Scene Inversion Slows the Rejection of False Positives through Saccade Exploration During Search

Kathryn Koehler (koehler@psych.ucsb.edu)
Department of Psychological and Brain Sciences
University of California, Santa Barbara
Santa Barbara, CA 93106-9660

Miguel P. Eckstein (eckstein@psych.ucsb.edu)
Department of Psychological and Brain Sciences
University of California, Santa Barbara
Santa Barbara, CA 93106-9660

Abstract

The effect of face inversion has been heavily studied, whereas fewer studies have investigated inversion in scenes. We investigated the influence of scene inversion on decisions and contextual guidance of eye movements during visual search. A saccade contingent display termination paradigm was used to assess the temporal dynamics of the effect. Observers searched for a computer mouse in office scenes and performed a yes/no detection task. Observers' sensitivity (d') was lower for inverted images relative to upright. Observers' false positive rate decreased with additional eye movements when they viewed upright images, but remained constant during the first three eye movements when viewing inverted images. The average distance of observers' eye movements to the target location was greater for inverted than upright scenes. We interpret that inverting an image disrupts the rapid extraction of scene gist, subsequently disrupting guidance in eye movement behavior and slowing the process of rejecting false positives.

Keywords: scene context; contextual guidance; eye movements; visual search; scene understanding; scene inversion; scene gist

Introduction

A complete effort to investigate human scene understanding should include assessment of the effects of impoverished scene information on behavior. Such endeavors help us identify the processes that contribute to scene understanding, and the conditions in which they break down. Given our lack of experience in navigating an upside-down world, one such instance of impoverished information arises when scenes are inverted. Compared to the amount of work dedicated to understanding the effects of face inversion (Farah, Tanaka, & Drain, 1995; Valentine, 1988; Yovel & Kanwisher, 2005), scene inversion has been relatively less studied. Identifying the resulting impact of scene inversion may allow us to draw conclusions about what processes mediate scene understanding, in the same way that face inversion research has served as evidence for holistic processing of faces (Tanaka & Farah, 1993) and as evidence for (Kanwisher, Tong, & Nakayama, 1998) and against (G. A. Rousselet, Macé, & Fabre-Thorpe, 2003) a dedicated face perception module in the brain.

Existing work using inverted scenes has heavily focused on change blindness (Kelley, Chun, & Chua, 2003; Shore & Klein, 2000) and generally assumed that inversion affects the extraction of meaning or context from a scene (Brockmole & Henderson, 2006; Kelley et al., 2003) based on evidence of the perceptual deficits caused by orientation changes of stimuli (Klein, 1982; Rock, 1974). Other work has similarly shown the negative impact of scene inversion on the categorization of scene type (Walther, Caddigan, Fei-Fei, & Beck, 2009), a result that suggests scene gist may be affected by inversion.

There has been less work assessing the effects of scene inversion on decisions and eye movements during search. In particular, scene gist, context, and information about objects that co-occur with a target are rapidly extracted (Greene & Oliva, 2009; Henderson & Hollingworth, 1999; G. Rousselet, Joubert, & Fabre-Thorpe, 2005), guide eye movements (Castelhano & Heaven, 2011; Eckstein, Drescher, & Shimozaki, 2006; Mack & Eckstein, 2011; Neider & Zelinsky, 2006; Oliva & Torralba, 2007; Preston, Guo, Das, Giesbrecht, & Eckstein, 2013; Torralba, Oliva, Castelhano, & Henderson, 2006) and facilitate behavioral decisions (Castelhano & Heaven, 2011; Eckstein et al., 2006; Mack & Eckstein, 2011; Neider & Zelinsky, 2006).

The current work focused on understanding the effect of scene inversion on behavioral performance and eye movement guidance during a visual search task. We assessed how this effect unfolds temporally by utilizing a viewing paradigm that terminates scene presentation based on the number of saccadic eye movements executed by the observer. This paradigm is similar to that used by Hsiao and Cottrell (2008) to investigate the number of fixations required to recognize a face. We are particularly interested in how scene inversion disrupts the extraction of scene context and guidance of eye movements. By including trials in which there was no target present, but that contained scene cues (i.e., other objects predictive of the target location), we were able to evaluate the effect of scene inversion on eye movement guidance by scene context in isolation from guidance by target information.

