Intrinsic and contextual features in object recognition

Derrick Schlangen

Psychology Department, Florida Atlantic University,
Boca Raton, FL, USA ✉

Elan Barenholtz

Psychology Department, Florida Atlantic University,
Boca Raton, FL, USA 🏠✉

The context in which an object is found can facilitate its recognition. Yet, it is not known how effective this contextual information is relative to the object's intrinsic visual features, such as color and shape. To address this, we performed four experiments using rendered scenes with novel objects. In each experiment, participants first performed a visual search task, searching for a uniquely shaped target object whose color and location within the scene was experimentally manipulated. We then tested participants' tendency to use their knowledge of the location and color information in an identification task when the objects' images were degraded due to blurring, thus eliminating the shape information. In Experiment 1, we found that, in the absence of any diagnostic intrinsic features, participants identified objects based purely on their locations within the scene. In Experiment 2, we found that participants combined an intrinsic feature, color, with contextual location in order to uniquely specify an object. In Experiment 3, we found that when an object's color and location information were in conflict, participants identified the object using both sources of information equally. Finally, in Experiment 4, we found that participants used whichever source of information—either color or location—was more statistically reliable in order to identify the target object. Overall, these experiments show that the context in which objects are found can play as important a role as intrinsic features in identifying the objects.

Introduction

In the natural world, objects typically appear within a rich and complex surrounding scene. Because certain objects tend to appear with higher frequency within certain contexts than other contexts (Greene, 2013), the scene in which an object appears may carry information about its identity, which may facilitate recognition. Numerous studies have found that detection and recognition of a target object is faster and more accurate when it is presented within a consistent contextual scene (Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce, Pollatsek, & Rayner, 1989; Davenport & Potter, 2004; Palmer, 1975). The precise nature and extent of such facilitation remains controversial (Henderson & Hollingworth, 1999). One possibility is that the scene activates perceptual processing specifically appropriate to the target stimulus, making the recognition process more efficient (Biederman et al., 1982; Boyce & Pollatsek, 1992; Boyce et al., 1989; Palmer, 1975). Alternatively, the presentation of a context may lead to a reduced criterion of feature matching between the target stimulus and some representation stored in memory (Friedman, 1979). Finally, some have argued 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 and that context and object recognition are functionally isolated (Henderson & Hollingworth, 1999; Hollingworth & Henderson, 1998).

A common feature of the above studies is that the target stimuli were identifiable even without the context. However, context can play a more direct role in recognition, providing information about the object's identity when the image of the object is degraded. Bar and Ullman (1996) showed that people could identify segmented portions of stylized line drawings more accurately when the segments were shown in the appropriate spatial relations to one another. Cox, Meyers, and Sinha (2004) found that images of faces that were degraded to the point of unrecognizability on their own activated the fusiform face area (FFA) when shown in the context of the rest of the person. More recently, Barenholtz (2014) found that showing photographed objects in their original contextual setting greatly reduced the minimum resolution needed to identify them. The benefits of context were greatly enhanced when the contextual environment in which the objects were located was familiar to the partici-
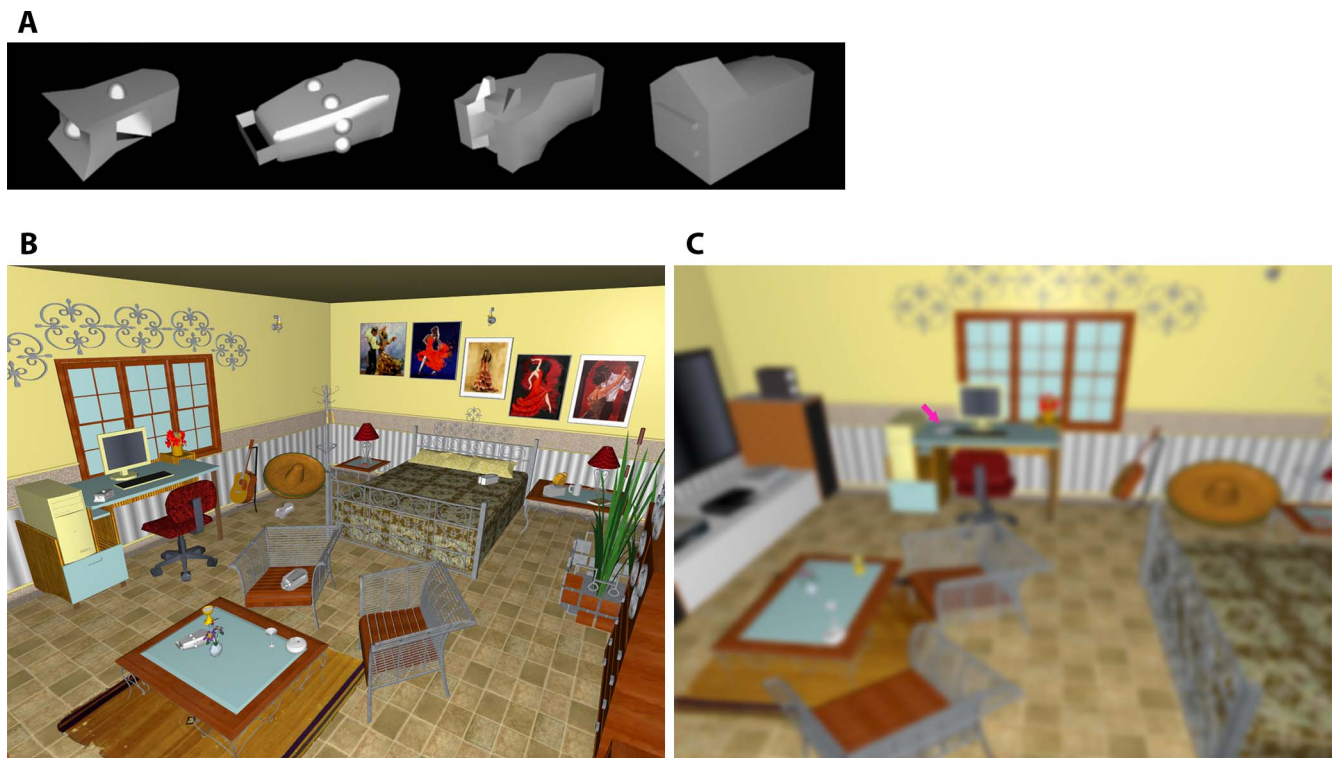
Figure 1. Sample stimuli used in Experiment 1. The scene viewpoint was different on each trial. (A) Examples of the rendered novel objects used as targets and distractors in the search task. (B) Sample stimulus from the search phase in Experiment 1. Participants searched the scene and indicated by keyboard response whether the target object was present or absent in the scene. Both context (i.e., scene location) and, in Experiments 2, 3, and 4, intrinsic features (color) of the target objects were manipulated in the search phase; see text for details. (C) Sample stimulus from the identification phase in Experiment 1, with an arrow pointing at the target object. The scene was blurred such that the target objects' distinguishing shape information was eliminated but contextual location (and the color of the object in Experiments 2, 3, and 4) was still visible.

pants, a finding consistent with studies showing that people encode the specific locations of objects in previously viewed scenes (Hollingworth, 2005, 2006, 2007).

