution images, but much less so than those without context. Experiment 2 found a much smaller effect of familiarity without context, suggesting that recognition in familiar contexts is primarily based on object-location memory.

Keywords: Object recognition; Context; Scenes; Objects; Contextual facilitation.

Many theoretical and experimental approaches to object recognition have considered conditions in which the target object appears in isolation. In the natural world however, objects typically appear within a rich and complex contextual scene. Context may present an additional challenge to recognition, because objects must be selected and segmented from the background. However, contextual information may also provide significant benefits for performing

Please address all correspondence to Elan Barenholtz, 777 Glades Road, Boca Raton, FL 33431, USA. E-mail: elan.barenholtz@fau.edu

The author is grateful to Howard Hock, the editor, and three anonymous reviewers for their helpful comments on the manuscript. Portions of this work were presented at the annual meeting of the Vision Sciences Society in Naples, Florida 2009 and 2010.

This research was sponsored by an NSF Award to Elan Barenholtz [grant no. #BCS-0958615].

recognition since the scene in which an object appears carries statistical information about the probable identity of the object.

Numerous studies have reported more accurate detection or faster/more accurate naming when a briefly presented target object appears within, or is preceded by, a coherent and semantically consistent contextual scene. Virtually all of the previous research on contextual facilitation of object recognition has considered performance under conditions in which the target object is fully recognizable without the context. Under such conditions, the potential role of context is in allowing for more rapid recognition under constrained viewing times. For example, a number of studies have used a detection paradigm, in which the participant must determine whether a previously specified target was present or absent in a stimulus. Under these conditions, several studies have found that detection is more accurate when the object and scene type are consistent (Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce, Pollatsek, & Rayner, 1989). However, Hollingworth and Henderson (1998) reported that this detection advantage disappeared after appropriately correcting for response bias and proposed that previously observed advantages may have been due to guessing. A different paradigm that has been employed consists of subjects freely naming briefly presented objects. Under these conditions, several studies have found a facilitory effect of semantically appropriate contexts (Auckland, Cave, & Donnelly, 2007; Davenport, 2007; Davenport & Potter, 2004; Palmer, 1975). Several other studies have used a methodology in which the stimulus is presented until identification, with reaction time as the dependent measure. Here too, several studies have found an advantage (i.e., lower RTs) when the presented object is shown in conjunction with, or preceded by, semantically consistent objects or scenes (Ganis & Kutas, 2003; Gronau, Neta, & Bar, 2008). Finally, several recent studies have investigated the time course of contextual influence on object recognition. Here, participants performed a speeded categorization task in which they had to determine whether a rapidly presented stimulus was a member of a given category ("animal") or not with the finding that identification was faster and more accurate when the target object was shown in a congruent contextual scene (Joubert, Fize, Rousselet, & Fabre-Thorpe, 2008; Sun, Simon-Dack, Gordon, & Teder, 2011).

Despite the preponderance of data suggesting some form of contextual facilitation of object recognition, the precise nature and extent of such facilitation remains controversial (see Henderson & Hollingworth, 1999, for an extensive discussion of these issues). One possibility is that a schema activated by the scene facilitates processing specifically appropriate to the target stimulus, making the recognition process more efficient and, therefore, taking less time (Biederman, 1981; Biederman et al., 1982; Boyce & Pollatsek, 1989; Palmer, 1975). Another possibility is that perceptual analysis is identical regardless of context, but that the presentation of a context leads to a reduced criterion of feature matching between the target stimulus and some representation stored in

memory (Friedman, 1979). Finally, some have argued that context and object recognition are "functionally isolated", and that participants in many of these studies were simply *guessing* the presence of a consistent versus inconsistent target, based on prior knowledge of the occurrence of specific objects of the scene (Henderson & Hollingworth, 1999; Hollingworth & Henderson, 1998).

As noted, in all of the previous studies, the target stimuli could be readily identified in isolation—that is, even without the context. Thus, any potential role of context in identifying the target object may only be indirectly observed, based on speeding up the recognition process. Indeed, a number of these studies used detection paradigms in which the performance measure (*d*') explicitly aims to eliminate the possibility of the context providing task-relevant information, which would be factored out as response bias (Biederman et al., 1982; Hollingworth & Henderson, 1998). However, a more direct potential role for contextual facilitation exists when the image of the object itself is insufficient to uniquely identify it. For example, in Figure 1, the blurred images shown on the left are highly ambiguous when viewed in isolation (left), but can be readily recognized when shown in their original context (circled, right). This facilitation is clearly due to the fact that the object's location, combined with the contextual scene, provides information about its likely identity, which can compensate for the lack of information in the image of the object itself. However, the location is not in itself sufficient—that is, one can't simply "guess", based on contextual information alone; there are many other objects that are consistent with lying on an office desktop. Instead, the top-down contextual information appears to be combined with the visible bottom-up features of the target image in order to

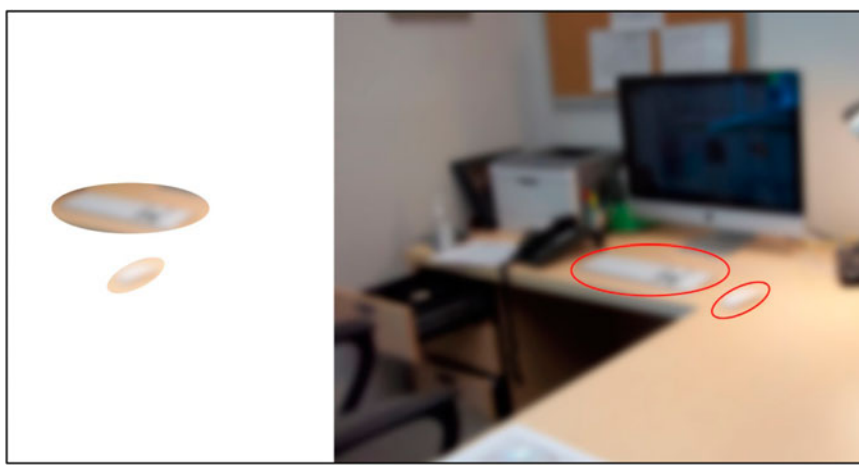

Figure 1. The blurred objects on the left are highly ambiguous and difficult to recognize when viewed in isolation but can be readily identified in context (circled, right). To view this figure in colour, please see the online issue of the Journal.

uniquely specify the object's identities. With regard to the previously described theories of contextual facilitation, this phenomenon appears to be best described as a shift in criterion at which a set of visible features may be matched to a stored representation.

The potential of context as a means of disambiguation is not limited to cases in which the image is artificially degraded; context could potentially play a role in many everyday situations that produce poor resolution, such as low lighting, occlusion, and, perhaps most ubiquitously, peripheral viewing. For example, despite poor resolution, people are able to recognize objects in their far periphery, an ability that is easy to demonstrate by trying to recognize objects in your periphery in a familiar setting such as your office. In recent years, the potential role of context as a source of information in the recognition of degraded or otherwise ambiguous stimuli has received a considerable amount of attention within neuroscience (Cox, Meyers, & Sinha, 2004) and in computer vision (Torralba, Murphy, Freeman, & Rubin, 2003). However, only a single study, by Bar and Ullman (1996), has examined the role of contextual information in the identification of degraded/ambiguous stimuli from a behavioural standpoint. In this study, participants tried to identify segmented portions of highly stylized line drawings. They found a large advantage when the segments were shown in the appropriate spatial relations to one another. However, the pictures used in these experiments were highly unrealistic, consisting of stylized black and white line drawings. Thus, it is difficult to assess to what extent these facilitation effects may extend to naturalistic conditions. In addition, although the target objects were "degraded" in some sense, based on their stylized composition, there is no obvious parallel to the kinds of degradation, such as blurring, that occur under natural conditions.