Methods

Participants

Eye-tracking and behavioral response data were collected from 48 undergraduates (ages 18-23) at the University of California, Santa Barbara with normal or corrected-to-normal vision who received course credit for participation. Informed written consent was collected from all participants.

Stimuli and Design

A total of 80 greyscale photos of office and home-office scenes were shown to each participant. Image sizes varied in height from 12.9° to 23.7° and width from 13° to 24.7°. Half of the images contained a computer mouse, used as the target for the search task, and half did not. Images contained a computer monitor or laptop in both mouse present and absent images. Sample images are shown in Figure 1.

Latin-square counterbalancing was used to assign participants to conditions, thus determining whether a given image would be shown to the participant upright or inverted and how many fixations they would be allotted during image viewing. The experiment was therefore a 2 (image orientation; upright or inverted) × 4 (fixation allowance; 1, 2, or 3 fixations, or 3 second allowance) repeated measures design. Presentation of conditions was not blocked, i.e., the order of image presentation was randomized.

Apparatus

Stimuli were displayed on a 1024 × 768 pixel resolution LCD Barco MDRC-1119 monitor, calibrated to native settings, with each pixel subtending 0.037° of visual angle. Eye tracking data were recorded using a tower-mounted Eyelink 1000 system (SR Research Ltd., Mississauga, Ontario, Canada) monitoring gaze position at 250 Hz. Fixations were calibrated and validated using a nine-point grid system. Initial fixation was controlled on every trial and monitored to ensure error never exceeded greater than 1°. Recalibration was performed in the case of large head or body movements. Saccades were classified as events where eye velocity was greater than 22°/s and eye acceleration exceeded 4000°/s^2.

Procedure

Observers were instructed to determine whether a computer mouse was present in a series of images. They were to make a simple yes/no decision, and were told there was a 50% likelihood of target presence in each image. On each trial, observers began with an initial fixation outside of the boundaries of where the image was to appear, 1° above the bottom of the screen and 1.4° to 6° degrees from the border of the image, centered horizontally. Once observers initiated a trial with a key press, they were required to maintain initial fixation for 500-1500 msec before the image would appear. They were told that the image would appear for a variable amount of time, during which they could search for the mouse. Once a fixation was made inside of the image boundaries on the screen, the image was removed after one, two, or three fixations or a time limit of three seconds. Participants were naïve to the specifics of the fixation-dependent termination and a post-experiment questionnaire confirmed that they did not infer the experimental manipulation. Following image termination, two buttons were displayed, and the observer used the computer mouse to select whether they believed the target was present or absent. Feedback was not provided.

Participants were not informed that display termination was contingent upon their eye movements, but were simply told that the stimulus would appear for a variable amount of time. A debriefing questionnaire confirmed that all participants remained naïve to this manipulation, but trial conditions were randomized during the experiment to ensure that participants could not anticipate whether an image would appear upright or inverted and how long they would have to view the image.

Figure 1: Sample images from mouse search task. White dots represent fixations from all observers on a target absent image shown upright (a) and inverted (b), as well as a target present image shown upright (c) and inverted (d). Red dots indicate the expected or actual target location. Examples (b) and (d) were presented inverted to participants, but are shown upright here for ease of comparison.

Data Analysis

Behavioral Performance For each of the image orientation conditions (inverted or upright) and fixation allowances we calculated the proportion of target present trials in which the observers correctly detected the target (hit rate) and the proportion of target absent trials in which the observers incorrectly reported the target to be present (false alarm rate). Hit rate and false alarm rate were analyzed with a 2 (inverted or upright) × 4 (fixation allowance) repeated measures ANOVA. The hit rates and false alarm rates were transformed to standard signal detection measures of criterion (c) and sensitivity, d' (Green & Swets, 1989) and standard errors were calculated using bootstrap resample methods (Efron, 1979). We calculated the criterion and sensitivity differences between the upright and inverted conditions for each of the 10,000 individual bootstrap resamples across each fixation allowance. We then calculated the proportion of criterion/sensitivity differences above or below zero to estimate the probability of observing differences larger/smaller than zero.