These studies demonstrate that the context in which an object is found can serve as a direct source of information for the purposes of recognition, similar to the "intrinsic" features of an object (e.g., shape and color). However, because these previous studies used stimuli in which both contextual and intrinsic information were always present at the same time, it is not possible to determine the scope and importance of context in recognition compared with intrinsic information. To address this, the current study employed a paradigm in which participants were trained on a set of novel, computer-generated objects—each defined by a unique shape—embedded in a rendered three-dimensional environment (Figure 1). We independently manipulated the target objects' "intrinsic" features (in this case, color) as well their "contextual" information (in this case, location within a scene) as participants searched for them within the scene. After this search task, participants performed a surprise identification task in which one of the novel objects was presented—in context—in a highly blurred image (Figure 1C). The

blurring served to eliminate the target object's shape as a basis for identifying it, forcing participants to use the object's color, location, or both (depending on the experimental condition) in order to identify it. Using this methodology, we examined four questions concerning the relative roles of contextual location and intrinsic objects features in object recognition: Can contextual location serve as a basis for identification on its own, in the absence of any intrinsic features that specify the objects' identity (Experiment 1)? Will participants combine location and color information to determine the identities of objects (Experiment 2)? Is there a bias for color or location information when they are each equally informative about identity but are in conflict with one another (Experiment 3)? Is there a bias for color or location information when one is more informative about identity than the other (Experiment 4)?

Experiment 1

In the first experiment, we investigated whether participants would identify objects based solely on their

previously learned locations in the absence of any distinguishing intrinsic object features. While previous studies have demonstrated that context can serve to supplement recognition of objects with intrinsic identifying features, no study to date has tested experimentally whether the identity of objects can be determined in the absence of such features. Thus, this is a critical test of the validity of contextual location as a basis for recognition.

We created a set of novel, computer-generated objects to be used as search targets, designed to be somewhat similar in gross visual properties but also visually distinct on closer inspection (Figure 1A). In the experiment, participants first performed a search phase, which consisted of a visual search task for one of these objects within a rendered bedroom scene that also contained several other novel objects in addition to typical bedroom accouterments (Figure 1B). The participant's task was to determine, on each trial, whether the specific target object, designated before the search, was present in the scene. Unbeknownst to participants at the beginning of the experiment, the locations of the various novel objects were experimentally manipulated throughout the search phase. One set of four objects always appeared in the same locations in the scene whenever they were present ("fixed location"). For example, object 1 always appeared on the computer desk throughout the search phase, whereas object 2 always appeared on the coffee table. The other target objects each appeared in one of four interchangeable locations whenever they were present in the search stimulus ("variable location"). For example, both objects 5 and 6 may have appeared on the dresser, center chair, bed, or entertainment center. We predicted that participants would show overall faster search times for the fixed-location objects compared with the variable-location objects. This prediction is based on previous evidence of contextual cueing, a phenomenon in which search times decline across repetitions of a search stimulus (Chun & Jiang, 1998, 2003). In particular, several studies found a similar phenomenon when the search stimuli consist of naturalistic scenes, such as the ones used in the current study (Brockmole, Castelhano, & Henderson, 2006; Brockmole & Henderson, 2006). (It is worth noting, however, that the current study is the first, to our knowledge, to test for such a phenomenon across variable viewpoints in a naturalistic scene.) Any observed contextual cueing for the fixed-location objects may be taken as evidence that participants were learning these object–context pairings.

After this initial search phase, each participant performed a surprise identification phase in which they were briefly presented with blurred images of the bedroom scene, each with a single test object shown in one of the locations that had been occupied by a fixed-location object during the search phase (Figure 1C).

The images were blurred to a degree such that the overall layout of the scene could still be easily discerned but the target object could not be identified based on its shape information (see Supplementary Experiment 1a). The participant's task was to choose in a forced-choice task which target object was present in the blurred image. If participants identified the blurred target as the object that had been associated with that location during the search phase, this would demonstrate that contextual location alone is sufficient to drive such identification behavior. It is important to note that while the blurring ensured that the target object was not identifiable based on its visual characteristics, the participants were never given this information directly and were therefore (presumably) under the assumption that there was a "correct" answer about which object was actually represented in the image. This experimental design was intended to induce behavior more reflective of natural recognition rather than a strategy of pure guessing.

Method

Participants

A total of 23 Florida Atlantic University undergraduate students participated in this experiment, satisfying a course requirement. All subjects had normal or corrected-to-normal vision.

Stimuli

An example stimulus is shown in Figure 1B. Each stimulus consisted of a rendered three-dimensional bedroom environment that included typical objects that might appear in a bedroom (e.g., a computer desk, flowers, lamps). The same bedroom scene, shown from different viewing angles, was used in generating all stimuli. Each stimulus (160 total) showed a different viewpoint of the scene. In addition, eight novel objects were generated to be used as both targets and distractors (see Figure 1A for examples) and were designed to appear as realistic but unfamiliar objects.

A total of 160 rendered scene images were included in the search phase: 10 target-present trials and 10 target-absent trials for each of the eight objects. There were eight locations in the bedroom scene where the target objects could appear. The four fixed-location target objects appeared only in their own designated location (i.e., target object 1 on the computer desk, target object 2 on the coffee table, target object 3 on the floor between a guitar and a sombrero, and target object 4 on the nightstand), with a small amount of variation within that area. For example, target object 1 could be found on either side of the computer desk, as long as it always appeared on the computer desk. The four variable-
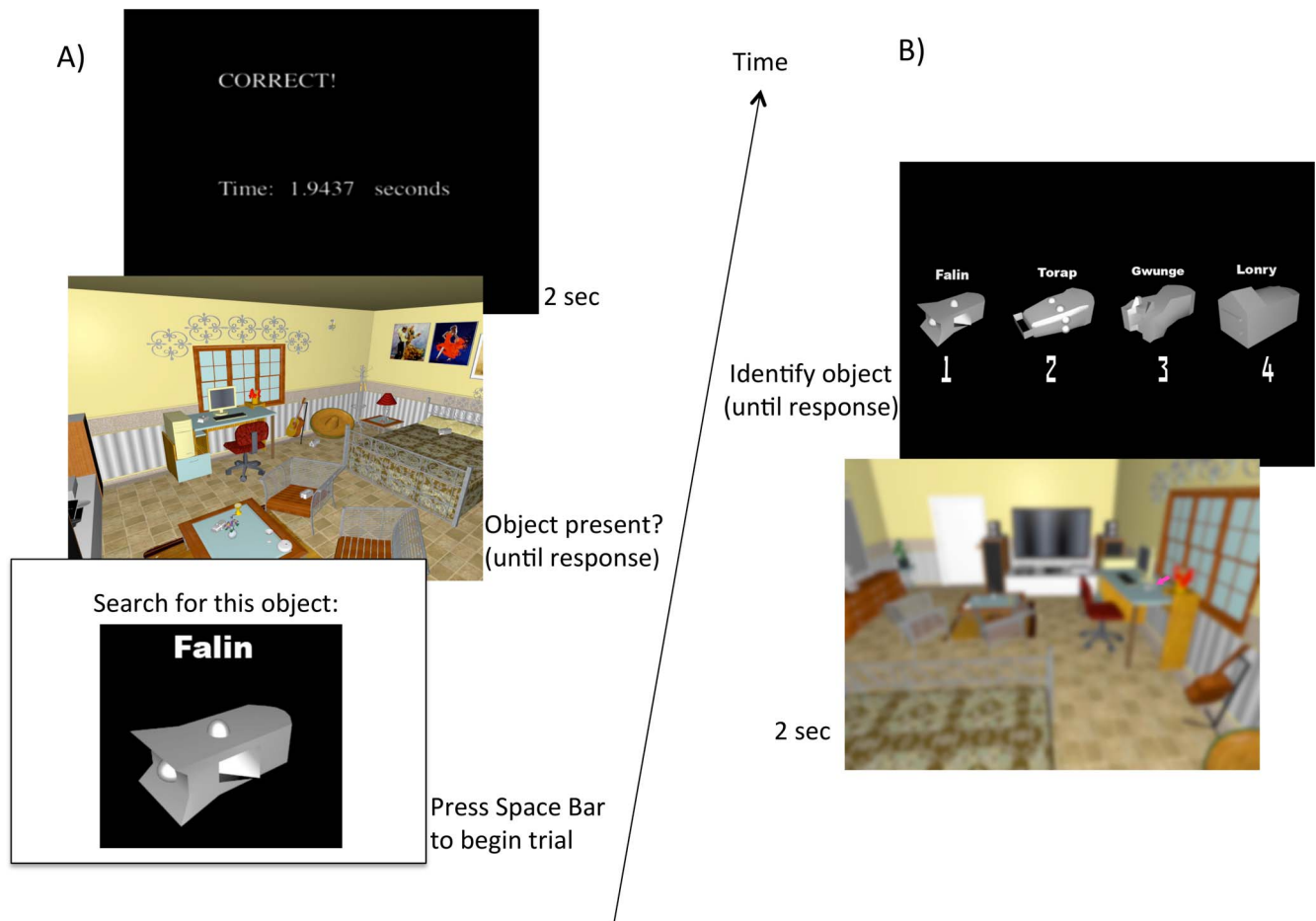
Figure 2. (A) Schematic of a single trial in the search phase of Experiments 1, 2, 3, and 4. The search cue was presented until the participant pressed the keyboard. Then, the search stimulus was presented until the participant responded. Finally, a feedback screen appeared for 2 s. (B) Schematic of a single trial in the identification phase of Experiments 1, 2, 3, and 4. The blurred scene (target designated by a pink arrow) was presented for 2 s, followed by the four objects from which the participant must choose when identifying the target.