Thus, to date, no attempt has been made to measure whether, and to what degree, context can facilitate *natural* object recognition, by providing information relevant to the object's identity. This is a significant omission, for a number of reasons. First, as noted earlier, there remains significant debate as to whether object and scene processing and recognition operate independently of one another or whether they interact (Henderson & Hollingworth, 1999). However, since previous research has only considered cases in which context isn't actually *needed* to identify the target object, it has essentially ignored the domain where interaction is most likely to occur—that is, when both sources of information are needed to successfully identify the object. Second, previous theoretical approaches to recognition, such as the influential theories of Marr (1982) and Biederman (1987) have taken as their goal the recognition of objects independent of context. As such, the program of these theories has been to attempt to establish bottom-up features that are general enough to specify individual object labels/categories under highly variable viewing conditions—such as changes in lighting and viewpoint—independent of any other sources of information. Indeed, the assumption that extracting such features is a basic goal of visual

processing has served as a guiding principle in considering lower level visual processes, such as edge detection and figure–ground assignment. However, examples such as those in Figure 1 suggest that, under appropriate contextual conditions, extracting a set of features that can support recognition on their own may not be necessary, given appropriate contextual constraints. Since recognition almost always takes place in contextual settings, such theories may be limited in characterizing how recognition typically takes place in the real world.

In order to assess the potential implications of contextual facilitation vis-à-vis empirical and theoretical approaches to object recognition, it is first necessary to determine how much information concerning object identity is actually available in the contextual scenes in which objects are typically found. Although it is certainly possible to hand select or artificially construct cases in which context provides critical information (e.g., Bar & Ullman, 1996), it remains an open question as to how much information context actually carries in natural, real-world settings. The current study aimed to address this by quantitatively assessing the contribution of context in identifying objects in photographs of real-world living environments. In an attempt to capture the natural statistics of objects and their contextual settings, these environments were photographed "as is" and target objects were selected with very few constraints from the photographs, without consideration of their "fit" within the contextual scene. Then, "quantized" images of these target objects were presented to participants, whose task was to resolve them until they were able to successfully recognize them (Figure 2). The size of the quantized blocks making up the image determined its "resolution", with larger blocks yielding reduced resolution and the minimum resolution at which participants could identify the target object served as a dependent measure. Using this metric, the present experiment compared performance when the object had to be identified in isolation versus

Figure 2. Sample stimuli from the three experimental conditions. Participants sequentially "resolved" the target image by decreasing the size of the quantized blocks. In the two context conditions—unfamiliar and familiar—the target image (pointed to by a pink arrow in the top leftmost square) appeared within its contextual scene, whereas in the no-context condition the same target image appeared in isolation. To view this figure in colour, please see the online issue of the Journal.

when it had to be identified embedded within its natural context. The difference between these may be taken as a measure of the informational content of context that is available object recognition.

Previous studies of contextual effects on object recognition have only considered "schema"-level contextual information, based on generally shared expectations concerning the presence of specific object types in certain categories of scenes (e.g., chickens are expected in farm scenes, not city scenes). However, people also can learn specific information about the locations of objects in particular scenes based on personal observation. Several studies by Hollingworth (2005, 2006, 2007) found that people could retain highly specific information about the location and appearance of objects within viewed scenes. Since people often spend a large portion of their waking hours in the same, highly familiar environments (e.g., home, office), such individual-level knowledge could play a large role in typical recognition. Indeed, it is possible that people in familiar environments can rely primarily, if not completely, on memory when identifying certain objects.

In order to address these potentially different sources of contextual information, I also compared recognition performance when the target object appeared within a familiar context (i.e., stimulus pictures were obtained from the participant's own home) versus an unfamiliar context (i.e., an indoor environment with which the participant was not personally familiar).

The same set of objects was used as stimuli across the three experimental conditions (familiar context, unfamiliar context, and no context). Thus, any differences in performance could be attributed to the difference in contextual condition itself. Because the target objects were chosen with few constraints on their selection (see Methods section for a description of the selection criteria), the stimuli consisted of a wide array of objects. Thus, it was likely that some objects would be judged as "belonging" in their contexts to a greater extent than others, such as a toaster in a kitchen versus a toaster in a bedroom. As noted, the same set of objects was used across all three experimental conditions. Thus, this potential interobject variability was not confounded with the basic experimental manipulation. However, it is likely to add considerable variability to performance overall and also may affect some conditions more than others. In order to capture the impact of such interobject variability on performance, each object used as a target was rated along several dimensions that were likely to affect performance. These included three measures of "goodness-of-fit" (GOF) between the object and the context in which it appeared, which essentially aimed at assessing to what extent the object and the context "belonged together":

1. *Consistency*: This rating was intended to assess the consistency of the target object with the context in which is appeared, i.e., given a certain category of object, is this the kind of scene you would expect it to appear in?

2. *Position*: This rating was intended to assess the consistency of the target object's specific position within its surrounding scene, i.e., given a certain kind of object and scene, to what extent is this the specific location within the scene where you would expect to find it?
3. *Frequency*: This rating was intended to measure how frequently the target object would be encountered within the context in which it appeared (Note: this is not the same as consistency, since an object can be highly consistent with a scene but not very frequently encountered. For example, a juice maker is highly consistent with a kitchen but may not be rated as very frequent.)

In addition to these goodness-of-fit ratings, several additional factors that might potentially affect performance were considered. These included the absolute size of the target image, which varied widely across target stimuli based on the physical size and distance of the object in the photograph. Although the dependent measure of block size was independent of the overall size of the image, larger stimuli contained more blocks, and thus more visual information, which could potentially make them easier to identify. In addition, the percentage of the rectangular target image occupied by target object was measured. This latter variable varied widely depending on the shape and convexity/concavity of the target object (for example, a thin object with a wide base, such as a lamp, takes up a much smaller percentage of the target image than a rectangular object such as a refrigerator). Objects making up a lower percentage of the image require more segmentation in order to isolate the target object, which might require higher resolution.

The general hypothesis of the current study was that the more contextual information participants had available to them, the lower the resolution they would need in order to identify the target object. Thus, in the familiar-context condition—where participants likely have highly specific knowledge about the locations of objects within their own homes—participants should need the least visual information to recognize the target stimuli. It is important to note that, in the studies of object/scene memory mentioned earlier, participants explicitly learned the location of objects within the scene as part of the experimental task; however, the current study gauged participant's memory of environments that were encountered during everyday activity. Thus, the level of precision and detail of those memories was unknown. Performance in the familiar-context condition is likely to be followed by performance in the unfamiliar-context condition, in which participants have access to general information regarding the probability of specific objects appearing in specific contextual settings and locations, without precise details. Finally, the worst performance was predicted in the no-context condition. If context indeed reduces the amount of information needed from the object itself, this implies that the information is instead being provided by the context. Thus, the reduction in resolution at which the target

objects are identified can serve to quantify the informational contribution of context in object recognition.

A secondary set of hypotheses concerned the ratings. First, we may expect that participants in the Unfamiliar-context condition will show better performance for objects that had greater GOF, since this may reduce the criterion at which the target object can be considered a match to the stored representation in memory. The effect of GOF is likely to be muted or absent in the familiar-context condition, however, since these participants are likely to depend, at least in part, on their individual memories and thus should be less sensitive to factors that depend on schema-level expectations. However, it is possible that familiar participants will be affected by GOF as well. For example, they could use a hybrid strategy that depends on both individual memory as well as schema-based strategies, particularly in cases where they do not have precise memory of the object in a given location. In addition, GOF between an object and its location could, theoretically, facilitate memory of the object as well. Finally, no impact of GOF was expected in the no-context condition, since the context was not available to participants. However, other factors, such as image size and percentage of the image containing the object are likely to be important in the no-context condition, perhaps more so than in the contextual conditions.

