Journal of Experimental Psychology:
Human Perception and Performance
2007, Vol. 33, No. 1, 20–30


0096-1523/07/$12.00 DOI: 10.1037/0096-1523.33.1.20

# Object-Intrinsic Oddities Draw Early Saccades

Mark W. Becker
Lewis & Clark College

Harold Pashler
University of California, San Diego

Jeffrey Lubin
Sarnoff Laboratories

The authors investigated whether anomalous information in the periphery of a scene attracts saccades when the anomaly is not distinctive in its low-level visual properties. Subjects viewed color photographs for 8 s while their eye movements were monitored. Each subject saw 2 photographs of different scenes. One photograph was a control scene in which familiar objects appeared in their canonical form. In the other picture, objects were altered in a way that rendered them deviant without introducing any obvious changes in low-level visual saliency. In Experiment 1, these alterations involved rotating an object in an unnatural fashion (e.g., an inverted head on a portrait, a truck parked on its front end). In Experiment 2, colors were distributed over objects in a way that was either reasonable or anomalous (e.g., a green cup vs. a green hand). Subjects fixated the anomalous items earlier (both in time and in order of fixations) than the nondistorted objects, suggesting that violations of canonical form are detected peripherally and can affect the likelihood of fixating an item.

*Keywords:* eye movements, attention, gaze, scene viewing

When one passively surveys a scene, what kind of analysis does the visual system perform on the contents of the visual periphery to determine the destination of the next saccadic eye movement? Is the “meaning” (identity, significance) of a remote part of the scene processed even before the eye perches upon it? To put the question slightly differently, how extensively does the brain analyze objects that people see only “out of the corner of their eyes”?

Although these questions are intriguing as well as fundamental, research on the topic is scant and seemingly contradictory. A few points seem clear, however. First, it has been known for a long time that in passive viewing tasks, the eyes rarely fixate on blank walls or empty sky (Buswell, 1935; Yarbus, 1967). One cannot infer too much from this observation, however; it might simply reflect the fact that these empty regions lack elementary visual features like edges or vertices, or are deficient in high spatial frequency content. Second, it has been noted that when people are given the opportunity to inspect scenes containing out-of-place objects, the eyes will spend more time fixated on the deviant objects than on other objects (Henderson, Weeks, & Hollingworth, 1999; Loftus & Mackworth, 1978). However, this finding might reflect perceptual analysis that occurs after the first fixation on an object rather than before: Even if the visual system never detects the oddity of objects in the parafovea or periphery, once the eye lands on a deviant object it may linger there, or return to it later, on the basis of analysis that occurs while the object is foveated.

The most intriguing question, then, is whether objects that are odd in some way are fixated earlier than other objects. On this point, the answer seems to depend on the sort of oddity involved. There is little doubt that fixations are drawn to objects with discrepant low-level visual characteristics. However, if an item is discrepant, not because of its low-level features but because its identity is discrepant from the theme of the scene (e.g., an octopus in a farm scene), the evidence suggests that anomalous objects may not be fixated earlier than other objects. Thus, at the extremes it seems that low-level oddity is detected peripherally, whereas higher level semantic anomaly is not detected peripherally. Here we investigate whether an anomaly that requires an intermediate level of processing, one that is based on an object violating its stored perceptual properties (e.g., a green hand), is analyzed peripherally and thus draws early saccades.

## Background

There appears to be a general consensus that visual discrepancies at the featural level are detected by visual processing that is parallel across the visual field. One possible mechanism for this computation was described by Itti and Koch (Itti, 2000; Itti & Koch, 2000, 2001), who proposed that local, competitive interactions between visual neurons result in a neural signal that is biased in favor of visually discrepant features. Implemented in a computational model, this model has been shown to make reasonably accurate predictions of where a human observer will look within both static and dynamic scenes.

Will an object that is odd according to its higher level characteristics also attract early saccades? To our knowledge, this issue has been examined in only two studies, both investigating the semantic mismatch between an object and surrounding objects. We refer to this as *object-context oddity*, meaning that the object is

Mark W. Becker, Department of Psychology, Lewis & Clark College; Harold Pashler, Department of Psychology, University of California, San Diego; Jeffrey Lubin, Sarnoff Laboratories, Princeton, New Jersey.

This work was supported by National Science Foundation Grant SBR 9729778 and National Institute of Mental Health Grant R01-MH45584. We thank Alex Holcomb and Ed Hubbard for their comments on drafts of this article.

Correspondence concerning this article should be addressed to Mark W. Becker, Department of Psychology, Lewis & Clark College, 0615 Southwest Palatine Hill Road, Portland, OR 97219. E-mail: mbecker@lclark.edu

anomalous because it does not fit within the theme of the scene in which it is found.

In studies by Loftus and Mackworth (1978) and Henderson et al. (1999), subjects viewed line drawings of scenes while their eye movements were tracked. In a subset of the images, an object that was consistent with the semantic theme of the scene was replaced with an object that was inconsistent with the theme. For example, Loftus and Mackworth replaced a tractor in a farm scene with an octopus, and Henderson et al. replaced a cocktail glass in a bar scene with a microscope.

The two studies arrived at opposite conclusions concerning the effect of object-context oddity on eye movements. Loftus and Mackworth (1978) found that subjects tended to make long saccades ($M > 7°$) to deviant objects and to fixate these objects earlier within the viewing period. Thus, they concluded that semantic information is analyzed in the periphery (at least 7° from fixation) and the eyes are drawn to semantically anomalous objects. Henderson et al. (1999) found no such effect. In their experiment, the saccades to the object were on average short ($\sim 3°$), and the anomalous object was fixated no earlier than a control object in the same location. Thus, they concluded that the initial saccade to an object was unaffected by semantic analysis of peripheral information. Only after the object was first fixated did its anomalous nature affect eye movements, with observers being more likely to return to the location.

Why did these studies produce discrepant results? Several possibilities exist. On the one hand, Henderson et al. (1999) proposed that the semantically anomalous stimuli used by Loftus and Mackworth (1978) may also have been anomalous in their low-level properties, confounding their results. Henderson et al. pointed out some indications of low-level anomalies. For example, when the octopus replaced the tractor, the resulting scene featured a single object with curvy contours in a drawing composed largely of rectilinear, man-made structures (barns, fences, etc.). Henderson et al. suggested that it may have been these low-level visual discrepancies that attracted the long early saccades observed by Loftus and Mackworth, rather than the semantic anomalies.

Seeking to avoid confounding low-level visual discrepancies with semantic discrepancies, Henderson et al. (1999) inserted the same object in two different scenes in the same experiment. In one display, the object was in a "semantically consistent" context, and in the other display it was in a "semantically inconsistent" context. For example, a fire hydrant appeared in a street corner scene (nonanomalous) and a living room (anomalous). They reasoned that having the same object occur in both conditions would equate the effect of visual anomaly.