Fixation Analysis To assess the guidance of eye movements, we calculated the average distance of each fixation from the target in the target present trials. The mode of the coordinate locations of the expected target location reported by five independent observers was used as the expected target location to compute distance measures on target absent trials. The results of this analysis were analyzed with a 2 (inverted or upright) × 6 (fixation number) × 2 (target present or absent) three-way repeated measures ANOVA.

Results

Behavioral Responses

Figure 2a presents observers' hit rates as a function of their fixation allowance. Observers' hit rates increased significantly with increasing number of fixations ($F(3,141) = 53.38$, $p < 0.001$) but were significantly lower for inverted than upright images ($F(1,47) = 71.80$, $p < 0.001$). When observers were allotted three seconds to search the image, the difference between their hit rates on upright and inverted images did not reach significance.

Figure 2b shows that the false alarm rate for the upright condition decreases with increasing number of fixations ($p < 0.05$ for all pairwise comparisons except the difference between two and three allowed fixations). In contrast, for the inverted scene condition the reduction in false alarm rate is only present in the three second presentation with more than four fixations ($t(47) = 3.648$, $p = 0.001$). Across the first three fixations there was no reduction in false alarm rate for the inverted scene condition ($p > 0.05$ for all pairwise comparisons). In addition, for the first fixation there is a trend for a higher false alarm rate for the upright scenes than the inverted scenes ($t(47) = 1.423$, $p = 0.161$).

Signal detection analysis allows us to separate effects on sensitivity or target detectability (d') from propensity to make a "target present" decision (criterion). Results (Figure 2c) show higher sensitivity at detecting the target when the scenes were upright than when they were inverted ($p < 0.02$). In addition, not surprisingly, sensitivity increases with longer exploration ($p < 0.001$).

Arguably more surprising but consistent with the false alarm rate are the results related to the criterion (Figure 2d). As scene exploration unfolds, observers' criterion for upright scenes increases steadily, whereas criterion for inverted scenes stays relatively constant throughout the first three fixations. With up to three seconds to explore the scene, the difference between upright and inverted criteria did not reach significance ($p=0.1$) Also, interestingly, for observers' first fixation there is a trend for a lower criterion for upright images than for inverted images ($p < 0.06$).

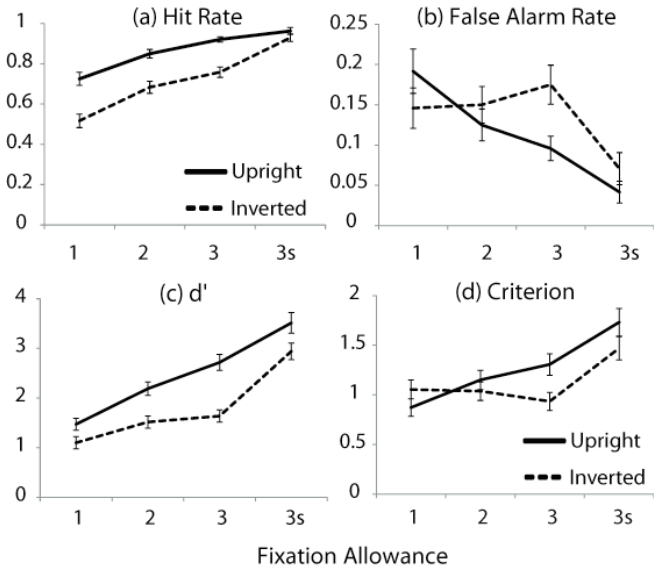

Figure 2: hit rate (a), false alarm rate (b), sensitivity (d'; c), and criterion (d) as a function of fixation allowance for upright and inverted images.

Eye Movement Analyses

Fixation analyses show a significant effect of inversion on the ability of the observers to fixate the target or expected target location. Figure 3 (black lines) shows the average distance of each fixation to the target location for inverted and upright scenes for target present trials. Distance measures were averaged across fixation allowance conditions after verifying that the experimental manipulation of display termination did not influence distances to target location of preceding fixations.