EXPERIMENT 1

Methods

Participants. Fifty-one Florida Atlantic University undergraduate students participated for course credit. Participants were divided up into three groups. The unfamiliar-context group had 23 participants (12 male), as did the no-context group (nine male). The familiar-context group, from whose homes the photographs used to generate the stimuli were derived, had five participants (three male). In addition, five additional undergraduate students (three male), participating in the lab as student researchers, performed the ratings task. Each of the participants only performed the experimental task on the stimuli from a single home. Thus, each home in the unfamiliar and no-context conditions had seven (42/6) participants assigned to it, and the familiar-context condition had one participant per home.

Stimuli. Stimuli consisted of modified photographs of indoor living spaces of single-family homes. The original photographs of each home were collected by a photographer who was not a participant in the experiment who was instructed to take pictures in each of the main living spaces of the house including kitchens, living-rooms/dens, studies, and bedrooms. Bathrooms and garages were not photographed. The procedure for photographing each room consisted of standing in the centre of the room and taking a rotational series of

nonoverlapping, vertically oriented photographs that covered the entire space of the room. The individual stimuli were then generated by selecting an object for inclusion as a target stimulus in the photographs. Only a single object was selected for inclusion from each photograph and only a subset of photographs were used to generate target stimuli. Overall, each home environment was used to create a total of 10–15 stimulus objects for a total of 72 target stimuli. Inclusion criteria for target objects were that the target object had to be fully or almost fully visible (not strongly occluded or cut off in the pictures), had to be an object with a familiar name or category, and had to be judged to be a permanent or semipermanent feature of the environment, rather than a transient object, such as disposable or food items.

Once the target objects had been identified, each target stimulus was created by first selecting the smallest rectangular region that contained the entire target object. Then, this rectangular image was used to generate the series of quantized target images by dividing the image into equally sized square blocks, and filling each block with the average red-green-blue (RGB) value of all of the pixels contained within that region of the original image (the rightmost and bottommost blocks were truncated as needed to fit into the rectangular region). This quantization method for assessing recognition performance is similar to the method first used by Harmon and colleagues (Harmon, 1973; Harmon & Julesz, 1973) to assess facial recognition capabilities and, more recently, Torralba (2009) for scene recognition. Each target image was used to generate a series of these quantized images, with blocks that ranged from the largest size (125 × 125 pixels per block) to the smallest (1 pixel per block, i.e., the original image). Each series consisted of 49 images in total, beginning with the lowest resolution and incrementing by five pixels per block dimension (i.e., 125 × 125; 120 × 120; 115 × 115 …) until the 30 pixels per block level; after that, the series continued in one pixel per block increments (i.e., 30 × 30; 29 × 29, etc.) until the original image (1 × 1). This led to a sequence in which the resolution of the image systematically increases as the size of the blocks decreases (Figure 2).

As shown in Figure 2, the stimuli in the two context conditions consisted of the quantized target image shown embedded within the context of the rest of the photograph from which it was taken, whereas in the no-context condition, the quantized target image was shown in isolation, in the centre of the screen. It is important to note that, because the quantized target image was generated from the smallest rectangular region containing the target object, it always contained some of the local background along with the target object. This method allowed presentation of the very same image both within its context and in near-isolation (i.e., in the no-context condition), without artificially segmenting the target object from its local background and providing information about its contour shape. This meant that participants performing the experimental task had to perform local perceptual organization processes—such as segmentation, figure–

ground, and perceptual grouping—in order to define the portions of the image that made up the target object, just as they would under natural viewing conditions. However, although this method did not completely isolate the target object, care was taken so that (1) the local context typically did not include any recognizable information that could serve as a meaningful context in the no-context condition and (2) it was very clear what constituted the target object in the image, based on its central position and the proportion of the total target image it occupied.

The images constituting the visual scenes (shown in the familiar- and unfamiliar-context conditions) measured ~12.5 degrees on each side, whereas the target-object stimuli (which were identical across all three conditions) varied in size from around one degree to four degrees in both width and height.

Procedure. The procedure was identical across all three conditions. Each participant was presented with all of the objects from a single home (between 13 and 15 objects, depending on the home) in random order. On each trial, participants were first presented with the lowest resolution target image (125 pixels per check) of the current object, either in isolation (in the no-context condition) or embedded within its contextual scene (in the two context conditions). Participants could then "resolve" the image by depressing a key on the keyboard that incremented the image to the next smaller check-size. Participants were instructed that there was always a single "primary object" within each image, which was fully or almost fully visible (this was necessary since portions of other objects and surfaces appeared within the stimulus image) and were instructed to identify this primary object. Participants had two chances to identify the object: once when they "thought" they knew what the object is and once when they were "sure" they knew what the object is. This method is related to that used by Bruner and Potter (1964), in which participants brought blurred images into focus until recognition. The quantization manipulation used here is similar, in principle, to blurring, but provides a simple metric by which the resolution (block size) and informational content (number of blocks) in the image can be measured. Unlike Bruner and Potter's study, in which the image sequence was presented at a fixed interval, in the current study participants moved through the sequence in a self-paced manner, allowing them to fully consider the contextual information in the image without any constraints. In addition, because participants had two chances to identify the object, they were encouraged to try to identify the object as early as possible. These two methodological techniques were used (rather than, for example, a single, time constrained, response) so that the maximal degree of contextual information available could be measured. This reflects the goal of the current study, which was to assess how much information is actually available in contextual scenes, rather than an attempt to assess how such information is accessed under

behaviourally constrained conditions. Participants entered their responses about the object's name/identity by typing their response (the text of their response appeared as they typed on the left side of the screen, separated from the stimulus image) and pressing the "Return" key. After their second response, a white screen replaced the image on the screen for a 2 s interval and then the next object-sequence began.

Analysis. Each response was scored based on size of the blocks making up the image that first produced a correct identification of the target object. For purposes of analysis, the size of the blocks was recorded as the number of pixels in one dimension of each square block, or pixels per dimension. This provides a simple measure of resolution that is independent of the size of the original image from which the quantized target images were created. The lowest presented resolution was 125 pixels per block, and the highest was one pixel per check, i.e., the original image. If the object was never correctly identified, a score of 0 was administered. Scoring of all responses was performed by independent raters who had not participated in the experiment and were blind to the experimental condition, based on a subjective determination that the label was both correct and specific enough to constitute a correct identification of the object. Each response was scored based on the largest block size at which the participant correctly identified it, either in the first or second response. Obvious mistakes (e.g., typos) were included in the analysis as long as the intended response of the participant was clear.

Each object was independently assessed by five raters along three GOF dimensions: consistency, position, and frequency. Raters viewed each object, shown circled within its surrounding context, in full resolution and chose a value from 1 to 7 (with 7 being the highest) for each dimension. Each rating was described to the rater with a question and an example. The consistency text was: "How consistent is this type of object with the scene in which it appears? For example, a toaster would be highly consistent with a kitchen scene but not a living room scene." The position text was "How typical is the position of this type of object within the scene in which it appears? For example, an alarm clock would typically be on a night table, but not on the floor." The frequency text was: "How frequently would you expect to see this type of object in the scene in which it appears? Note that this is not the same as consistency. For example, a Jacuzzi might be consistent with a bathroom scene but would have a lower *frequency* than a sink."