location target objects were rotated among four locations within the bedroom scene (see above for specific locations) and did not appear in any of the fixed target locations. For example, target object 5 may have appeared on the entertainment center on one trial but could also appear on the dresser, bed, or chair (i.e., on any of the variable target locations) on subsequent trials. Importantly, the variable-location objects appeared only in these four locations, and several of these objects were present on every trial. This ensured that any search advantage for the fixed-location objects was not simply due to learning which locations typically held target objects. The bedroom was shown from many different viewpoints over the course of the experiment; participants were exposed to a different viewpoint of the room on each trial to search for the target object. All of the bedroom items except for the target and distractor objects remained in the same location.

In the identification phase, four new blurred images were generated for each of the four fixed locations for a total of 16 stimuli (Figure 1C), each showing a different viewpoint of the scene. Variable locations were not tested because they lacked an exclusive object pairing. See above for specific details of the identification stimuli.

Procedure

Before beginning the search phase, participants were first shown each of the eight target objects, repeated twice, in order to familiarize them with the objects' shapes. In order to enhance familiarity and recognition of the different objects, each was paired with a novel name (e.g., "Lonry," "Torap"); participants were shown a preview of each object and its name before the initiation of the search phase. Next, participants performed the search phase. A sample trial sequence is shown in Figure 2A. Participants first viewed a picture of the target object (a single fixed-location or variable-location target object) with its name, and then pressed

the space bar when they were ready to continue. Next, they were shown the search stimuli, half of which contained the cued target object and half of which did not. The search stimulus was presented until participants made a response. The participants' task was to respond by pressing *Y* on the keyboard if the target object was present in the scene or pressing *N* if the target object was not present, as quickly and as accurately as possible. Finally, to encourage accuracy and promptness in responding, a feedback screen displaying response time and accuracy of response was shown for 2 s. The search phase was divided into 10 blocks of 16 trials for a total of 160 trials. Each block showed a picture of each target object twice, once where the target object was present and once where the target object was absent (i.e., distractor-only trials). Thus, half of the trials contained the target object and the other half of the trials did not.

Participants then performed a surprise identification phase consisting of a total of 16 trials (Figure 2B). On each trial of the identification phase, participants were presented with a blurred image of the same bedroom scene used for the search phase with an arrow pointing to the target object. The stimulus image was shown for 2 s in order to simulate a quick glance at an object. The stimulus image was then replaced by a lineup of the four fixed-location target objects, and participants chose which object had been presented in the previous image by pressing *1*, *2*, *3*, or *4* on the keyboard. The objects were assigned the same number and position in the lineup on each trial.

Supplementary Experiment 1a: In order to determine whether there was any information in the blurred images of the objects that might bias participants to choose one object over the other, a different set of 15 independent participants performed the same identification phase task (with no time constraint) without exposure to the search phase and thus no knowledge about the locations of the objects.

Results

Search phase

One participant was removed from the analysis because of chance-level performance (50%) in the search phase. Overall accuracy in the search phase was 74.51%, 95% confidence interval (CI) [71.54, 77.47], *SD* = 6.69. For the purpose of investigating reaction times, incorrect responses were removed from the analyses and a 5% trimmed mean was computed to remove extreme scores. In order to determine whether participants were learning the fixed-location objects during the search phase, we first compared search reaction times in the search phase for the fixed-location objects with those for the variable-location objects. Figure 3 shows mean reaction times for the fixed-location target

objects and the variable-location target objects across the 10 experimental blocks. Comparing overall reaction times from blocks 2 through 10, fixed-location target objects were located in 2.09 s ($SD = 0.37$), whereas variable-location target objects were located in 2.45 s ($SD = 0.42$). The mean reaction time difference between conditions was 0.355 s, 95% CI [0.253, 0.457]. A paired-samples t test revealed that the reaction time difference was significant, $t(21) = 7.23$, $p < 0.001$. Thus, fixed-location target objects were located faster overall than variable-location target objects. Additionally, in order to assess the rate of learning, we compared the decrease in reaction time between block 1 and block 10 for the fixed-location and variable-location target objects. A paired-samples t test showed that fixed-location target objects (−0.731 s) had a significantly steeper decrease in reaction than variable-location target objects (−0.268 s), $t(19) = 2.71$, $p = 0.014$, mean difference decrease in reaction time = 0.463 s, 95% CI [0.105, 0.882]. These results (a) demonstrate that participants learned the locations of the fixed-location objects and used this information to facilitate their search and (b) represent a novel extension of the contextual cueing effect (Brockmole et al., 2006; Chun & Jiang, 1998) to a case in which the context is defined in a viewpoint-independent manner based on location within a three-dimensional environment.

Identification phase

The primary analysis of interest for Experiment 1 involved the identification phase results. In the degraded image identification phase, participants' choices were consistent with identifying the test object based on location information on 62.8% of the trials ($SD = 30.33\%$). This was significantly greater than would be expected by chance (25%), $t(21) = 5.85$, $p < 0.001$. The mean difference effect size was 37.8%, 95% CI [24.35, 51.25].