The first fixation shown in Figure 3 is the landing point of the first saccade *into* the image from the forced initial fixation outside of the image. The second fixation shown in Figure 3, which is the landing point of the first saccade *within* the image, is the first fixation counted toward the allowance on every trial. Therefore, at least two fixations

were recorded for every observer on every trial (except on a small percentage of erroneous trials due to tracker or observer error). More generally, on a trial where n fixations were allowed, n+1 fixations within the image will be recorded.

On target absent trials, we computed the average distance of each fixation to the expected target location (Figure 3, grey lines) estimated by the selections of five separate observers that did not participate in the search task. Fixations are significantly closer to target locations for upright scenes than inverted scenes in both target present and absent images ($F(1,47) = 263.98$, $p<0.001$). Representative examples of observer fixations are shown in Figure 1. The increase in average distance to target location on the fourth, fifth, and sixth fixations for upright, target present trials likely occurs due to exploratory eye movements made after target localization in the 3 second viewing time condition.

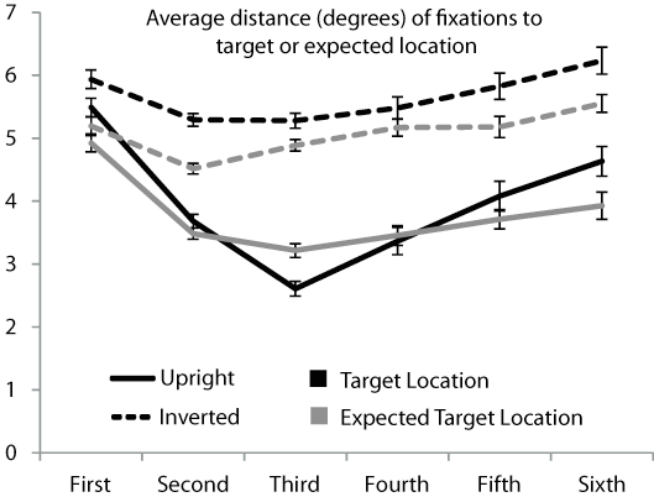

Figure 3: Average distance of fixations to target (during target present trials) or expected target location (during target absent trials) for upright and inverted images. It is important to note that increasingly fewer fixations are included in the averages for later fixations due to the fixation allowance manipulation.

Discussion

The objective of the current paper was to investigate the influence of scene inversion on search decisions and eye movements. The first result is that scene inversion reduces observers' sensitivity (d') to detect a target. This reduction in detectability could be solely attributed to an increased difficulty to detect an inverted target in both the fovea and visual periphery. However, the disruption of eye movement guidance toward expected target locations with scene inversion suggests that the reduction in target detectability is also partly due to an inability to rapidly extract scene context and foveate potential target locations.

Arguably the more interesting finding is the influence on observers' decision criterion. Our results display a marginally lower criterion (see Figure 2d, $p < 0.06$) following the first fixation when the image was upright than when it was inverted. As scene exploration evolves through the first three saccades, the false alarm rate decreases significantly when the scene is upright but not when it is inverted. In addition, the analysis of the fixations in target absent images shows that the eye movements in the upright scene condition are more guided toward expected target locations. Taken together we interpret these results as suggesting that observers fixate likely target locations in a scene and as they reject individual likely target locations they become less likely to make a target present judgment when the target is not there. When the scene is inverted, rapid extraction of scene context is disrupted, preventing typical guidance of eye movements toward likely target locations, thereby delaying the process of rejecting candidate target locations and the lowering of the decision criterion.

A possible explanation for the marginally lower criterion on the first fixation is the role of rapid scene gist extraction (Greene & Oliva, 2009; Henderson & Hollingworth, 1999; Rousselet et al., 2005) in setting the initial decision criterion. This explanation is consistent with the finding by Hillstrom et al. (2012) that after a 250 ms preview of a scene, only the first two eye movements during unconstrained scene search are influenced by gist information. When observers are first presented with the image, they recognize the office space as a likely scene to contain a target mouse and lower their decision criterion. In this interpretation, scene inversion would disrupt the rapid extraction of scene gist and not automatically lower the criterion. This automatic adjustment of decision criterion based on scene gist extraction would be a useful strategy if the observer is presented with images from different categories (e.g., office space, jungle, beach, etc.) and lowers their criterion for scenes semantically related to the given target.