Each individual rating was scored in terms of its relative difference from the mean of that rater for that particular dimension, yielding a bias-independent, positive or negative "difference score". The average of the five rater's difference scores was then calculated, yielding the final ratings for each dimension of every object.

Each target object's was measured in terms of its overall size (measured as total pixels in the rectangular image), which ranged from 2301 pixels to 167,750 pixels, with an average of 35,413. In addition, each object's was measured in terms of the proportion of the rectangular image it occupied, which ranged from 14% to 94%, with an average of 63%.

Results

The percentage of trials on which participants correctly identified the target object—at any resolution—was calculated for each condition. In the no-context condition it was at 85% (*SD* = 10%). In the unfamiliar-context condition, accuracy was at 96% (*SD* = 6%), while in the familiar-context condition, participants identified the objects with 100% accuracy (*SD* = 0). A one-way ANOVA found that there was a significant effect of condition, $F(2, 87) = 14.456$, $p < .001$, partial $\eta^2 = .413$. Post hoc analysis using a Bonferroni corrected *t*-test found a significant difference between the no-context condition and each of the two context conditions (all *p*s < .001) but no significant difference between the two context conditions. As described in the Method section, participants had two chances to respond on each trial. Considering only trials in which they accurately identified the object, participants in the unfamiliar-context condition did so in their first response on 74% of trials versus 26% on the second trial. For participants in the no-context condition, the breakdown was 64% first response, 36% second response. In the familiar-context condition participants always identified the object correctly in their first response.

Figure 3 shows the mean block size at which participants correctly identified the object in each of the three contextual conditions, measured by the number of screen pixels making up a single dimension of each block in the quantized image. The average block size at which participants correctly identified the target object in the familiar condition was 96.6 pixels per dimension (*SD* = 15.23), 22.3 pixels per dimension in the unfamiliar-context condition (*SD* = 13.4) and 4.5 pixels per dimension in the no-context condition (*SD* = 6.5). A one-way ANOVA yielded a significant and large effect of contextual condition, $F(2, 87) = 175.5$, $p < .0001$, partial $\eta^2 = .881$. A one-way ANOVA yielded a highly significant effect of contextual condition, $F(2, 87) = 175.5$, $p < .0001$. Bonferroni post hoc analysis found highly significant differences between each of the three contextual conditions (*p*s < .001 in all comparisons). Another way to represent performance that amplifies these differences further is to consider the *number* of blocks making up the target image at which participants correctly identified the target object. This was calculated by taking the total number of pixels in the target image and dividing by the number of pixels in each of its quantized blocks (i.e., the square of the pixels-per-dimension measure used previously). This may

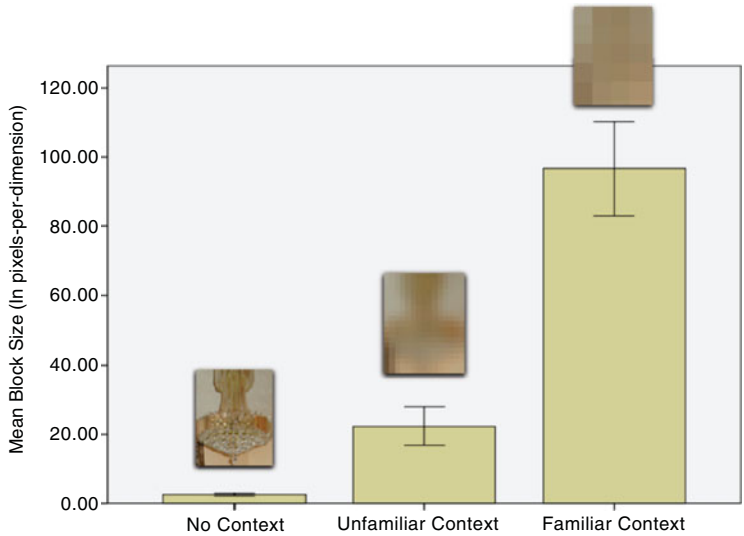

Figure 3. Performance across the three contextual conditions in Experiment 1. Each bar represents the average check size (in number of pixels per block dimension) at which participants in that condition successfully identified the target object. Bars represent one standard error the mean. The images above each of the bars shows a sample experimental stimulus at the level of resolution connoted by the bar. To view this figure in colour, please see the online issue of the Journal.

be thought of as a measure of the amount of visual "information" making up each quantized image, as it indicates the number of RGB values that would be needed to fully reconstruct the image. In the no-context condition, the median[1] image had 1045 blocks (rounded to the nearest integer); in the unfamiliar-context condition, it had 175 blocks per image; and in the familiar-context condition, it had 10 blocks per image.

The extremely low resolution in the familiar-context condition suggests that participants may have often been able to correctly identify the object without resolving it at all. This was tested by measuring the percentage of trials in which participants correctly identified the object at the lowest resolution level, without resolving it at all (125 pixels per block). In the familiar-context condition, participants did this on 71% of trials; participants in the unfamiliar-context condition did so on 4% of trials; and in the no-context condition this never occurred. Looking only at trials in which participants resolved the image before identifying it, familiar participants identified the target objects at an average resolution of 29.6 pixels per block (SD = 19), and unfamiliar participants did so

[1] The median was calculated, rather then the mean, because the number of blocks increases exponentially with pixels per dimension, making it highly vulnerable to the effect of outliers, such as trials in which respondents did not get the correct response.

at 22.36 pixels per block (SD = 13.39). The difference that was not significant by t-test ($p > .1$)

Ratings data. The pairwise correlations between each of the five rater's values for each of the three GOF measurements were assessed to measure interrater reliability. For consistency, all of the pairwise comparisons across raters were significantly positively correlated with one another, with an average r-value of .43. For frequency, 80% (16/20) of the pairwise comparisons were significantly positively correlated with one another with an average r-value of .43. For position, 70% (14/20) of the pairwise comparisons were significantly positively correlated with an average r-value of .35. As described earlier, a single average score for each dimension of every object was computed from the various raters. This was used to calculate correlations between the ratings of each dimension and performance, as measured by the block size at which each object objects was, on average, correctly identified. Block size showed a significant positive correlation with consistency, $r(71) = .320$, $p = .007$ and frequency, $r(71) = .347$, $p = .003$, and position reaching marginal significance, $r(71) = .241$, $p = .051$. In both the familiar-context and no-context conditions, block size was not significantly correlated with any of the ratings measures (all ps > .05).

The three ratings measures were themselves highly positively correlated, with pairwise r-values ranging between .79 and .91 (all ps < .001). The sum of the three rated dimensions were combined into a single measure of GOF for each object. Figure 4 shows performance as a function of GOF score in the unfamiliar and familiar context conditions. Block size was positively correlated with GOF in the unfamiliar-context condition, $r(71) = .321$, $p = .006$, and the familiar-context condition, $r(71) = .311$, $p = .019$.

Inspection of the familiar-context graph in Figure 4 shows that there were a higher proportion of objects recognized at the maximum block size (125) at the higher GOF levels, which may have been the sole basis of the positive correlations. To test this, a separate correlation was performed excluding these objects, which yielded no significant correlations between GOF and block size.

Although there is a reasonably strong correlation between GOF and performance in the unfamiliar-context condition, performance for many of the low-GOF objects in the unfamiliar-context condition still appear to be superior to performance in the no-context condition, where the mean block size was around five pixels in size. This was tested by selecting only objects that rated below the mean GOF score and assessing performance for those objects in the unfamiliar condition (M = 11.9, SD = 9) and the no-context condition (M = 2.8, SD = 0.42), a difference that was highly significant, $t(44) = 5.45$, $p < .0001$.