However, this control does not necessarily equate low-level salience. Low-level salience has long been recognized as reflecting primarily the differences between an object and its surround (Itti & Koch, 2000, 2001; Nothdurft, 1993, 2000; Titchener, 1908). The same visual features will be visually salient (or not) depending on the scene context in which they appear. Thus, using the same set of objects as semantically anomalous targets in one set of scenes and nonanomalous targets in another set of scenes does not ensure that low-level visual discrepancies are equated across conditions. It is possible that some targets were visually salient in semantically consistent scenes whereas other targets were visually salient within semantically anomalous scenes. Distributing the effects of these low-level visual discrepancies between conditions could potentially override the effects of semantic anomaly per se.

The specific findings of Henderson et al. (1999) may also be questioned because of their use of fairly crowded line drawings. Recognition of line drawings relies exclusively on high spatial frequency contour information, which is subject to especially severe loss in the periphery (Hilz & Cavonius, 1974). The use of cluttered displays also increases the contribution of lateral masking (crowding), which would be expected to further diminish the ability to identify the targets in the periphery (Bouma, 1970; Ehlers, 1936; Flom, Heath, & Takahashi, 1963). Thus, it is possible that the negative results of Henderson et al. might not generalize to real scenes for this reason.

Although it would be possible to follow up on the issue of object-context oddity, it is not clear that a definitive study is really possible. Absent a complete enumeration of elementary visual features, there is no way to ensure that introducing a new item in two scenes does not change low-level salience. Counterbalancing stimuli, as done by Henderson et al. (1999), does not solve this problem, for the reasons described above. Any experiment investigating object-relational oddities requires swapping out entire objects; thus, it seems extremely difficult to dissociate low-level visual discrepancy from semantic discrepancy.

The alternative strategy used here to examine the possible influence of higher level processing on saccadic control involves transforming one or several objects rather than substituting one object for another in a scene. The transformations are selected to be ones that introduce object-intrinsic oddity rather than object-context oddity. By *object-intrinsic oddity* we mean that we transform an item in such a way as to make the altered object violate the stored canonical form of the object (e.g., we transform the color of a hand to green).

This object-intrinsic manipulation also changes the visual features of the scene and thus might produce low-level visual discrepancies. Unlike the object-context manipulation, however, this method alters a single visual property of the scene. Thus, we can chose manipulations that, on the basis of previous research, appear unlikely to introduce low-level discrepancies (Experiment 1), or we can introduce the same manipulations in a control object where the manipulation introduces the same visual properties without producing object-intrinsic oddity (Experiment 2). In short, this method allows for more experimental control over the visual aspects that have been changed in the scene.

Finally, it is worth noting that switching from an object-context oddity to an object-intrinsic oddity may alter the level of processing required to detect the odd item. Detecting an oddity on the basis of object-context discrepancy is likely to occur later in the processing sequence than detecting an oddity on the basis of object-intrinsic oddity. The former requires the identification of both the odd item and the theme of the scene, activation of semantic information about the types of items that are consistent with the theme, and some mechanism to compare this semantic information with the odd object. By contrast, the object-intrinsic oddity might be detected at an intermediate level of processing; it requires only activation of the canonical representation of the object and a comparison of the structure of the visible object with that canonical representation. Thus, detection of this object-intrinsic oddity may represent an intermediate level of processing, higher along the processing stream than low-level visual properties but lower than the level of processing required to detect an object-context oddity.

## Present Experiments

We began with a rotation manipulation in Experiment 1, because rotating faces and vehicles out of their canonical orientation plainly produces results that are grossly anomalous in just about any scene. As Corballis (1988) pointed out, detecting that an object is in an unusual orientation must normally be a relatively late process, occurring after the object has been identified and compared with a stored memory of the canonical or proper orientation of the object. Consistent with this notion, prior research suggests that rotating an object within the picture plane does not result in preattentive pop-out (Enns, 1990) or provide for efficient selection in a partial-report task (Von Wright, 1968), implying that the low-level changes caused by rotation are not computed in parallel and without capacity limits across the visual field.

Wolfe (2001) also investigated how rotating objects might affect the allocation of attention. His subjects searched either for one upside-down ("dead") animal among upright ("live") distractors or for one live animal among dead ones. Notice that in both cases the target was rotated 180° relative to the distractors. If 180° of rotation produced low-level visual discrepancies capable of attracting attention to the target, subjects should have been equally fast in detecting both types of targets. Instead, Wolfe found a search asymmetry in which search for one dead animal among live distractors was more efficient than a search for a single live animal among dead animals. This asymmetry suggests that rotation alone did not result in low-level discrepancies that attract attention to a rotated item.[1] Instead, a rotated target rapidly attracted attention only when the rotation produced a target that violated the canonical representation for that item. This finding suggests that in the absence of low-level visual discrepancies, discrepancies based on a mismatch between the target's orientation and the canonical orientation of that object may be able to influence the allocation of attention.

Detection of this latter form of discrepancy is likely to rely on a relatively late stage of processing, after object identification allows access to the stored representation of the canonical form of the object. Indeed, Ballaz and colleagues found that this type of search asymmetry based on canonical orientation depends critically on the ability to identify the objects in the scene (Ballaz, Boutsen, Peyrin, Humphreys, & Marendaz, 2005). They found that the search asymmetry disappeared when object identification was made impossible by low-pass filtering the displays. In addition, Ballaz et al. (2005) showed that the search asymmetry did not occur for a patient with visual agnosia. Taken in sum, these studies suggest that rotation fails to produce low-level visual discrepancies capable of attracting attention to a target, but that higher level discrepancies based on a mismatch between an object and its canonical orientation may be able to draw attention.

Finally, it has been noted that rotating a picture of a face that is illuminated from the top provides shape-from-shading information that can lead to preattentive pop-out (Braun, 1993; Enns & Shore, 1997). To attempt to minimize this possibility, we used a portrait photograph that was lit from the front rather than from above.

Although it is logically possible that rotation could introduce orientation discrepancies between an object and its background, this seemed quite unlikely in the present situation (see Figure 1). To provide more assurance on this point, in the second experiment we used a different manipulation that was not subject to this potential limitation. Here, the transformation consisted of coloring one object in the scene green. When the object was one that should not have been green (a person's hand or a stop sign), this introduced an object-intrinsic anomaly. In control scenes, the objects that were colored green were ones that could reasonably be green (a coffee mug and a license plate). Thus, if the color green is either inherently salient or salient in the context of these scenes, then any early fixations on green hands and stop signs should also appear for green coffee mugs and license plates.[2]

There were several other notable precautions incorporated into our studies that were not featured in the earlier work in the area. One is that the deviant objects were pretested to be large enough to be readily resolved even with a very distant point of regard within the scene. Another is that our observers saw only one image that contained an anomaly. Both Henderson et al. (1999) and Loftus and Mackworth (1978) had subjects view an entire series of images, with half of the images containing something that was semantically anomalous. Given the abundance of anomalous items in these experiments, it would seem plausible that observers might begin searching for odd items, in which case the viewing behavior revealed would not necessarily reflect the habitual tendencies that arise when people view a scene with no particular expectation of encountering oddities.