Figure 4 shows the percentage of trials on which each subject selected the object consistent with the location information, ordered from lowest to highest. It is apparent from the graph that the average performance (62.8%) is not actually representative of most participants; instead, there was a wide diversity of performance. A minority of participants (5 out of 22) used location on all trials; a similarly sized group did not use the location information at all; and a majority appeared to use the location information on some but not all trials, with a wide variety of frequency across participants. More than half of participants (13 out of 22) chose the object consistent with location on only ≤75% of trials.

One possible interpretation of the variability in performance in the identification task is that it reflects different levels of learning and memory for the object's
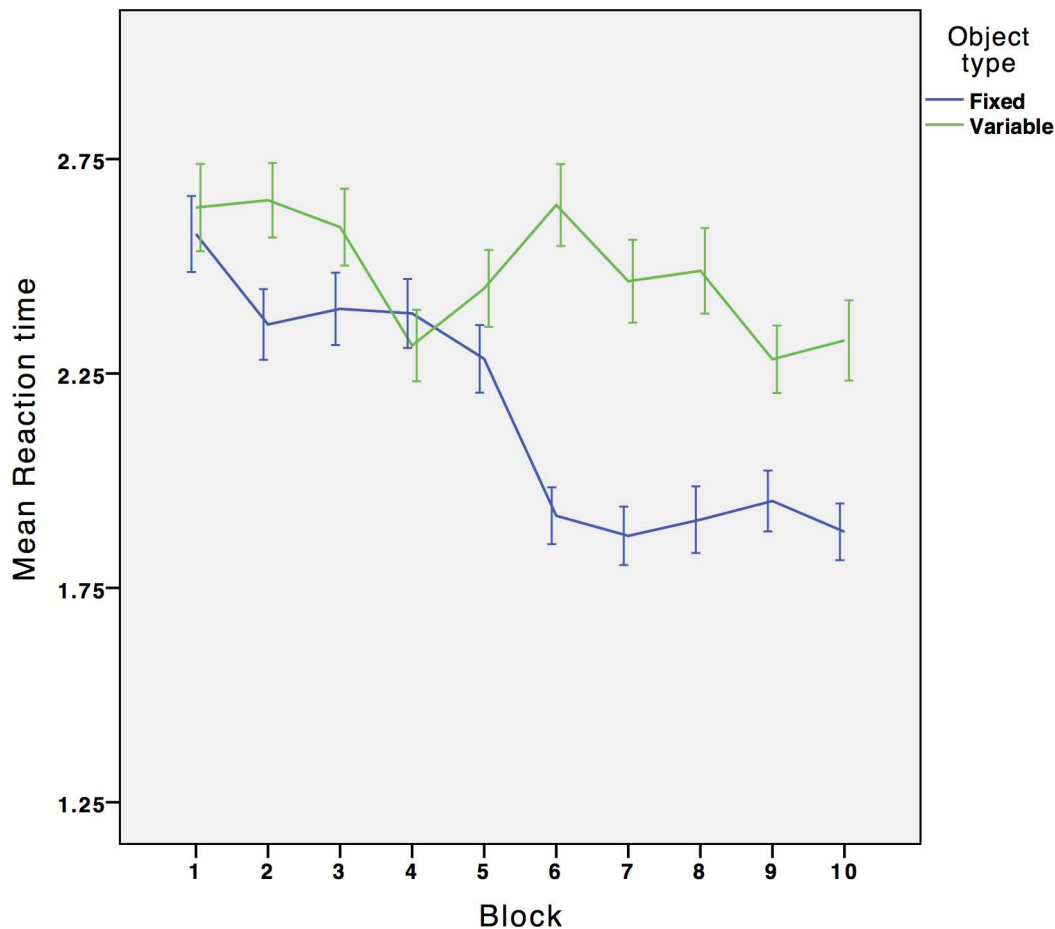
Figure 3. Reaction time data (s) in the search phase of Experiment 1 for the fixed and variable target objects. Participants located fixed-location target objects faster than variable-location target objects in the later blocks of the search phase. Error bars represent ± 1 *SE*.

locations. Alternatively, it could be that what varied most across participants was not the degree to which they learned the locations but rather the extent to which they applied this information to the identification task. To further examine this issue, we assessed the difference in reaction time between the fixed-location and variable-location objects for each participant (i.e., a measure of contextual cueing). Out of the 22 participants, 19 demonstrated a difference of 100 ms or higher, suggesting that almost all the participants showed some learning of the objects' locations. However, some participants showed a stronger contextual effect than others. Therefore, we tested to see whether there was a correlation between the contextual cueing effect size and a tendency to choose the identity consistent with location in the identification phase. The correlation was not significant, $r(20) = 0.194, p = 0.388$. Similarly, the tendency to choose the identity consistent with location was not significantly correlated with the overall accuracy in the search phase, $r(20) = 0.240, p = 0.283$. Finally, the decrease in reaction time from block 1 to block 10 for the fixed-location target objects was

not correlated with the tendency to identify the object consistent with location, $r(18) = 0.188, p = 0.427$. These results suggest that the variability in identification performance was not simply the result of variability in initial learning of the objects' locations.

As discussed earlier, the design of this experiment was intended to induce a strategy in which the participant believed there is a correct answer to the identification task in the identification phase. This raises the question of whether the participants in this experiment were engaging in a behavior that is similar to more typical recognition—that is, trying to identify the object based on the belief that there is a correct response—or were guessing based on location in a manner that is not typical of real-world recognition. If participants were engaging in pure location-based guessing, then we would expect them to respond identically to each location across different trials (e.g., seeing the blurred object on the computer desk would always result in choosing object 1). However, as noted, the choice behavior of most participants was quite variable—that is, choosing based on location on many
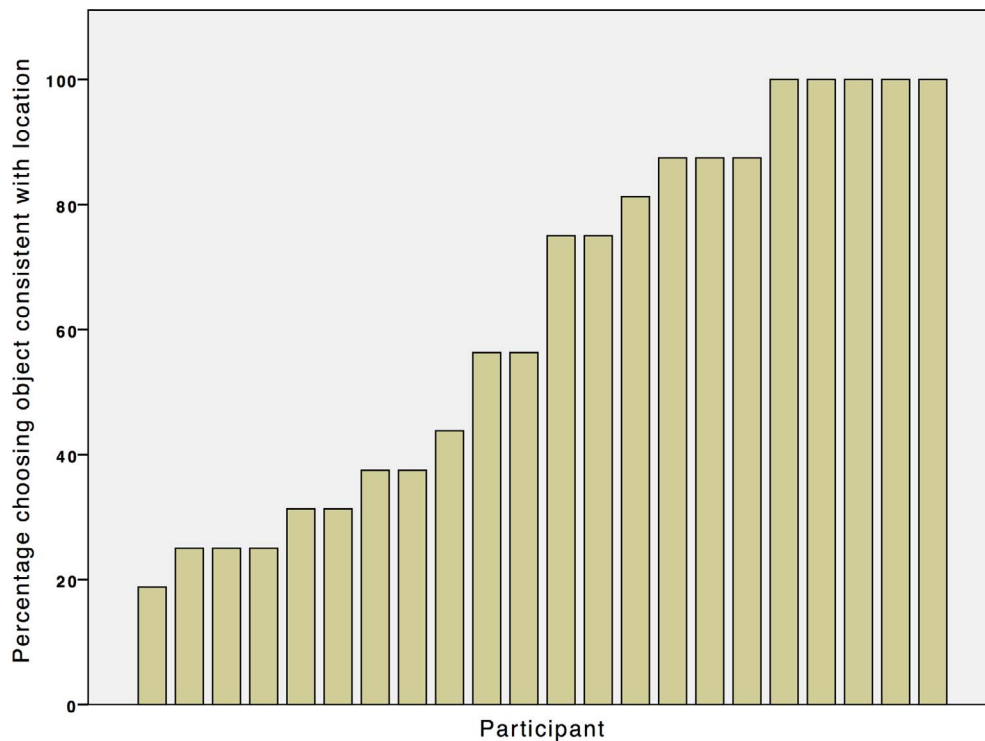
Figure 4. Percentage of trials in the identification phase of Experiment 1 in which participants selected the object that was consistent with the location information from the search phase. Each bar represents a single participant's performance.