Performance was positively correlated with image size in both the unfamiliar-context condition, $r(71) = .13$, $p < .05$, and in the no-context condition, $r(71) = .348$, $p < .01$, but not in the familiar-context condition (all ps > .05). The

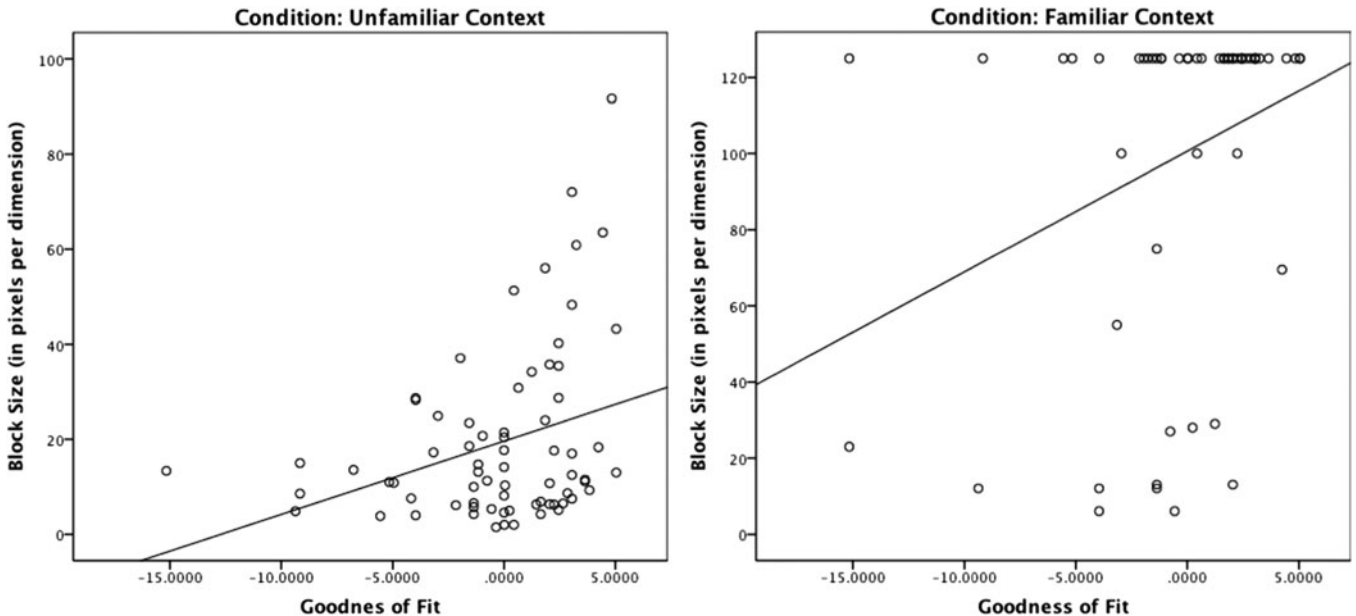

Figure 4. Scatterplots showing the relationship between the rated goodness-of-fit of each object and the average block size at which that object was accurately identified in the two context conditions. The lines represents the linear regression fit to the data.

percentage of the target image occupied by the target object was not significantly correlated with block size in any of the conditions (all *p*s > .05).

Discussion

The results of Experiment 1 demonstrate that context can play a powerful role in object recognition, providing information that allows objects to be accurately identified under greatly reduced resolution. On average, participants in the unfamiliar-context condition correctly identified the target objects at approximately one-quarter the resolution required by participants in the no-context condition; participants in the familiar-context condition correctly identified the target objects at approximately one-fifth the resolution required by participants in the unfamiliar-context condition. These results suggest that very different strategies were at work in each of the three conditions, based on the differential availability of contextual information. In short, context, and particularly a familiar context, can be a powerful source of direct information in object recognition.

Breaking down the results by condition, in the no-context condition, participants typically needed to resolve the image to almost its original resolution in order to successfully recognize it and could not identify it at all on almost 20% of trials. However, although performance was generally poor, it was positively correlated with image size. Thus, as visual information increased (with more blocks making up the larger images), participants did better, suggesting that they were successfully using whatever visual information was available to perform the task. To date, very few studies have assessed recognition performance of nonsegmented naturalistic images. The current results suggest that recognition of such images, in isolation of the broader context in which they appear, is very difficult, requiring near-optimal resolution to be performed successfully and sometimes failing even with high resolution. Thus, the current results may be taken to suggest that, under conditions of poor resolution, such as dim lighting or peripheral viewing, contextual factors likely play a very large role in object recognition. Of course, this does not suggest that people would perform similarly poorly if the target object were fully segmented from their backgrounds, as is the case in many previous studies of object recognition, in which objects are shown in isolation on a white background. But, in the real world, objects are not segmented from their background and the current results may be assumed to reflect these conditions more closely than under artificial segmentation.

In the unfamiliar-context condition, participants almost always had to resolve the image—usually to a fairly large degree—until they were able to correctly identify it but to a much lesser degree than in the no-context condition. Like the familiar-context participants, performance improved as the goodness of fit between the object and its context increased, suggesting that participants were accessing schema-based expectations in attempting to determine the object's

identity. However, although the contextual scene clearly provided facilitation, participants were not simply guessing the correct answer, using these schema-based expectations alone. They almost never chose the right answer at, or close to, the minimum resolution, as would be expected if they were simply guessing correctly. Instead, they waited until they were reasonably confident of the object's identity before responding, as indicated by the fact that they usually made the correct choice in their first response.

The correlation between GOF and performance suggests that participants were accessing their schematic knowledge in performing identification. But even low-GOF objects were identified, on average, at considerably lower resolution in the unfamiliar-context condition compared with the no-context condition. It should be noted that the ratings scale only represents relative, not absolute GOF. Thus, even objects with low GOF scores may "belong" in their contexts to the extent that the context provided some semantic information about their identity. Furthermore, there was a high degree of variability in the ratings themselves; correlations between raters, although significant, were still not very high. However, it is also the case that there are additional potential benefits to viewing an object in context that do not depend on schema-based expectations First, the context can provide information about the physical size of the object, as long as the target image can be localized in depth relative to other, recognizable objects. Second, as noted, the local surrounding portions of the image were included with the target image, which means participants had to segment the target object from its immediate surround, requiring several processes of perceptual organization such as figure–ground segmentation and perceptual grouping. Viewing the object within a larger context may make such processes easier by allowing the participant to identify portions of the target image that belong to other objects or by allowing background properties (such as texture) to be computed and then subtracting the background from the image (Dongxiang, Hong, & Ray, 2008). Relatedly, the context may allow segments of other, nontarget objects that are encroaching into the target image to be grouped with the other objects and thus identified as not part of the target (for example, in Figure 3 a section of the blinds, which are not part of the target object, encroach into the target image). However, it is important to note that the percentage of the target image occupied by the target object did *not* correlate significantly with performance. This suggests that difficulty in determining which part of the image "belonged" to the target object was not a dominant determinant of performance, and thus was unlikely to be a significant factor in the different performance in the no-context and contextual conditions.

In the familiar-context condition, participants correctly identified the object based on minimal resolution on 71% of trials. Across all trials, the median number of blocks needed to identify the image was only around 10 blocks, which contains little more than some of the coarse average colour distributions in the original image. This suggests that participants in this condition were often

not relying on the visual information in the target image at all and were instead identifying the object based solely on their memory of what object they had previously seen in that location. This result is consistent with previous studies showing that people can retain detailed information about the location and appearance of objects in viewed scenes (Hollingworth, 2005, 2006, 2007). However, unlike in these previous studies, in the current study, participants were never given instructions to actively memorize the locations of objects in the scenes but instead learned them through everyday interaction with their environment. Thus, these results suggest that people can and do build up rich representations of scenes they encounter on a frequent basis and that this information is available for the purposes of recognition.