## General Method

### Stimulus Materials and Design

Four color photographs served as stimulus materials. Each image subtended 33° × 24° of visual angle. Two of the photographs were of one scene (the front of a house) and two were of a different scene (a portrait of four people). One of the pictures in each pair was a control image, and the second picture (the altered picture) was identical to the control photo except that one object had been altered (using Adobe Photoshop) to make it deviant. Each subject saw only two pictures: one control picture and the altered version of the other picture. The presentation order of control and altered images was counterbalanced across subjects.

### Display

The images were shown on a 17-in. Sony Trinitron Multiscan 17SE II computer monitor set at a resolution of 1,024 × 768 pixels, with 24-bit color.

### Eye Tracking

Eye positions were monitored using an Eye Tech Digital Systems Quick Glance eye tracker running Eye Science SDK Version 3.1 software (Chap-

[1] Wolfe (2001) found that searching for one upside-down elephant among upright elephants was extremely efficient (5 ms per item). However, the distractors in these displays were extremely homogeneous, and thus, rotating a single item might have produced an abundance of low-level visual anomalies that might not occur with more heterogeneous displays. In fact, when Wolfe's subjects searched for a dead (unspecified) animal among a heterogeneous display of live animals, the slope of the search function increased to 28.6 ms.

[2] This color manipulation necessarily introduces a visual change that may or may not be visually discrepant. However, by changing the color of an existing object in the scene rather than adding a new item, we can be somewhat confident that the only visual properties that change are color and perhaps brightness. Thus, we can control for the effect of the salience of these visual attributes by changing another object in the scene to the same color without making that object semantically anomalous.

*Figure 1.* Approximate images used in Experiment 1 (with faces replaced for copyright reasons). All images were color photographs and subtended 33° × 24° of visual angle. At the beginning of each trial the subject was fixated on the center of the image. Each subject saw only two images, either the top left and bottom right photos or the top right and bottom left photos.

pell, 2000). The device has infrared lights mounted on each side of the computer monitor and a video camera mounted at the bottom of the monitor. The infrared lights illuminate the eye and provide a reference point for the eye-tracking video monitor. The eye tracker converted the eye position data to $x$ and $y$ screen coordinates. The eye tracker was set to obtain an eye position every 50 ms, and these $x$ and $y$ coordinates and the time of the reading were stored for later analysis.

### Procedure

Subjects were informed that the goal of the experiment was to determine how people normally view scenes, and that they would view photographs while their eye movements were monitored. Subjects anticipated an unspecified type of memory test following the viewing period. They sat with their chin in a chin rest while looking at the computer monitor. The room lights were dimmed to keep stray light from interfering with the eye tracker, and the experimenter positioned and adjusted the infrared lights of the eye tracker until the software that came with the eye tracker indicated that the tracking was "good." The experiment then began.

The first event was a calibration check. Subjects were instructed to look directly at a small fixation cross that appeared on the screen. The cross appeared in one quadrant of the screen, remained there for 2 s, and then moved to a different quadrant. After the cross had appeared in all four quadrants, a fixation cross appeared in the center of the screen and remained there for 1 s. The first photograph replaced the fixation cross and remained present for 8 s, followed by another 1-s fixation cross and then 8 s of viewing the second picture. All photographs subtended the entire screen. Following the second picture the calibration check was run once again.

### Data Analysis Routines

*Calibration data.* The data from the calibration portion of the experiment were used to identify and eliminate subjects for whom the eye tracker was not performing well. An analysis program (written in Macromedia Director) ignored the first 300 ms of calibration data following the appearance of each cross to allow time for the subjects to fixate on the new cross. It compared the measured coordinates of the remaining 1,700 ms of eye-tracking data with the actual position of the red crosses that the subject was fixating. A subject's data were discarded if more than 10% of the calibration measures, on either the original calibration or the calibration that occurred at the end of the experiment, were more than 1° from the center of the fixation cross.

*Eye position data.* Assuming the calibration data were satisfactory, the same program then identified fixations within the picture viewing periods. The algorithm defined the beginning of a fixation as two successive eye positions that were within 0.5° of one another. The algorithm continued to compare successive eye positions and defined the end of a fixation as occurring when the eye moved more than 0.5° between successive samples. The location of the fixation was defined as the average of all of the eye positions within a fixation. The program recorded the start time, duration, and fixation number of each fixation.

In addition, the program recorded whether a fixation landed on the target object. To determine this, we defined a rectangular region that just covered the target object. Fixations were deemed to be on the target if the mean fixation location was within 0.5° of this rectangle.

## Experiment 1: Images With Rotational Violations

### Method

*Subjects.* One hundred twenty-five subjects with normal or corrected-to-normal vision participated for course credit. After the calibration data were used to filter out subjects for whom the tracker may not have been working accurately, 94 subjects remained. This high attrition rate was due to the relatively poor accuracy of our eye tracker and the fact that our calibration criterion was higher than the criterion of the eye tracker's setup software. However, the inaccuracy of the eye tracker should not be systematic, and thus it should only add noise to our data, making real effects more difficult to detect (as do the numerous other sources of noise that arise in behavioral experiments, such as subject distraction, fingers slipping off response keys, etc.).

*Stimuli.* One of the images was a photographic portrait of four people (see Figure 1). The original image served as the control image, and in the altered image, one of the people's heads was rotated 180°. The face that was rotated subtended about 5.5° × 4° of visual angle. The second image was a picture of a house with a number of people standing around and a tow truck in front of the house. In the altered condition the tow truck was rotated 90°, so that it was sitting on its front grill rather than its tires. The truck subtended about 5° × 3.3° of visual angle.

### Results

Both Loftus and Mackworth (1978) and Henderson et al. (1999) found that odd items were fixated more often and for a greater total duration than "normal" control items. We also found (see Table 1) that people made more fixations to the rotated objects than their upright controls, face: $t(76) = 5.458$; truck: $t(78) = 5.878$ ($p < .001$ for both). This resulted in a greater total duration looking at the rotated items than their controls, face: $t(67) = 5.586$; truck: $t(83) = 5.712$ ($p < .001$ for both). This finding suggests that the modifications we made to the images were perceived as odd; however, our main interest was not in the total number of fixations on the odd objects but instead on the very first fixation on the objects.