trials but also choosing the target that was inconsistent with location on a substantial number of trials. Additional analysis of these inconsistent trials suggests that these responses were distributed across different objects. Of the 13 participants who chose the target object consistent with location on only ≤75% of trials, 10 chose an inconsistent object across all four locations tested, while the other three participants chose an inconsistent object for three of the locations. These results suggest that participants were not simply applying a consistent strategy of guessing based on location.

Supplementary Experiment 1a: The independent participants who performed the identification phase task without being trained on the location information selected the location-based choice on 26.25% of trials, a number that was not significantly greater than chance (25%), $t(14) = 0.295$, $p = 0.772$. The mean difference was 1.25%, 95% CI [–7.83, 10.33]. Thus, the visible information in the blurred objects did not provide any information for performing the identification task.

Discussion

The results of Experiment 1 demonstrate that people rely on contextual information to identify an object when the object's intrinsic visual features are unavailable for the purposes of identification. In these situations, context may not only facilitate identification but also be used as the only source of information for identifying the object when it is unrecognizable by its own intrinsic features. While previous studies have suggested that contextual information may be combined with bottom-up visual information for the purposes of recognition (Barenholtz, 2014), to our knowledge this is the first study to demonstrate that context alone may guide identification behavior in the absence of other sources of information.

However, while the results clearly demonstrate a tendency to use contextual information, only a minority of participants chose the location-consistent object on all trials. Most participants showed a tendency to do so most often but made other choices on a significant number of trials. As noted above, this may reflect variable learning of the objects' locations in the first place across participants or it may reflect variability in participants' tendencies to apply the location information in the identification task. While by no means definitive, the lack of a correlation between the contextual cueing effect size and performance in the identification phase may be taken as evidence against the former theory; a larger contextual cueing effect likely reflects better learning of the object locations, yet it did not lead to higher frequencies of location-based identifications. However, it is important to note that contextual cueing (Chun & Jiang, 1998) is often marked by the absence of explicit

awareness of the contingency between location and context—although with naturalistic stimuli, explicit memory is present (Brockmole et al., 2006). Thus, some participants may have had only implicit knowledge of the object–location associations. This may have facilitated their search for the fixed-location objects during the search phase but not always translated this knowledge to the identification task, which may rely more on explicit knowledge of the object–location associations.

A different possibility is that what varied across participants was not their learning and memory of the object's locations per se but rather the extent to which they applied this knowledge in the identification phase. As noted, the instructions of the experiment were such that participants likely believed that there was a "correct" response in the identification phase. The fact that most participants chose the object that was not consistent with location on a substantial number of trials (while still showing a tendency to incorporate location on a majority of trials) suggests that participants were making considerations in addition to location in their choice behavior. We propose that this was most likely based on the (false) assumption that there was a correct answer and that some visual information was available, leading them to override location on some trials.

Experiment 2

Experiment 1 demonstrated that participants identify objects based on contextual location information alone when the objects were degraded to the point that they contained no intrinsic feature information that could be used to discriminate target identity. However, under typical conditions, both contextual and intrinsic information is available and may be combined for the purposes of recognition. Thus, in Experiment 2, we tested how location and color information are combined in identifying a degraded object. A paradigm similar to that used in Experiment 1 was used in Experiment 2. However, unlike Experiment 1, where color was uniform across all of the objects and only location served as the source of identification in the identification phase, in Experiment 2 both location and color carried information about object identity. Specifically, each object had a color that was shared with one other object and a location that was shared with a different object. However, the specific combination of color and location was unique to each object. For example, object 1 appeared on the computer desk and was white. Object 2 also appeared on the computer desk but was yellow. Finally, object 3 was white but

appeared on the nightstand. Thus, object 1 could not be identified based on being white or on the computer desk alone but only by the combination of these two dimensions. Thus, in the identification phase, where the shape information that distinguished target identity was eliminated, we assessed how often participants chose the object that was consistent with both color and location information in identifying the object.

Method

Participants

A total of 31 Florida Atlantic University undergraduate students with normal vision participated in this experiment, satisfying a course requirement.

Stimuli and procedure

The same bedroom scene stimuli and the same eight target objects used in Experiment 1 were used. However, in this experiment, four of the objects were used as target objects and the other four objects were used as distractor objects. Each target object shared a location with one other target object (but not on the same trial) and shared a color with another target object, but each target object had its own unique combination of location and color information. See above for an example of a target object's specific property information.

The target search cue was the same color as the target in the search scene. There were distractor objects—colored white, yellow, dark gray, or brown—in the target-present trials. Two of the distractor objects were always the same color as the target object on each trial. This made the search task more difficult by ensuring that participants could not simply use the color of the target object to identify it without examining the various objects in the scene. On target-absent trials, only distractor objects appeared in the search scene as any of the four colors, but they could not appear on the computer desk or nightstand, which were reserved for target objects only.