However, although the performance in the familiar-context condition clearly depended on personal experience and memory, it is interesting to note that there were positive correlations between performance in the experimental task and the GOF between the object and its context, which measured general, schema-level expectations. Even more surprisingly, these positive correlations appear to be driven by the fact that participants were able to identify higher GOF objects at the minimal resolution. Thus, these results do not support the idea that familiar participants employ a hybrid strategy, relying on schemas when they are unable to remember the object based on the location. This is because the image quality is so poor at minimal resolution that a schema-based model is unlikely to provide the correct answer, a fact supported by the fact that participants in the unfamiliar condition only achieved recognition at minimal resolution on 4% of trials. Instead, this result suggests that familiar participants were better able to *remember* what object was in a location (which is based on their individual experience), when that object also happened to "belong" there (which is based on general schemas). This somewhat surprising result has several possible explanations. First, it could be that objects that conform to schema-based expectations are actually initially encoded in memory more successfully than those that do not. Alternatively, it could be that schema-based expectations generated a set of initial guesses about the likely object, which served as a memory cue that led to successful retrieval of the correct object.

Regardless of the explanatory mechanism, it is clear that performance in the familiar-context condition depended, to a large degree, on memory of the scenes and objects that served as stimuli. There are actually several different sources of memory that could potentially have facilitated performance. First, and most likely foremost, these participants may have remembered which specific object that they had seen in the specific location in the scene. However, this object-location associative memory is not the only potential source of facilitation available to familiar participants. Even without being given a specific location, people familiar with an environment may recall which types of objects appear *anywhere* within their homes. In addition, people are likely to remember the visual *appearance* of the specific objects that are in their home (i.e., their

specific visual features). Both of these types of information could facilitate performance for familiar objects, even without knowing a specific contextual setting (beyond "my home"). Because Experiment 1 only presented familiar participants with objects in context, it is impossible to determine from their performance whether, and to what extent, these more general types of familiarity may have facilitated them, along with their location-specific memory. Experiment 2 aimed to address this.

EXPERIMENT 2

Experiment 2 tested recognition performance under the same conditions as the no-context condition in Experiment 1 but where a subset of the target objects were photographs of objects from the participant's own homes. The experiment used a new set of participants and stimuli and employed a design in which each participant was tested on his/her own objects and those of the other participants. This allowed us to compare performance for both familiar and unfamiliar objects using a within-subjects design. Because no context was provided, no GOF measures were assessed in Experiment 2 as they were in Experiment 1. Instead, two different measures, which did not depend on contextual factors, were assessed: "global frequency" was intended to measure how often would one expect to encounter an object belonging to the category of the target, anywhere within a home. In addition, a rating of "typicality" was intended to measure how typical the particular target object was of its category.

Methods

Participants. Ten Florida Atlantic University students (four male), who did not take part in Experiment 1, participated in Experiment 2 for course credit. Three additional students (two male) served as raters of the experimental stimuli.

Stimuli and procedure. Stimuli and procedure were similar to those used in the no-context condition in Experiment 1 but using a new set of stimuli derived from the homes of participants in Experiment 2. In addition, Experiment 2 used a within-subjects design, in which each participant performed the recognition task on both his/her own photographed objects (familiar) as well as all of the objects derived from the homes of the other participants (unfamiliar). Participants performed the familiar and unfamiliar tasks in two, sequential experimental sessions. Five of the participants performed the familiar task first, while the other five performed the unfamiliar task first.

Analysis. The three raters scored each object for its global frequency and typicality on a scale of 1 for least frequent/typical to 7 for most frequent/typical.

As in Experiment 1, each individual rating was scored in terms of its relative difference from the mean of that rater for that particular dimension, yielding a bias-independent, positive or negative "difference score". The average of the three rater's difference scores was then calculated, yielding the final ratings for each dimension of every object.

Results and discussion

Figure 5 shows the mean block size at which participants, on average, correctly identified the object in the two conditions. A within-subjects *t*-test found that performance was significantly better in the familiar ($M = 5.5$, $SD = 0.9$) versus the unfamiliar condition ($M = 7.52$, $SD = 1.63$), $t(9) = 2.755$, $p = .025$, $d = .941$. Because participants in the familiar condition did not know where in their home an object had come from, the source of the advantage for the familiar objects was likely based on one, or both, of two factors: (1) their knowledge of which kinds of objects they have in their home and (2) familiarity with the specific appearance of those objects. If the familiarity advantage stemmed, at least in part, from knowledge of which classes of objects are present, then it might be specific to lower global-frequency objects whose low a priori likelihood may raise the criterion for correctly identifying them. Similarly, if the familiarity advantage stemmed, at least in part, from knowledge about the specific *appearance* of the object, then the familiarity advantage might be more

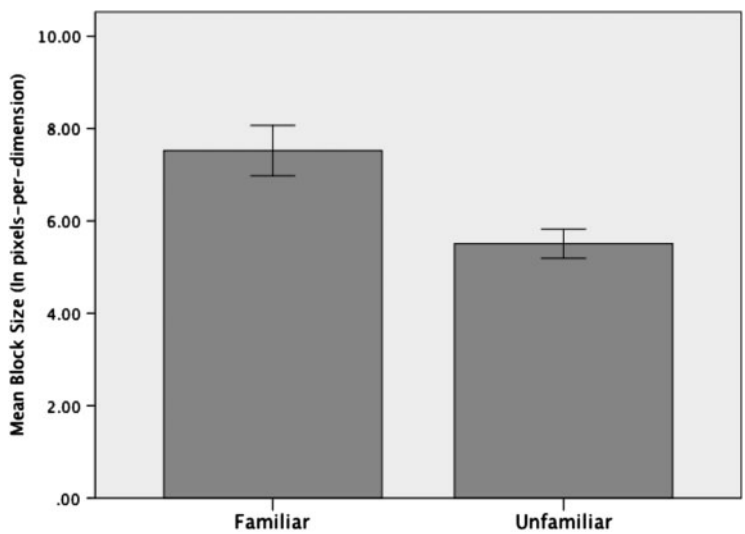

Figure 5. Performance in the two experimental conditions in Experiment 2. Each bar represents the average check size (in number of pixels per block dimension) at which participants in that condition successfully identified the target object. Bars represent one standard error of the mean.

pronounced for lower-typicality objects, whose features are less consistent with the correct object category, than for the higher typicality objects.

To first measure interrater reliability, the pairwise correlations between each of the three rater's values for each of the three GOF measurements were assessed. For the two ratings, global frequency and typicality, all of the pairwise comparisons across raters were significantly positively correlated with one another, with average r-values of .51 and .57, respectively; all p-values < .0001. Next, the objects were binned into upper and lower halves of rated global frequency and typicality; separate t-tests were then conducted on the high and low rated objects comparing familiar and unfamiliar performance. For the lower global-frequency objects, there was a significant difference between performance for the familiar objects (M = 7.16, SD = 1.62) and unfamiliar objects (M = 5.10, SD = 0.45) by within-subjects t-test, $t(9) = 4.56$, $p = .003$, $d = 3.35$. However, for the higher global-frequency subset of objects there was no significant difference between performance in the familiar condition (M = 6.77, SD = 1.9) versus the unfamiliar condition (M = 5.99, SD = 1.13) objects, $t(9) = .966$, $p > .1$. Similarly, for the lower-typicality subset of objects, there was significantly better performance in the familiar condition (M = 7.77, SD = 1.84) versus the unfamiliar condition (M = 5.53, SD = 1), by within-subjects t-test, $t(9) = 3.783$, $p = .007$. $d = 1.379$. However, for the higher-typicality subset of objects, there was no significant difference in performance in the familiar condition (M = 6.94, SD = 1.5) versus the unfamiliar condition (M = 5.79, SD = 0.87), $t(9) = 1.82$, $p > .1$.