The rotated objects were fixated substantially earlier than the same objects in their correct orientation. This was true whether one evaluated the time to first fixate the object (see Figure 2, top) or the ordinal fixation number of the first fixation on the object (see Figure 2, bottom). Independent-sample $t$ tests comparing the object in its normal orientation with its rotated orientation confirm these findings for both the time to first fixate the object, face: $t(93) = 4.10$; truck: $t(73) = 4.56$ ($p < .001$ for both), and the fixation number of the first fixation on the object, face: $t(72) = 6.31$; truck: $t(69) = 4.65$ ($p < .001$ for both).

Although the rotated objects were fixated earlier, the duration of the first fixations to the rotated objects was not significantly different, face: $t(93) = 0.15$; truck: $t(77) = 0.55$ ($p > .20$ for both), than the duration of fixations to the same objects in their correct orientation. People did, however, tend to make more fixations on the rotated objects before fixating on a new object, resulting in longer first gaze durations, face: $t(72) = 2.04$; truck: $t(88) = 2.32$ ($p < .05$ for both), on rotated than on correctly oriented objects. Furthermore, the length of the first saccade to the rotated objects was no longer, face: $t(84) = 0.40$; truck: $t(81) = 0.61$ ($p > .20$ for both), than the first saccades made to these objects when they appeared in their correct orientation. It is worth noting that these saccades were much larger than those reported by Henderson et al. (1999) and slightly larger than those reported by Loftus and Mackworth (1978). The relatively large image size and the fact that our images were not very cluttered may have contributed to people making larger saccades.

### Discussion

The main findings of this experiment were that observers fixated rotated objects earlier than objects that appeared in the common orientation; however, rotated objects were not fixated immediately, and the initial saccades to rotated images were about the same length as the initial saccades to upright objects in the same location. The finding that people fixated the anomalous items earlier in the pattern of fixations suggests that the mechanism responsible for choosing fixation locations is indeed affected by an object-intrinsic anomaly. However, the finding that the saccades to these anomalous items are not immediate and are no longer than saccades to nonanomalous objects suggests that this mechanism may analyze information over a limited area around fixation (Motter & Belky, 1998) or that this analysis may be subject to capacity limitations. Although the area of analysis clearly extends into the periphery at least some of the time (mean saccade length of ~8°), our data do not warrant concluding that the entire image is analyzed for semantic content prior to making a saccade. It is also worth noting that the mean fixation number on the face was earlier (regardless of orientation) than the mean fixation number on the truck. To foreshadow our more general conclusions discussed below, this may reflect the fact that items within the area analyzed by the mechanisms subserving oculomotor guidance compete for the next fixation, and this competition involves information about object-intrinsic anomalies together with other semantic and nonsemantic biases about where to look next (e.g., areas with high information content, areas close to the present fixation).

Though it is unlikely that the rotation of stimuli introduced discrepant low-level visual properties that in turn caused rotated items to be fixated early, it is difficult to completely rule out this

Table 1
*Means and Standard Deviations for Experiment 1*

| Variable | Face | | | | Truck | | | |
|---|---|---|---|---|---|---|---|---|
| | Upright | | Inverted | | Upright | | Inverted | |
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Total no. of fixations on item | 4.56 | 2.36 | 8.04** | 3.55 | 2.09 | 1.17 | 4.02** | 1.94 |
| Total duration fixating item (ms) | 1,393.31 | 660.59 | 2,558.24** | 1,257.24 | 616.22 | 409.95 | 1,226.64** | 610.62 |
| Start time of first fixation on item (ms) | 1,715.81 | 956.64 | 932.45** | 907.53 | 3,767.43 | 2,204.72 | 2,065.42** | 1,252.70 |
| No. of first fixation | 5.92 | 2.59 | 3.23** | 1.39 | 11.43 | 6.72 | 6.29** | 3.51 |
| Length of first saccade to item (degrees of visual angle) | 8.45 | 2.06 | 8.24 | 2.80 | 9.89 | 5.01 | 9.28 | 4.29 |
| Duration of first fixation on item (ms) | 319.46 | 220.12 | 326.15 | 212.94 | 275.86 | 154.66 | 291.96 | 120.62 |
| Duration of first gaze on object (ms) | 536.60 | 405.93 | 783.51** | 724.95 | 526.52 | 360.14 | 723.73** | 443.48 |

** $p < .01$.

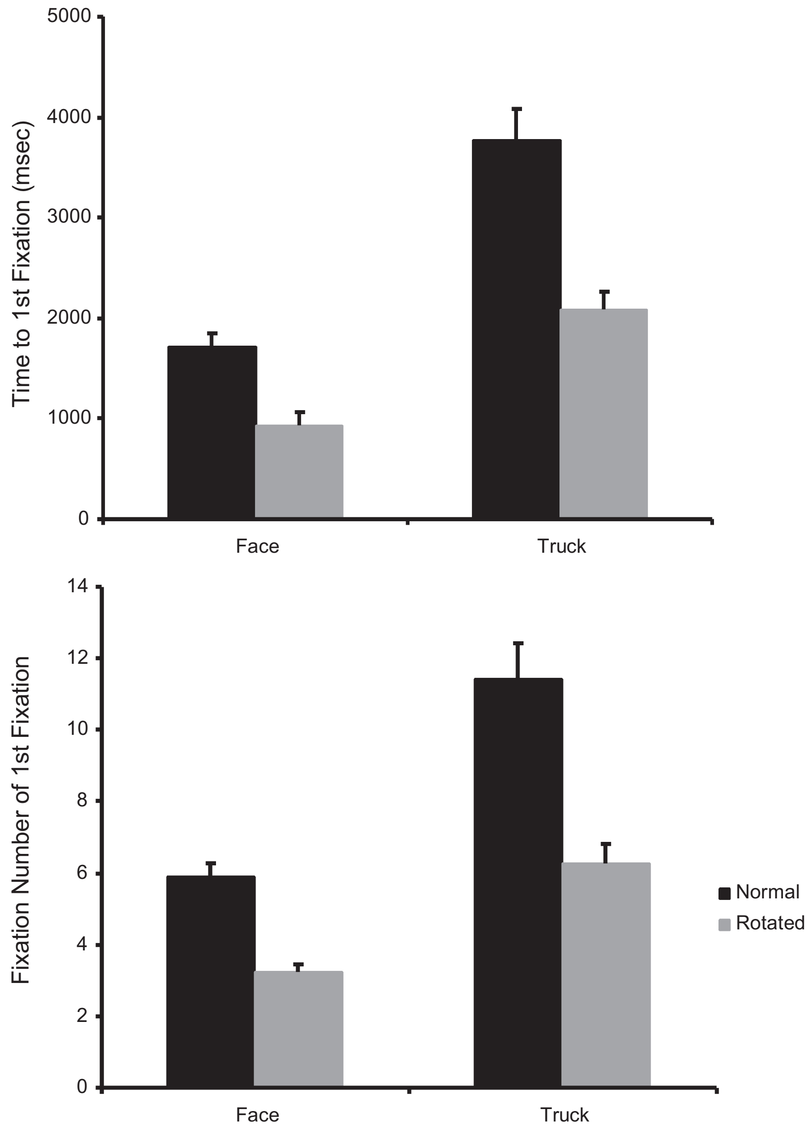


*Figure 2.* Objects are fixated earlier when they appear in an anomalous orientation. This is true whether one measures the time needed to first fixate the object (top) or one counts the number of fixations made prior to fixating the item (bottom). Error bars represent the standard error of the mean.