For the identification phase stimuli, four additional images were created for each of the four target objects for a total of 16 test stimuli. As in Experiment 1, each identification phase stimulus showed a blurred image containing a single test object in its location within the scene (i.e., no distractor objects were used). Both the location and color information could be easily discerned in these images. After viewing the test stimulus, participants had to identify the object from a lineup of the four target objects. All of the target objects in the lineup picture were shown in a base gray color so that
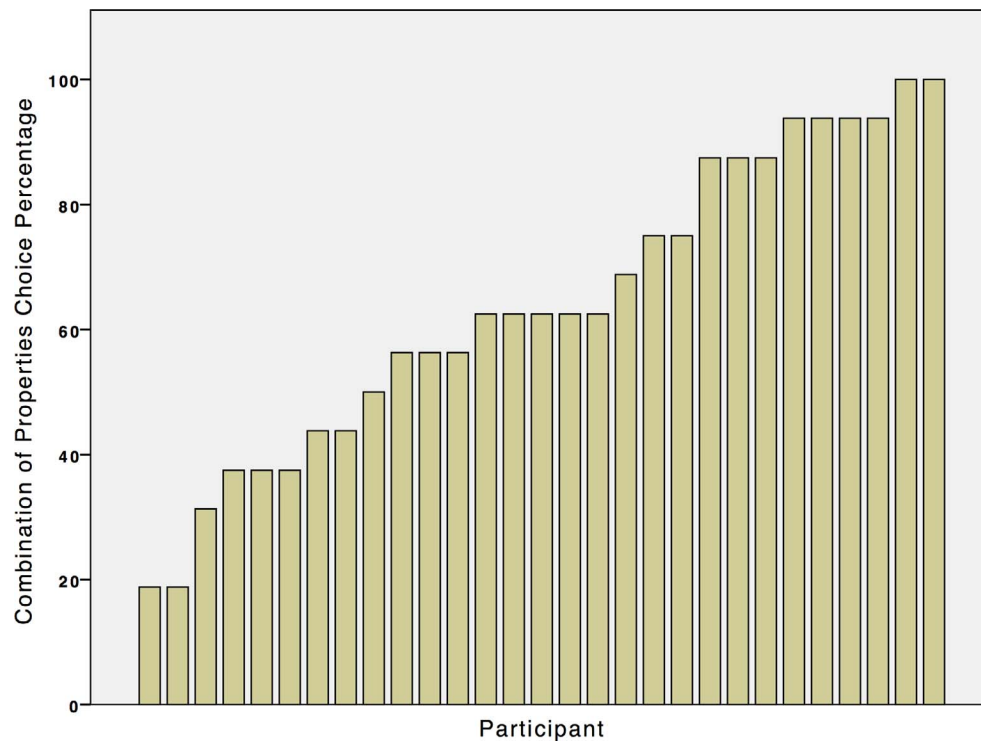
Figure 5. Percentage of trials in the identification phase of Experiment 2 in which participants selected the object that was consistent with the combination of location and color information from the search phase.

participants had to use their memory of the color information in order to perform the identification.

The procedure was identical to that in Experiment 1. There were 96 trials in the search phase and 16 trials in the identification phase.

Results and discussion

Two participants were removed from further analysis because they performed at chance accuracy in the search phase. The overall accuracy in the search phase was 86.97%, 95% CI [83.64, 90.26]. The overall reaction time for correct trials in the search phase was 1.94 s, 95% CI [1.77, 2.12].

In the identification phase, participants chose the target object that was consistent with the combination of the color and location of the test object on 64.03% of the trials ($SD = 24.30$). It is important to note that even if participants were choosing based on either color or location information alone, they would be expected to choose the "combined" object on half of the trials because only two of the four target choices would carry the correct value within any single property. Thus, we assessed whether participants chose the combination object more often than 50% of the time. This was found to be significant by t test, $t(28) = 3.11$, $p < 0.01$. The mean difference effect size was 14.03%, 95% CI [4.79, 23.27]. That is, participants showed a significant

tendency to use both sources of information in combination to identify the degraded target objects.

Figure 5 shows the percentage of trials, ordered from lowest to highest, on which each subject selected the object consistent with combining the color and location information. Visual inspection of the graph demonstrates that, as in Experiment 1, there was a high degree of variability across participants. About one-third of the participants identified the object consistent with the conjunction of properties at a high level (75% or higher), and a small minority (between 2 and 3 out of 29) appeared to engage in random guessing (i.e., chance performance). Finally, about half of the participants chose the object consistent with a combination of properties in the range of 40% to 60% of trials in the test phase. As noted above, 50% performance is predicted by identifying the object on the basis of color or location but not both. Thus, these participants may have been relying on one source of information or the other while ignoring the conjunction of features. Alternatively, these participants may have been using conjunction information when it was available to them but were unable to remember or apply this information in all cases. Supportive of this possibility, there was a significant positive correlation between accuracy in the search phase and choosing the object of the combined color and location in the identification phase, $r(27) = 0.65$, $p < 0.001$. However, no similar significant correlation was present between reactions times (RTs)

in the search phase and choice in the identification phase, $r(27) = 0.10$, $p = 0.607$.

Experiment 3

Experiment 2 showed that participants combine both location and color information to identify a degraded object. This raises an interesting question: Is one source of information more heavily weighted than the other when both are present? To test this, Experiment 3 used a paradigm similar to that used in Experiment 2 with the exception that during the search phase of Experiment 3, each of the target objects had high- and low-frequency locations in which they could appear as well as high- and low-frequency colors. During the identification phase, we tested behavior when the test object's color and location were in conflict (i.e., more consistent with one object than another). For example, on most trials target object 1 appeared on the computer desk and was most often white. However, it also appeared with lower frequency on the coffee table and appeared with a lower frequency with a color of brown. Target object 2 had the opposite property information: It usually appeared on the coffee table and with a brown color but sometimes appeared on the computer desk and with a white color. Then, on the critical trials in the identification phase, we examined identification behavior when the color and location information were in conflict; for example, a brown object (more consistent with object 2) on the computer desk (more consistent with object 1). In this case, either choice is equally consistent with the information in the search phase. Thus, a consistent tendency to choose one or the other would indicate a bias for that source of information. It may seem reasonable to assume that participants would give more weight to intrinsic features of an object when tasked with identifying an object because these properties should be more stable. However, the location of an object may be more salient than the color because it could indicate something about the function of the object.

Method

Participants

A total of 28 Florida Atlantic University undergraduate students participated in this experiment, satisfying a course requirement.

Stimuli and procedure

The same bedroom scene and eight objects from Experiments 1 and 2 were used. A total of 192 images (half for target-present trials and half for target-absent trials) were created as search phase stimuli. Four of the novel objects were used as target objects and the other four were used as distractor objects. In this experiment, both the location and color of the target objects were variable but statistically predictive of the objects' identity. Each of the target objects had a high- (i.e., 70.83%) and low- (i.e., 29.17%) frequency location as well as a high- (70.83%) and low- (29.17%) frequency color. The high-frequency properties of target object 1 were the computer desk location and the color white, and the low-frequency properties were the coffee table location and the color brown. Object 2 had the reverse property frequencies of object 1. The high-frequency properties of target object 3 were a location on the floor next to the guitar and the color yellow, and the low-frequency properties were the nightstand location and the color dark gray. Target object 4 had the reverse property frequencies of object 3. Each of the target objects had 24 trials in the search phase, which comprised the target-present trials. Of the 24 trials for each target object, 12 trials featured the high-frequency location and high-frequency color. Five trials showed the high-frequency location and low-frequency color, five trials showed the low-frequency location and high-frequency color, and two trials showed the low-frequency location and low-frequency color. Distractor objects could be any of the colors used for target objects (white, brown, dark gray, or yellow). They were placed in random locations throughout the scene but never appeared in the locations reserved for target objects.

For the identification phase stimuli, six additional images were created for each of the four target objects for a total of 24 test stimuli. The images were blurred so that only the color and location of each target object were visible. On 12 of the trials, the blurred test objects' location and color were both consistent with the high-frequency location and color of a single target object; this tested whether participants could identify the object when both color and location pointed to a specific object and essentially replicated the identification trials in Experiment 2. On 12 of the trials the color and location information were in conflict—that is, the location was consistent with the high-frequency location of one target object and the color information was consistent with the high-frequency color of a different target object. This allowed us to test whether participants had a bias for favoring location or color in the identification task.

The procedure was identical that in to Experiments 1 and 2 with the exception that the target search cues were always the same neutral-gray color while the actual target objects in the scenes were white, brown, dark gray, or yellow. Participants were informed that the target objects in the scenes would be "colored"

Figure 6. Breakdown of the proportion of trials in which each participant selected the object in the identification phase of Experiment 3 that was consistent with the color, the location, or neither property of the object from the search phase.

versions of the search cues. This ensured that participants had to learn the distribution of color for the various target objects during the search rather than learning the color distribution from the search cues themselves. There were 192 trials in the search phase and 24 trials in the identification phase.