The results of Experiment 2 demonstrate that there is a significant effect of familiarity in object recognition, even when the objects are shown without context. However, this advantage was limited to lower-frequency/typicality objects. These results suggest that the familiar advantage is likely due both to the participant's overall knowledge of which objects are within his/her house (which allowed participants to identify unusual, or low-frequency, objects from their own homes more successfully than other people's homes) as well as their knowledge of the particular appearance of those objects (which allowed them to identify objects with unusual appearances, or low-typicality, from their own homes). However, even though the familiar advantage in this experiment was statistically significant, performance was still dramatically worse than in the familiar context condition in Experiment 1. Although a direct comparison between the two experiments is not appropriate, because they used different sets of stimuli, the relative scale of the familiarity effects in the two experiments suggest that the brunt of the advantage in the familiar-context condition in Experiment 1 was not due to overall familiarity with objects in the environment or their appearance. Instead it was based on memory of the objects based on their specific locations. Put another way, it seems that knowing that an object comes from a familiar environment, without knowing where in the that environment it comes from, does not represent a specific enough source of contextual information to dramatically reduce the criterion at which objects are identified.

GENERAL DISCUSSION

The results of the two experiments reported here demonstrate the profound role that context can play in visual object recognition, serving to dramatically reduce the amount of bottom-up information needed to identify an object. This is the first study to measure the information content of context in nonmanipulated, real-world settings, and the current finding suggests that contexts can provide a large portion of the information needed for object recognition. It should be noted that these results likely reflect an upper bound of contextual facilitation, since participants had unlimited time to consider the objects and their contextual surrounding, a condition that may not hold under typical viewing conditions. As stated, the goal of the current study was to measure the degree of information available within contexts, not to assess how efficiently such information is processed. To that end, these results make clear that contexts do carry a high degree of information that is available for the purposes of object recognition. It stands to reason that the visual system takes advantage of this information when possible. However, future studies will be needed to assess how efficiently this information is utilized under time-constrained conditions of the sort used in earlier research of contextual facilitation.

The observed contextual facilitation in the unfamiliar condition was driven, at least in part, by schema-based expectations, as evidenced by the modulation of performance by the GOF between objects and their context. However, context still facilitated performance even for low GOF objects, which may suggest a contribution of nonsemantic information contained within the context, such as the target object's size or figure–ground cues. In the familiar condition, the very poor resolution at which participants were usually able to identify the object suggests that their performance was largely driven by their individual memory of specific object–location associations. This is the first study to demonstrate that people retain sufficiently detailed memories of their indoor living environments to be able to identify objects based almost exclusively on their locations within the environment. However, even for familiar participants, recognition wasn't entirely memory based and instead interacted with schema-based expectations, as evidenced by the fact that there was a reduced but significant effect of GOF. Finally, Experiment 2 demonstrated that these benefits of familiarity extend beyond location-based memory to more general knowledge/memory of what objects are present within a broader environment (in this case, the individual participant's home). These benefits specifically applied to objects that were less common or had an unusual appearance and were considerably smaller than those due to contextual location in Experiment 1.

The current results are consistent with previous studies showing contextual facilitation of object recognition. In particular, the results of the unfamiliar-context condition in Experiment 1 provide support for "criterion-modulation"

models of facilitation, in which context reduces the amount of bottom-up information required to trigger a match to a stored representation by generating schema-based hypotheses about which objects are likely to be present in a given scene (Friedman, 1979). However, although these results clearly demonstrate that people can combine context and object information in object recognition, it is important to note this conclusion does not bear directly on the experimental question at issue in many of these earlier studies, which were concerned with whether context can speed up recognition of fully recognizable images. In the current study, participants were able to spend as long as they wanted performing the experimental task and the facilitation came in the form of being able to correctly identify the object on the basis of reduced visual information, not reduced time. Thus, the standard mechanisms of perceptual priming, which are usually measurable by speeding up cognitive processes, were unlikely to be a factor. Furthermore, because the current study used real-world images and objects, rather than artificially generated stimuli, the context was truly informative about the identity of the object. This may be contrasted with previous studies in which the design was such that the contextual category did not predict whether a particular target object was present or not (Biederman et al., 1982; Hollingworth & Henderson, 1998). Nevertheless, with regard to the basic theoretical question of whether or not object recognition can be influenced by contextual scene processing, the current study provides the strong evidence that context can be highly influential. Indeed, one way of thinking about the current results is that the context in which an object appears can serve as an important "feature" for purposes of recognition just like the visible properties of the object itself.

The current results do not directly address brain mechanisms, but several previous studies may shed some light on the likely neural mechanisms underlying the contextual facilitation effects observed here. Because the facilitation in the familiar and unfamiliar both depended on spatial and object memory, both are likely to involve medial-temporal (MTL) structures, particularly the hippocampus and associated structures, which have long been known to be important in encoding and storing memory (Squire & Zola-Morgan, 1991). However, the schema-based memory in the unfamiliar-context condition may rely on some different structures outside of MTL as well, including frontal and occipitotemporal regions. In particular, Bar and colleagues (Bar, 2003, 2004; Bar et al., 2006) have proposed a theory of neural activation during object recognition whereby low spatial-frequency visual representations are transmitted from occipital cortex to the Orbital Frontal Cortex (OFC) in the frontal lobe; this structure uses the coarse representations to generate guesses about the object's identity, transmitting these back to occipitotemporal regions to guide further visual processing. Bar and colleagues (Bar, 2003; Bar & Aminoff, 2003) have also proposed that *scene*-based contextual information is subserved by a distinct cortical network in MTL encompassing the parahippocampal cortex (PHC) and

the retrosplenial complex (RSC). Both of these regions have previously been shown to be involved in the processing of location information (Aguirre, Zarahn, & D'Esposito, 1998); one specific region in PHC, sometimes referred to as the "parahippocampal place area" or PPA, shows selective activation to visual scenes (Epstein & Kanwisher, 1998). The RSC, part of the cingulate gyrus in medial cortex, has been found in human imaging studies to be involved in episodic memory processing (Svoboda, McKinnon, & Levine, 2006) and navigation (Maguire, 2001). Thus, this proposed network may be thought of as forming the neural basis for experience-based scene schemas which may be activated in order to facilitate object recognition. The degraded objects used as stimuli in the current study are likely similar to the LSF images; that is they may be used to generate hypotheses about the objects' likely identity, whereas the presence of scene information is likely to activate the PHC/RSC network in generating schema-based expectation. If so, the interaction between top-down, schema-based, expectations and bottom-up information in the current study may depend on a convergence of medial-temporal, frontal, and occipitotemporal regions.