possibility. As a further check on whether our manipulations introduce gross low-level anomalies, we used Itti and Koch's (2000; Itti, Koch, & Niebur, 1998) computational model to derive saliency maps of each of our images.[3] These saliency maps are gray-scaled images in which brightness indicates each point's saliency value. We then calculated the mean brightness values for the areas that corresponded to the manipulated objects. Using this method we found that the upright face ($M$ = 89.01) was slightly more visually salient than the inverted face ($M$ = 84.43). The visual saliency of the truck was virtually unaffected by the rotation manipulation ($M$ = 118.59 upright, $M$ = 118.79 inverted). These values suggest that our rotations did not produce a radical shift in the low-level visual saliency of the objects. Even so, in Experiment 2 we tried another kind of visual distortion to provide a converging test for the idea that object-intrinsic anomalies rather than low-level visual properties draw earlier fixations.

## Experiment 2: Semantic Color Discrepancies (or the Lure of Green Flesh)

### Method

*Subjects.* One hundred twenty-eight subjects with normal or corrected-to-normal vision participated for course credit. After the calibration data were used to filter out subjects for whom the tracker was not working accurately enough, 81 subjects remained.

*Stimuli.* In one picture (see Figure 3) a person's hand (subtending about 5° × 4° of visual angle) was changed from flesh color to green. In the other a stop sign (subtending about 3° × 3° of visual angle) was changed from red to green. Both of these alterations create object-intrinsic oddity. However, they

[3] We are grateful to the Itti Lab for allowing us to use the iLab Neuromorphic Vision C++ Toolkit's bottom-up, saliency-based visual attention algorithm.

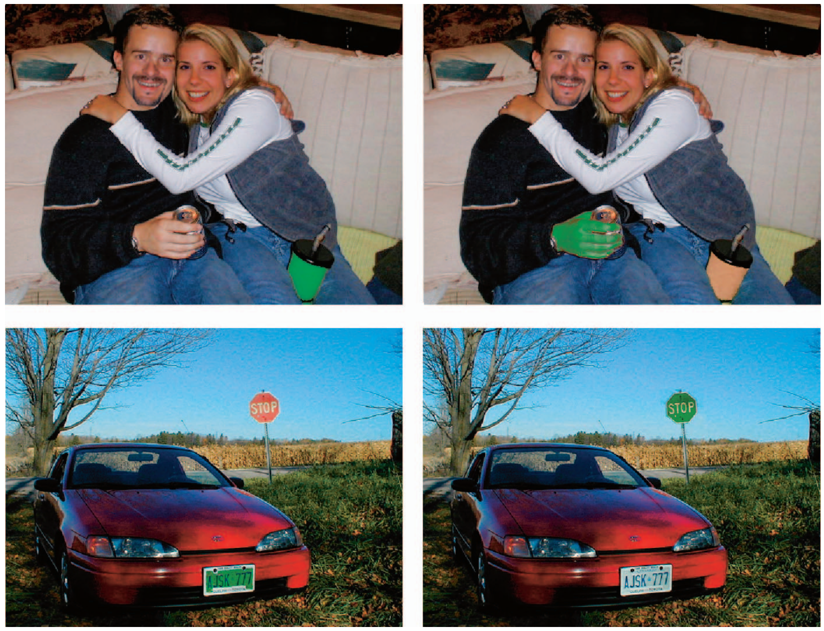


*Figure 3.* Approximate images used in Experiment 2 (with faces replaced for copyright reasons). All images were color photographs and subtended 33° × 24° of visual angle. At the beginning of each trial the subject was fixated on the center of the image. Each subject saw only two images, either the top left and bottom right photos or the top right and bottom left photos.

also introduce a low-level feature into the display (namely the green color itself). To be sure that this shade of green was not somehow attention grabbing in and of itself, we introduced the same green into the control images. Here, this alteration changed a neutral object to the same color green as the green in the experimental manipulation. The neutral objects were a large (subtending about 5° × 3.2° of visual angle) coffee mug in the photo of people and a license plate (subtending about 2° × 4° of visual angle) in the image of a car and stop sign. These objects are described as "neutral" to indicate that although they were not originally green, they could perfectly well have been that color. This also controls for the possibility that any alteration of the photo inadvertently introduced some very subtle low-level cue.

## Results and Discussion

As seen in Table 2, observers tended to make more fixations on the green hand, $t(79) = 3.374$, $p < .001$, and to fixate on it for a longer total duration, $t(77) = 3.449$, $p < .001$, than the flesh-colored hand. In addition, the hand was fixated earlier when it was green than when it was flesh colored. This was true whether one looked at the time to first fixate the hand, $t(77) = 2.2$, $p < .05$ (see Figure 4) or the fixation number of the first fixation on the hand, $t(69) = 2.67$, $p < .05$. It is our contention that the green hand was fixated earlier because hands do not normally appear green and thus the green hand is perceived as anomalous. However, it is possible that in this particular scene, any green object would have been visually anomalous and thus fixated earlier, more often, and for a greater duration.