Results and discussion

The overall accuracy in the search phase was 79.69%, 95% CI [77.03, 82.35]. The overall reaction time for correct trials in the search phase was 2.56 s, 95% CI [2.42, 2.70].

In the identification phase, on trials where the location and color were both consistent with the high-frequency color and location of a single target object, participants identified the test object consistent with the high-frequency properties on 61.61% of trials, those consistent with the low-frequency properties on 20.5% of trials, and those consistent with neither location nor color on 17.8% of trials. A Friedman's analysis of variance (ANOVA) revealed a significant overall difference in the identification choices, $\chi^2(2) = 22.81$, $p < 0.001$. A follow-up planned comparison showed that participants identified the object consistent with the high-frequency properties significantly more than with the low-frequency properties, $\chi^2(1) = 16.33$, $p < 0.001$. This result demonstrates that participants encoded the

frequency of the location and color information and chose the targets consistent with the higher frequency information when making an identification.

On conflict trials, participants' identification choices were consistent with location on 45.2% of trials, with color on 39% of trials, and with neither property on 15.8% of trials. Friedman's ANOVA revealed a significant overall difference in identification choices, $\chi^2(2) = 15.63$, $p < 0.001$. A follow-up planned comparison between location and color choices showed no significant difference between these property choices, $\chi^2 = 0.727$, $p = 0.394$. Thus, there was no significant bias for identifying the degraded object by either location or color.

Figure 6 shows the distribution of identification choices for each participant on the conflict trials. Visual inspection of the graph shows that some participants identified the object primarily by location, while a few others primarily used color. Most of the participants, however, used a mix of both location and color properties to identify the target object. Like the variability of responses in Experiment 1, these data suggest that participants were not engaging in an explicit guessing strategy based on either location or color but rather were attempting to recognize the object based on all of its available properties. The relatively low number of choices based on neither property in the conflict trials provides evidence that both the color and location were encoded and used for the purpose of

identification. However, neither source of information was given priority in the identification task.

Experiment 4

Experiment 3 examined identification behavior when two statistically equal properties were in conflict with one another. Experiment 4 further examined identification behavior under a condition of conflict in which the statistical reliability of the two properties was not equal. Here, each target object in the search phase of Experiment 4 had one property—either color or location—that was fixed and one property that was variable. The fixed property was the same on all trials, while the variable property had both a high- and low-frequency value. Like in Experiment 3, we tested to see what would happen when these different properties were in conflict with one another. In particular, we were interested in cases where the conflict was between the fixed property of one object and the high-frequency—but not fixed—property of a different object. For example, target object 1 was always white (fixed color) and appeared on the computer desk with high frequency but also appeared on the coffee table with lower frequency (variable location). Target object 2 was always brown (fixed color) and appeared on the coffee table with high frequency but also appeared on the computer desk with lower frequency (variable location). In the identification phase, for example, participants would be presented with a white object on the coffee table. Here, the fixed property is consistent with object 1 but the variable property is more consistent with object 2. Thus, the goal was to determine whether participants would show a preference for making an identification consistent with the property with greater reliability.

Method

Participants

A total of 33 Florida Atlantic University undergraduates with normal or corrected-to-normal vision participated in this experiment, which satisfied a course requirement.

Stimuli and procedure

The same objects and bedroom scene from the previous experiments were used. For the search phase stimuli, 160 total search scenes were created; half of the trials were target present and the other half of the trials were target absent. Either the location or the color of the target objects remained the same throughout the

duration of the search phase. The property that was not fixed was systematically varied so that for each object it had a high-frequency occurrence (70% of target-present trials) and a low-frequency occurrence (30% of target-present trials). Target object 1 was always white, with a high-frequency location on the computer desk and a low-frequency location on the coffee table. Target object 2 was always brown, with a high-frequency location on the coffee table and a low-frequency location on the computer desk. Note that target objects 1 and 2 had opposite high- and low-frequency locations. The fixed location for target object 3 was on the floor next to the guitar, with a high-frequency color of yellow and a low-frequency color of dark gray. Finally, target object 4 had a fixed nightstand location, with a high-frequency color of dark gray and a low-frequency color of yellow. Target objects 3 and 4 had opposite high- and low-frequency colors. The other four objects used in the previous experiments were used as distractors on both target-present trials and target-absent trials. They were colored with any of the four target colors and appeared at locations throughout the room except for locations at which target objects 1 through 4 were shown.

For the identification phase, six images were created for each of the four target objects for a total of 24 trials. Similar to the previous experiments, each image was blurred such that only the color and location of the test object was visible; no shape information that distinguished the target's identity was visible. Half of the trials showed the target object with its fixed property and its high-frequency variable property. The other half of trials were "conflict" trials in which the target was shown with the fixed property of one object and the low-frequency variable property of that same object, which was also the high-frequency variable property of a different object. Thus, one object with a fixed property and another object with a different high-frequency property were in conflict with each other on these trials.

The procedure was identical that that in the previous experiments. There were 160 trials in the search phase and 24 trials in the identification phase.

Results and discussion

Four participants were removed from further analysis because they scored close to chance accuracy in the search phase. The overall accuracy in the search phase was 83.09%, 95% CI [79.50, 86.69]. The overall reaction time for correct trials in the search phase was 2.38 s, 95% CI [2.27, 2.50].

In the identification phase, when the blurred target object with its high-frequency variable property was tested, participants chose the target object with the

fixed property and high-frequency variable property on 78.45% of trials ($SD = 26.14$), which was significantly better than chance, $t(28) = 5.86$, $p < 0.001$. The mean difference effect size was 28.44%, 95% CI [18.5, 38.39]. This showed that participants could use the most reliable property cues (location or color) gleaned from the search task in order to identify the degraded objects in the identification phase.

When we tested the objects with fixed locations with the high-frequency color of the other fixed-location object, participants identified the object based on the location information on 72.41% of the trials, the high-frequency color of the other object on 16.67% of the trials, and neither property on 10.92% of the trials. A Friedman's ANOVA revealed that there was a significant difference between the identification choices, $\chi^2(2) = 24.15$, $p < 0.001$. A follow-up planned comparison between fixed location and high-frequency color was also significant, $\chi^2(1) = 13.37$, $p < 0.001$. That is, participants used the more reliable information—location, in this case—to identify the degraded object.

When objects with fixed colors were tested in the high-frequency location of the other fixed-color object, participants identified the object based on color on 59.2% of the trials, the high-frequency location of the other object on 28.16% of the trials, and neither property on 12.64% of the trials. A Friedman's ANOVA revealed a significant difference between the identification choices, $\chi^2(2) = 17.59$, $p < 0.001$. A follow-up planned comparison between fixed color and high-frequency location was significantly different, $\chi^2(1) = 5.54$, $p = 0.019$. Again, participants used the more reliable property—color, in this case—to identify the degraded object.

It is interesting to note that participants in Experiment 4 were able to identify the degraded object by fixed location on one trial and then by fixed color on the next trial, switching properties to choose the more reliable source of information. This suggests that people operate flexibly in using different sources of information for object identification.