The facilitation of the familiar participants in Experiment 1, however, likely depended exclusively on structures located in medial temporal cortex, particularly the hippocampus and several associated regions. In humans, several studies have demonstrated that patients with damage to hippocamapus structures exhibit amnesic deficits in remembering associations between objects and their locations (Smith & Milner, 1981; Stepankova, Fenton, Pastalkova, Kalina, & Bohbot, 2004). Similar results have been shown in animal studies where localized lesions to the hippocampus and nearby structures, such as perirhinal cortex, the fornix, and the parahippocmpal cortex, severely impair learning and/or memory of object–place associations (Bachevalier & Nemanic, 2008; Gaffan, 1994; Gaffan & Parker, 1996; Malkova & Mishkin, 2003; Parkinson, Murray, & Mishkin, 1988). More recently, single-neuron recording studies in rates have identified specific neurons in perirhinal cortex—an MTL structure that plays an important role in object memory (see Murray, Bussey, & Saksida, 2007, for a review)—that respond preferentially to specific object–location pairings after they have been learned (Rolls, Xiang, & Franco, 2005). Since the current results suggest that object recognition in familiar environments depends heavily on memory of object–location associations, it is likely that such recognition depends heavily on these hippocampal and associated brain regions.

CONCLUSION

The current results provide perhaps the most direct evidence to date that context can facilitate object recognition, in this case by reducing the amount of visual information needed to identify an object. In highly familiar environments, as in the familiar-context conditions, it appears that people often require very little

visual information beyond the location and approximate size of an object in order to identify it, relying mostly on memory. Because people often spend much of their time in a small set of physical environments (e.g., office, home), recognition under typical conditions may require much less of the kind of perceptual processing of the sort proposed by previous theories of recognition. Even in unfamiliar environments, these results suggest that context greatly reduces the visual information needed to recognize objects relative to without any context, likely based on schema-based expectations about likely objects, given the scene type. Because object recognition almost always takes place within a rich, constrained environment, the role of context is likely of paramount importance in everyday recognition. Thus, regardless of the mechanism underlying contextual facilitation, the current results reinforce the notion that theories of visual object recognition must take contextual factors into consideration.

REFERENCES

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron*, *21*(2), 373–383. doi:10.1016/S0896-6273(00)80546-2

Auckland, M., Cave, K., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin and Review*, *14*(2), 332–337. doi:10.3758/BF03194073

Bachevalier, J., & Nemanic, S. (2008). Memory for spatial location and object-place associations are differently processed by the hippocampal formation, parahippocampal areas TH/TF and perirhinal cortex. *Hippocampus*, *18*(1), 64–80. doi:10.1002/hipo.20369

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. doi:10.1162/089892903321662976

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. doi:10.1038/nrn1476

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*(2), 347–358. doi:10.1016/S0896-6273(03)00167-3

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the USA*, *103*(2), 449–454. doi:10.1073/pnas.0507062103

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, *25*(3), 343–352. doi:10.1068/p250343

Biederman, I. (1987). Recognition-by-components—a theory of human image understanding. *Psychological Review*, *94*(2), 115–147. doi:10.1037/0033-295X.94.2.115

Biederman, I., Mezzanotte R. J., Rabinowitz J. C., Francolini C. M., & Plude, D. (1981). Detecting the unexpected in photointerpretation. *Human Factors*, *23*, 153–164.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177. doi:10.1016/0010-0285(82)90007-X

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 556–566. doi:10.1037/0096-1523.15.3.556

Bruner, J. S., & Potter, M. C. (1964). Interference in visual recognition. *Science*, *144*(3617), 424–425. doi:10.1126/science.144.3617.424

Cox, D., Meyers, E., & Sinha, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science*, *304*(5667), 115–117. doi:10.1126/science.1093110

Davenport, J. (2007). Consistency effects between objects in scenes. *Memory and Cognition*, *35*(3), 393–401. doi:10.3758/BF03193280

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. doi:10.1111/j.0956-7976.2004.00719.x

Dongxiang, Z., Hong, Z., & Ray, N. (2008, June). *Texture based background subtraction*. Paper presented at the international conference on Information and Automation, China. pp. 601–605.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. doi:10.1038/33402

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*(3), 316–355. doi:10.1037/0096-3445.108.3.316

Gaffan, D. (1994). Scene-specific memory for objects: A model of episodic memory impairment in monkeys with fornix transection. *Journal of Cognitive Neuroscience*, *6*(4), 305–320. doi:10.1162/jocn.1994.6.4.305

Gaffan, D., & Parker, A. (1996). Interaction of perirhinal cortex with the fornix-fimbria: Memory for objects and "object-in-place" memory. *Journal of Neuroscience*, *16*(18), 5864–5869.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*(2), 123–144. doi:10.1016/S0926-6410(02)00244-6

Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience*, *20*(3), 371–388. doi:10.1162/jocn.2008.20027

Harmon, L. D. (1973). The recognition of faces. *Scientific American*, *229*(5), 70–83. doi:10.1038/scientificamerican1173-70

Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science*, *180*(4091), 1194–1197. doi:10.1126/science.180.4091.1194

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271. doi:10.1146/annurev.psych.50.1.243

Hollingworth, A. (2005). Memory for object position in natural scenes. *Visual Cognition*, *12*(6), 1003–1016. doi:10.1080/13506280444000625

Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 58–69. doi:10.1037/0278-7393.32.1.58

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 31–47. doi:10.1037/0096-1523.33.1.31

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, *127*(4), 398–415. doi:10.1037/0096-3445.127.4.398

Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. l. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*(13), 1–18. doi:10.1167/8.13.11

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavain Journal of Psychology*, *42*(3), 225–238. doi:10.1111/1467-9450.00233

Malkova, L., & Mishkin, M. (2003). One-trial memory for object-place associations after separate lesions of hippocampus and posterior parahippocampal region in the monkey. *Journal of Neuroscience*, *23*(5), 1956–1965.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Henry Holt & Co. Inc.

Murray, E. A., Bussey, T. J., & Saksida, L. M. (2007). Visual perception and memory: A new view of medial temporal lobe function in primates and rodents. *Annual Review of Neuroscience*, *30*, 99–122. doi:10.1146/annurev.neuro.29.051605.113046

Palmer, S. E. (1975). Effects of contextual scenes on identification of objects. *Memory and Cognition*, *3*(5), 519–526. doi:10.3758/BF03197524

Parkinson, J. K., Murray, E. A., & Mishkin, M. (1988). A selective mnemonic role for the hippocampus in monkeys: Memory for the location of objects. *Journal of Neuroscience*, *8*(11), 4159–4167.

Rolls, E. T., Xiang, J., & Franco, L. (2005). Object, space, and object-space representations in the primate hippocampus. *Journal of Neurophysiology*, *94*(1), 833–844. doi:10.1152/jn.01063.2004

Smith, M. L., & Milner, B. (1981). The role of the right hippocampus in the recall of spatial location. *Neuropsychologia*, *19*(6), 781–793. doi:10.1016/0028-3932(81)90090-7

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253* (5026), 1380–1386. doi:10.1126/science.1896849

Stepankova, K., Fenton, A. A., Pastalkova, E., Kalina, M., & Bohbot, V. R. D. (2004). Object–location memory impairment in patients with thermal lesions to the right or left hippocampus. *Neuropsychologia*, *42*(8), 1017–1028. doi:10.1016/j.neuropsychologia.2004.01.002

Sun, H.-M., Simon-Dack, S. L., Gordon, R. D., & Teder, W. A. (2011). Contextual influences on rapid object categorization in natural scenes. *Brain Research*, *1398*, 40–54. doi:10.1016/j.brainres.2011.04.029

Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia*, *44*(12), 2189–2208. doi:10.1016/j.neuropsychologia.2006.05.023

Torralba, A. (2009). How many pixels make an image? *Visual Neuroscience*, *26*(1), 123–131. doi:10.1017/S0952523808080930

Torralba, A., Murphy, K., Freeman, W., & Rubin, M. (2003). Context-based vision system for place and object recognition. *Proceedings of the IEEE Conference on Computer Vision*, *9*, 273–280.