To assess this possibility, we compared fixations on the coffee cup when it was green with fixations of the cup when it was not green. When the coffee cup appeared green it was fixated no earlier than when it was not. This was true whether one investi-

Table 2
*Means and Standard Deviations for the Image With the Green Hand and the Control Image With the Green Cup*

| | Hand | | | | Cup | | | |
|---|---|---|---|---|---|---|---|---|
| | Normal | | Green | | Normal | | Green | |
| Variable | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Total no. of fixations on item | 1.95 | 1.87 | 3.35** | 1.86 | 1.90 | 1.35 | 1.60 | 0.93 |
| Total duration fixating item (ms) | 482.98 | 682.81 | 960.68** | 558.94 | 360.22 | 365.76 | 495.85 | 421.15 |
| Start time of first fixation on item (ms) | 4,600.93 | 3,057.55 | 3,269.70* | 2,405.39 | 4,898.40 | 2,257.66 | 4,677.50 | 2,652.99 |
| No. of first fixation | 12.79 | 7.99 | 8.60* | 4.92 | 13.30 | 5.72 | 13.19 | 7.04 |
| Length of first saccade to item (degrees of visual angle) | 7.59 | 5.37 | 7.90 | 5.13 | 10.95 | 5.73 | 9.52 | 3.91 |
| Duration of first fixation on item (ms) | 276.92 | 111.56 | 311.14 | 118.88 | 321.00 | 126.45 | 298.14 | 137.78 |
| Duration of first gaze on object (ms) | 388.46 | 205.09 | 680.19** | 342.15 | 502.77 | 302.21 | 401.71 | 211.02 |

* $p < .05$. ** $p < .01$.

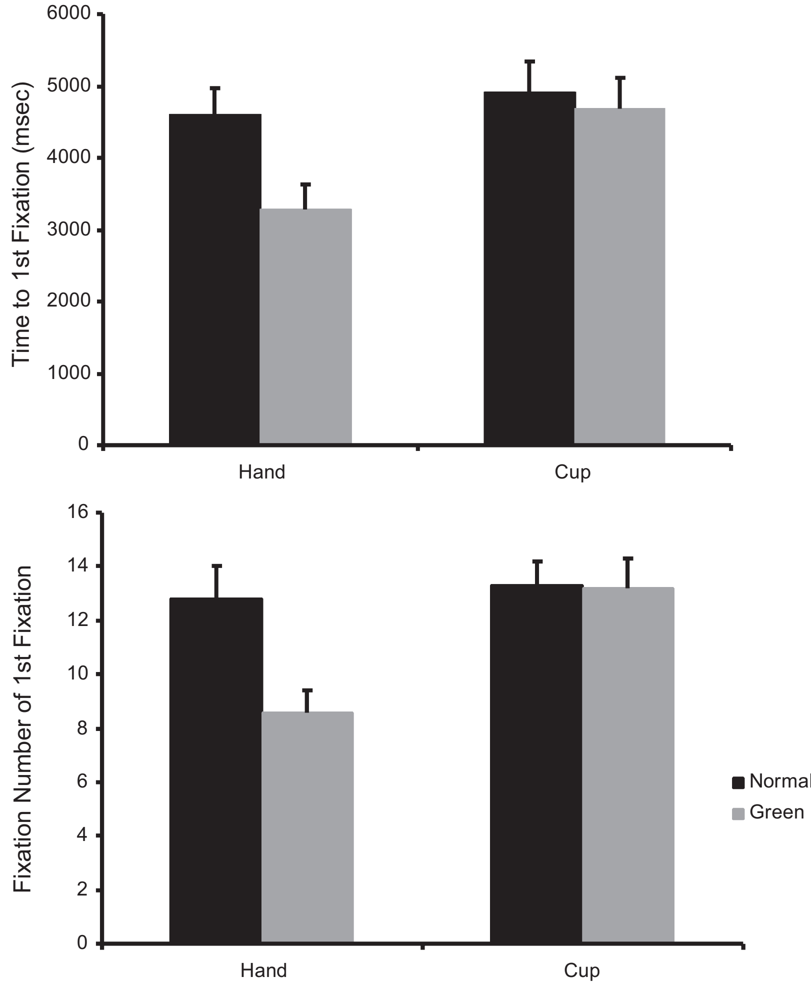


*Figure 4.* A green hand is fixated earlier than a flesh-colored hand. This is true whether one measures the time needed to first fixate the object (top) or one counts the number of fixations made before fixating the object (bottom). A cup that is not anomalous when it is green is fixated no earlier when it is green than when it is flesh colored. Error bars represent the standard error of the mean.

gated the time to first fixate (Figure 4) the cup, $t(79) = 0.41$, $p > .20$, or the fixation number of the first fixation (Figure 4) on the cup, $t(78) = 0.08$, $p > .20$. In addition the total number of fixations did not differ with the cup's color, $t(69) = 1.219$, $p > .20$, nor did the total time fixating on the cup, $t(77) = 1.546$, $p > .10$. These data are consistent with the view that the cup was not seen as anomalous when green and thus was not viewed earlier, more often, or for more total time.

Finally, we again used saliency maps derived from Itti and Koch's (2000) low-level computational model to calculate the saliency value for the area of the hand and the cup in each image. When the hand was green ($M = 88.94$), it was less visually distinct than when it appeared flesh colored ($M = 104.36$). Similarly, the green cup ($M = 74.36$) was less visually distinct than a flesh-colored cup ($M = 80.96$). Thus, a model based solely on this construal of low-level visual saliency should have predicted the flesh-colored hand to be fixated earlier than the green hand, but we found the opposite, suggesting that the object-intrinsic anomaly was indeed responsible for the early saccades to the green hand.

In the image containing the stop sign (see Table 3), the stop sign was fixated earlier (Figure 5) when it appeared green than when it appeared red, $t(79) = 1.89$, $p < .05$, and the fixation number of the first fixation on the green stop sign was marginally earlier than the fixation number of the red stop sign, $t(80) = 1.65$, $p = .051$. However, we also found that the time to first fixate the license plate was shorter when it appeared green than when it appeared white, $t(80) = 1.87$, $p < .05$, and the fixation number of the first fixation was earlier in the sequence of fixations when the license plate appeared green, $t(80) = 2.1$, $p < .05$. Thus, one possible interpretation of this result (but not the coffee cup result) is that any green object in this particular image is fixated earlier because of its low-level characteristics relative to the scene.

This interpretation is consistent with the data from a low-level saliency map of the images. The green license plate was more salient ($M = 135$) than the white license plate ($M = 125$), and the green stop sign was more salient ($M = 157$) than the red stop sign ($M = 87$). Thus, although we found the desired effect that the green stop sign was fixated earlier, it is possible that this effect was

Table 3
*Means and Standard Deviations for the Image With the Green Stop Sign and the Control Image With the Green License Plate*

| Variable | Stop sign | | | | License plate | | | |
|---|---|---|---|---|---|---|---|---|
| | Normal | | Green | | Normal | | Green | |
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Total no. of fixations on item | 1.73 | 0.15 | 2.12* | 1.21 | 2.03 | 1.37 | 2.55 | 1.68 |
| Total duration fixating item (ms) | 494.75 | 404.42 | 851.34** | 589.82 | 838.63 | 960.69 | 976.48 | 707.31 |
| Start time of first fixation on item (ms) | 3,883.85 | 2,367.08 | 2,912.24* | 2,281.58 | 4,008.07 | 2,847.88 | 2,842.65* | 2,787.76 |
| No. of first fixation | 10.23 | 5.93 | 8.00* | 6.26 | 10.83 | 7.43 | 7.55* | 6.68 |
| Length of first saccade to item (degrees of visual angle) | 8.25 | 4.68 | 9.97 | 4.23 | 4.07 | 3.02 | 4.36 | 3.03 |
| Duration of first fixation on item (ms) | 320.26 | 111.56 | 402.72 | 118.89 | 457.69 | 314.34 | 483.62 | 404.60 |
| Duration of first gaze on object (ms) | 365.85 | 233.96 | 556.53** | 317.05 | 694.71 | 589.40 | 665.97 | 482.39 |