General discussion

The current results are consistent with previous experiments showing that contextual information can facilitate the identification of objects in degraded images (Bar & Ullman, 1996; Barenholtz, 2014). However, in these previous studies, there was no way to independently assess the roles of context and intrinsic object features for identification. The current study shows that context may be as important as intrinsic features in recognition. First, Experiment 1 showed that context can be fully responsible for determining the identity of a degraded object when the intrinsic image of the object is insufficient for identification. Previous demonstrations have suggested that context can drive identification of ambiguous objects (Bar, 2004; Barenholtz, 2014). However, in these previous studies, some intrinsic information was available for performing the identification task, even if the image was substantially degraded. In Experiment 1 here, however, the target object contained no intrinsic information that could be used to discriminate between the choices in the identification task. Thus, these results demonstrate that context does not merely facilitate object identification; it can serve as the sole basis of identification.

The variability of responses in Experiment 1 suggests that the tendency to choose on the basis of location was probably not due to an explicit guessing strategy based solely on location, which would predict consistent responses across trials for specific locations. Instead, participants seemed to have been incorporating additional factors in their identification decisions, most likely based on the false assumption that there was task-relevant intrinsic information. These results suggest that participants were trying to engage in something akin to natural recognition (i.e., where there is true object identity) rather than employing an explicit guessing strategy.

Experiment 1 also showed that knowing the likely location of an object within a three-dimensional context can improve visual search time in three-dimensional scenes, even under conditions of varying viewpoints, a form of contextual cueing (Chun & Jiang, 1998, 2003). Most studies of contextual cueing, including those involving natural scenes (Brockmole et al., 2006), have considered only targets with fixed two-dimensional coordinates, where participants could have used the screen location to perform the task. One exception is Chua and Chun (2003), who tested the effect of viewpoint variation on contextual cueing using stimuli consisting of an array of bowling pins and cylinders (i.e., artificial scene stimuli). They found a decline in contextual cueing with rotations past 15° from training displays. In the current study, we found that contextual cueing extends to much larger rotational viewpoint changes in a naturalistic three-dimensional scene.

Experiment 2 showed that an object's contextual and intrinsic feature information can be combined and used to identify a degraded object. This result is consistent with those of Barenholtz (2014), who found that participants needed less resolution in order to correctly identify objects shown within their original contextual scene compared with objects shown in isolation. This effect of context was greatly enhanced when the participants were already familiar with the scene and object in question. In that study, participants likely

used their schemas, memory, or both in order to reduce the set of objects that were likely to be present in a given location and then used the available intrinsic information to identify the object. The results of Experiment 2 in the current study may be interpreted similarly; participants used the location information to narrow down their choices and then further discriminated between the two remaining options on the basis of the color information. Overall, these findings are most supportive of previously proposed theories of contextual facilitation that are based on "criterion modulation" or "matching" models, whereby context reduces the amount of visual information required to trigger a match to a specific object (Friedman, 1979). This is in line with behavioral (Auckland, Cave, & Donnelly, 2007; Davenport, 2007) and electrophysiological (Mudrik, Lamy, & Deouell, 2010; Mudrik, Shalgi, Lamy, & Deouell, 2014) results that support a criterion modulation interpretation. For example, Mudrik et al. (2014) found a pronounced frontocentral event-related potential (ERP) negativity starting as early as ~210 ms after stimulus onset for scenes presented with semantically incongruent target objects compared with scenes with congruent objects. This early contextual congruity effect is consistent with the notion that a scene context exerts influence over object identification processing before complete identification is achieved. Conversely, the results of Experiment 2 of the current study are inconsistent with a strong form of the functional isolation model (Hollingworth & Henderson, 1998), as participants used both an intrinsic object feature (color) and context concurrently to identify the object in the identification phase. Thus, context is clearly not isolated from the processing of other features in object identification. However, while the current results clearly demonstrate that people combine contextual and intrinsic information in object identification, it is important to note that this conclusion does not bear directly on the underlying question at issue in many of these earlier studies, which were concerned with whether context can speed up recognition of fully recognizable images.

Experiment 3 found that when color and location information were in conflict, with one property statistically favoring one object and the other property favoring another object, there was no consistent a priori bias for one source of information over the other. Instead, as the results of Experiment 4 show, participants chose the more reliable property (whether color or location) and even switched from one source to another across trials.

Previous studies have suggested that contextual cueing results found within a specific local context (e.g., located on a pillow) can transfer to a different global context (e.g., transfer from a pillow in a bedroom to a pillow in a living room; Brockmole & Vo, 2010). In relation to the current study, if local location information transfers to a different global context, then we would expect identification behavior to be similar if the target object appeared within the same local context (e.g., a computer desk) in a different semantic scene (e.g., a living room). However, it is also possible that identification based on location would be lost or reduced in the context of a different global scene. Future research is needed to address this question.

Conclusions

The current study examined possible roles that visual context can have in the identification of objects. Degraded objects were used in order to simulate real-world conditions when viewing conditions are not ideal, as can frequently occur in everyday perception. We found that location information (a) can be fully responsible for determining an object's identity when the intrinsic features of the object are insufficient for identification, (b) can be combined with intrinsic object features to determine an object's identity, (c) can be equally utilized as intrinsic features of the object to identify an object, and (d) can be given priority over intrinsic object features when the location information is a more reliable cue for identification. Overall, the results of these experiments suggest that location information is treated as simply another feature in object recognition, on par with more familiar intrinsic features of objects such as shape or color.

Keywords: context effects, cue combination, degraded object recognition

Acknowledgments

Corresponding author: Derrick Schlangen.
E-mail: dschlang2@gmail.com.
Address: Psychology Department, Florida Atlantic University, Boca Raton, FL, USA.

References

Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin & Review, 14*(2), 332–337.

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

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

Barenholtz, E. (2014). Quantifying the role of context in visual object recognition. *Visual Cognition, 22*(1), 30–56.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143–177.

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 531–543, doi:0278-7393/92.

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.

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 699–706, doi:10.1037/0278-7393.32.4.699.

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition, 13*(1), 99–108, doi:10.1080/13506280500165188.

Brockmole, J. R., & Vo, M. L-H. (2010). Semantic memory for contextual regularities within and across scene categories: Evidence from eye movements. *Attention, Perception, & Psychophysics, 72*(7), 1803–1813, doi:10.3758/APP.72.7.1803.

Chua, K. P., & Chun, M. M. (2003). Implicit scene learning is viewpoint dependent. *Perception & Psychophysics, 65*(1), 72–80.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*(1), 28–71, doi:10.1006/cogp.1998.0681.

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(2), 224–234, doi:10.1037/0278-7393.29.2.224.

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. L. (2007). Consistency effects between objects in scenes. *Memory & 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.

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.

Greene, M. R. (2013). Statistics of high-level scene context. *Frontiers in Psychology, 4,* 1–31, doi:10.3389/fpsyg.2013.00777.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243–271.

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:0096-3445/98.

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia, 48*(2), 507–517.

Mudrik, L., Shalgi, S., Lamy, D., & Deouell, L. Y. (2014). Synchronous contextual irregularities affect early scene processing: Replication and extension. *Neuropsychologia, 56,* 447–458.

Palmer, S. E. (1975). Effects of contextual scenes on identification of objects. *Memory & Cognition, 3*(5), 519–526.