* $p < .05$. ** $p < .01$.

due, at least in part, to low-level visual anomalies rather than higher level object-intrinsic anomalies. These saliency values limit the conclusions that can be drawn from this image. This finding, however, highlights that this method of control is able to distinguish between possible effects due to low-level visual characteristics and those stemming from higher level object-intrinsic oddity, thereby giving us additional confidence that the results from the other three images are not due to low-level factors.

For both sets of scenes, the duration of the first fixations to the target and control objects was no different (see Table 3) when the objects appeared green or in their normal color (all *p*s > .05). In addition, the length of the first saccade to target objects was not affected by the color of the target item (all *p*s > .05).

## General Discussion

Our results demonstrate that the first fixation on an object altered in a way that deviates from the normal appearance of that object occurs substantially earlier than the first fixation on the same location when it contains the same object without any comparable alteration. The length of the initial saccade to these anomalous items was about 8° of visual angle. Taken together, these two findings suggest that substantial processing occurs in the visual periphery (at least when acuity allows it and scenes are relatively simple, as was the case in our displays). Furthermore, when this peripheral analysis reveals object-intrinsic anomalies, it draws early fixations. These results were replicated for all four of our images across two different types of image manipulation (rotation and color alteration).

Although it is tempting to conclude that our results agree with those of Loftus and Mackworth (1978) and contradict those of Henderson et al. (1999), it should be noted that our object-intrinsic anomalies are noticeably different from the object-context anomalies examined in the earlier studies. In both the Henderson et al. and the Loftus and Mackworth studies, the anomalous items were objects whose presence would be improbable given the other objects present in the scene (e.g., a motorcycle in an opera house). In the present experiments, objects were transformed to assume a character that would be improbable in any context (with one notable exception: horror movies). The green hand, for instance, was anomalous because human hands are not normally green. The detection of this type of anomaly would seem to be something that can occur only at relatively late stages of the object-recognition process; to identify the green hand as anomalous, the observer must both identify the object as a hand and detect that the color differs from the canonical properties associated with that object category. This is a potentially important distinction between the present work and the earlier studies. One could imagine the possibility that object-intrinsic anomaly might draw early saccades even if object-context semantic anomaly does not.

Consistent with this view, neuropsychological research suggests that the abilities to detect object-intrinsic violations and object-context violations may operate independently. For instance, Riddoch and Humphreys (1987) presented a case study of patient J.B., who appeared to be able to detect object-intrinsic oddities but was unable to identify object-context oddities. J.B. could identify non-objects that were formed by interchanging parts of real objects (e.g., replacing a kangaroo's tail with a foot), leading Riddoch and Humphreys to conclude that J.B. could access stored knowledge about the structure of particular objects. However, when shown pictures of a hammer, chisel, and screw, J.B. could not choose which two pictures should be used together. Thus, it seems that J.B. lacked the ability to retrieve the semantic information required to detect meaningful relationships between objects. It appears, therefore, that J.B. was sensitive to object-intrinsic anomalies without being sensitive to object-context anomalies.

Although our data suggest that object-intrinsic anomaly is processed in the periphery, it is of interest that anomalous items are not fixated immediately and that saccades to these anomalous items are no greater in length than saccades made to control objects. These two findings are open to various interpretations. It might be that the semantic analysis that drives the anomaly effects observed in this study is restricted to some limited spatial region. This could occur because perceptual acuity in the far periphery limits the ability to resolve the object with enough detail to allow object identification. As mentioned in the introduction, there is reason to believe that the ability to detect the object-intrinsic anomalies we used probably relies on the ability to identify the object (Ballaz et al., 2005) and access the stored perceptual representation for this type of object (Corballis, 1988). In the far periphery, object identification may be impossible. However, it is worth noting that, although limited, our data suggest that this analysis of object-intrinsic oddity is rather expansive (extending at least 8° into the periphery, at least some of the time).

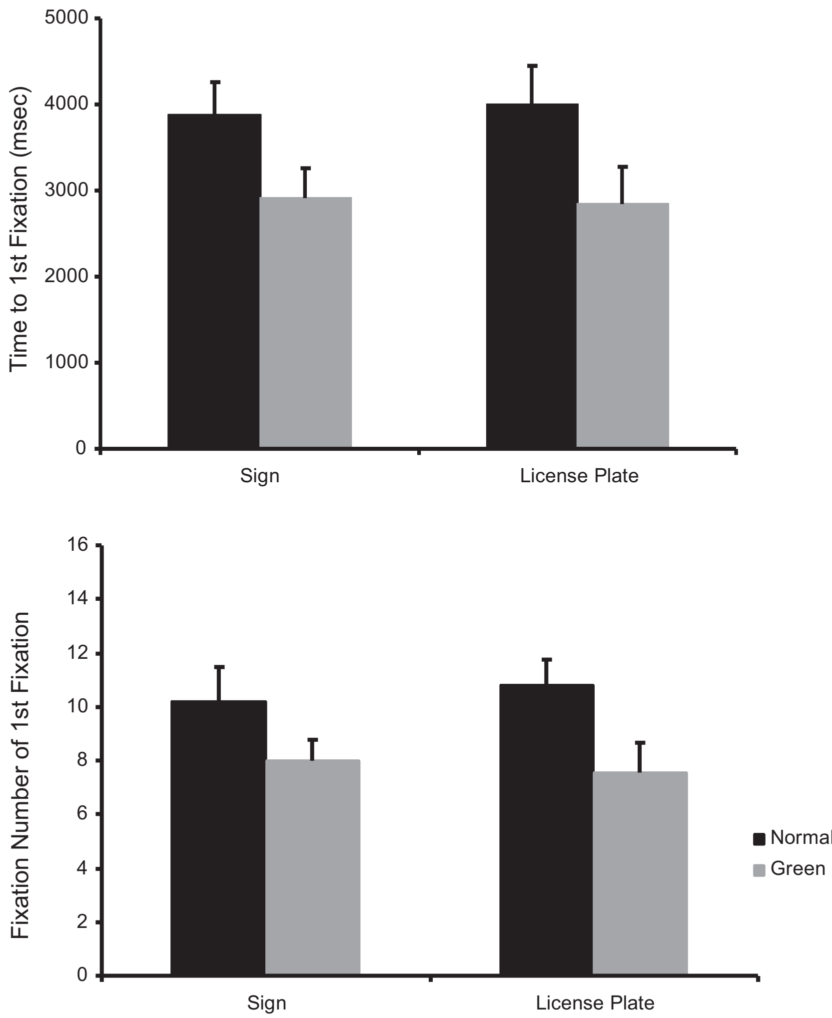


*Figure 5.* A green stop sign is fixated earlier than a red one, in terms of both the time needed to first fixate the item (top) and the fixation number of the first fixation (bottom). A green license plate is also fixated earlier than a white license plate. This may be because a green license plate is also perceived as anomalous in California (subjects were run in California where most plates are white). Error bars represent the standard error of the mean.

It should also be kept in mind that semantic anomaly is likely to be only one of many factors entering into saccade planning. Itti and Koch (2000; Itti, Koch, & Niebur, 1998) suggest that low-level featural discrepancies seem to draw saccades more powerfully when they are near to the current fixation; the same might be true of high-level anomalies. Thus, the influence that anomaly detection has on saccadic programming may decline as the distance between the object and the current fixation increases, resulting in limited saccade distances to anomalous items.

In either case, our results suggest that the appearance of an object-intrinsic anomaly in the periphery (up to at least 8° from fixation) can influence the competition for the next fixation. When the anomaly occurs sufficiently close to the current fixation, the probability that the next saccade will be to the anomalous item increases, and thus the object is fixated earlier in time (and earlier in the sequence of fixations) than a comparable nonanomalous object.

The hypotheses offered in the preceding paragraph could be incorporated within a modified saliency-competition model of saccade selection. In such an account, it would be assumed that the visual system generates a global saliency map of the visual scene, with the most salient location competing most effectively to attract the next fixation (Itti, 2000; Itti & Koch, 2000; Koch & Ullman, 1985; Kusunoki, Gottlieb, & Goldberg, 2000; Wolfe, 1994). Strength within the saliency map has been assumed to reflect activity at that location in a number of retinotopically organized cortical areas, each specialized for the analysis of particular visual features (e.g., V4 for color and middle temporal for motion). Models of this sort have only recently begun to postulate a role for semantic factors in this competition. For instance, some investigators have recently begun to add components to these computation models that take into consideration probable locations for relevant information given the scene context (Navalpakkam & Itti, 2005; Oliva, Torralba, Castelhano, & Henderson, 2003). Our results suggest that object-intrinsic oddity may alter the activity associated with a particular location within the saliency map.

Our results are limited to showing that an *object-intrinsic* oddity affects weightings within the saliency map. However, it is possible that other types of semantic information may also affect these weightings. For instance, saliency weights might be elevated for objects of special interest (e.g., faces), for emotionally charged or disturbing objects, for objects that have a momentary importance for the observer's current task, or for things related to the observer's ongoing thoughts (Moores, Laiti, & Chelazzi, 2003; Pashler & Shiu, 1999). Although the suggestion that these other factors may affect saliency weightings is speculative, allowing them to affect the saliency weighting of items might help these models better account for the distribution of fixations. For example, faces might be a common target of fixations not only because of their low-level visual properties but also because the semantic recognition that an object is a face may increase the weight assigned to that object within the saliency map. It is also, of course, possible that object-context anomalies increase the weight, although given the problems with the literature on this question (discussed in the introduction), we do not attempt to speculate on whether that is so.

## References


Ballaz, C., Boutsen, L., Peyrin, C., Humphreys, G. W., & Marendaz, C. (2005). Visual search for object orientation can be modulated by canonical orientation. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 20–39.

Bouma, M. (1970, April 11). Interaction effects in parafoveal letter recognition. *Nature, 226,* 177–178.

Braun, J. (1993). Shape-from-shading is independent of visual attention and may be a "texton." *Spatial Vision, 7,* 311–322.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art.* Chicago: University of Chicago Press.

Chappell, R. (2000). Eye Science SDK (Version 3.1) [Computer software]. Mesa, AZ: EyeTech Digital Systems.

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review, 95,* 115–123.

Ehlers, H. (1936). The movements of the eyes during reading. *Acta Ophthalmologica KjoBenhavn, 14,* 56–63.

Enns, J. T. (1990). Three-dimensional features that pop out in visual search. In D. Brogan (Ed.), *Visual search* (pp. 37–45). Philadelphia: Taylor & Francis.

Enns, J. T., & Shore, D. I. (1997). Separate influences of orientation and lighting in the inverted-face effect. *Perception & Psychophysics, 59,* 23–31.

Flom, M. C., Heath, G. G., & Takahashi, E. (1963, November 15). Contour interaction and visual resolution: Contralateral effects. *Science, 142,* 979–980.

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 210–228.

Hilz, R., & Cavonius, C. R. (1974). Functional organization of the peripheral retina: Sensitivity to periodic stimuli. *Vision Research, 14,* 1333–1337.

Itti, L. (2000). *Models of bottom-up and top-down visual attention.* Unpublished doctoral thesis, California Institute of Technology, Pasadena, California.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506.

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

Kusunoki, M., Gottlieb, J., & Goldberg, M. E. (2000). The lateral intraparietal area as a salience map: The representation of abrupt onset, stimulus motion, and task relevance. *Vision Research, 40,* 1459–1468.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 565–572.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6,* 182–189.

Motter, B. C., & Belky, E. J. (1998). The zone of focal attention during active visual search. *Vision Research, 38,* 1007–1022.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45,* 205–231.

Nothdurft, H. C. (1993). Saliency effects across dimensions in visual search. *Vision Research, 33,* 839–844.

Nothdurft, H. C. (2000). Salience from feature contrast: Additivity across dimensions. *Vision Research, 40,* 1183–1201.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. *Proceedings of the IEEE International Conference on Image Processing, 1,* 253–256.

Pashler, H., & Shiu, L. (1999). Do images involuntarily trigger search? A test of Pillsbury's hypothesis. *Psychonomic Bulletin & Review, 6,* 445–448.

Riddoch, M. J., & Humphreys, G. W. (1987). Visual object processing in optic aphasia: A case of semantic access agnosia. *Cognitive Neuropsychology, 4,* 131–185.

Titchener, E. B. (1908). *Lectures on the elementary psychology of feeling and attention.* New York: Macmillan.

Von Wright, J. M. (1968). Selection in visual immediate memory. *Quarterly Journal of Experimental Psychology, 20,* 62–68.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics, 63,* 381–389.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum Press.



Received July 18, 2005
Revision received June 13, 2006
Accepted June 15, 2006 ■