Our discussion has emphasized the use of scene context to guide eye movements and lower the decision criterion but it is possible that the process of false alarm rate reduction might also occur in the absence of eye movements though processes of covert attention to the visual periphery (Ludwig, Davies, & Eckstein, 2014; Posner, Snyder, & Davidson, 1980).

Additionally, we have used the term scene context broadly but it is possible that various different types of cues are guiding eye movements and are disrupted when the scene is inverted. Scene gist (Torralba et al., 2006), as discussed, is one commonly investigated example that may cue attention. Co-occurring objects, in this case the computer monitor, keyboard, or desk, can also serve as useful indicators of target location (Castelhano & Heaven, 2011; Eckstein et al., 2006; Mack & Eckstein, 2011).

Finally, viewing time is not equated across the levels of the fixation allowance and could partially explain the increased sensitivity with increased number of allowed saccades, although the target object (a computer mouse) is

difficult to detect in the visual periphery suggesting that eye movements are likely important in the increase in target detectability.

In conclusion, we have demonstrated that inverting a scene increases the likelihood that participants will falsely detect a target object, thus lowering their decision criterion as compared to that of upright scenes, likely due to the disruption of the extraction of scene gist. These results, as well as future results derived from the manipulation of scene orientation, are useful for constructing critical tests of scene understanding mechanisms, and for understanding their influences during the scene exploration process.

Acknowledgments

Support for this research was provided by the National Institute of Health (R21 EY023097) and the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

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

Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*. http://doi.org/10.3758/s13423-011-0107-8

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional Cues in Real Scenes, Saccadic Targeting, and Bayesian Priors. *Psychological Science*, *17*(11), 973 – 980. http://doi.org/10.1111/j.1467-9280.2006.01815.x

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 1–26.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 628.

Green, D. M., & Swets, J. A. (1989). *Signal Detection Theory and Psychophysics*. Peninsula Pub.

Greene, M. R., & Oliva, A. (2009). The Briefest of Glances The Time Course of Natural Scene Understanding. *Psychological Science*, *20*(4), 464–472. http://doi.org/10.1111/j.1467-9280.2009.02316.x

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*(1), 243–271.

Hillstrom, A. P., Scholey, H., Liversedge, S. P., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review*, *19*(2), 204–210. http://doi.org/10.3758/s13423-011-0205-7

Hsiao, J. H., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science*, *19*(10), 998–1006.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition*, *68*(1), B1–B11.

Kelley, T. A., Chun, M. M., & Chua, K.-P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *Journal of Vision*, *3*(1), 1. http://doi.org/10.1167/3.1.1

Klein, R. (1982). Patterns of perceived similarity cannot be generalized from long to short exposure durations and vice versa. *Perception & Psychophysics*, *32*(1), 15–18. http://doi.org/10.3758/BF03204863

Ludwig, C. J. H., Davies, J. R., & Eckstein, M. P. (2014). Foveal analysis and peripheral selection during active visual sampling. *Proceedings of the National Academy of Sciences*,

Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision*, *11*(9). http://doi.org/10.1167/11.9.9

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621. http://doi.org/16/j.visres.2005.08.025

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527. http://doi.org/10.1016/j.tics.2007.09.009

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, *109*(2), 160–174.

Preston, T. J., Guo, F., Das, K., Giesbrecht, B., & Eckstein, M. P. (2013). Neural representations of contextual guidance in visual search of real-world scenes. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(18), 7846–7855. http://doi.org/10.1523/JNEUROSCI.5840-12.2013

Rock, I. (1974). The perception of disoriented figures. *Scientific American*. Retrieved from http://doi.apa.org/psycinfo/1974-20167-001

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, *3*(6), 5. http://doi.org/10.1167/3.6.5

Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes? *Visual Cognition*, *12*(6), 852–877. http://doi.org/10.1080/13506280444000553

Shore, D. I., & Klein, R. M. (2000). The effects of scene inversion on change blindness. *The Journal of General Psychology*, *127*(1), 27–43.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225–245. http://doi.org/10.1080/14640749308401045

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, *113*(4), 766–786. http://doi.org/10.1037/0033-295X.113.4.766

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*(4), 471–491.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain. *The Journal of Neuroscience*, *29*(34), 10573–10581. http://doi.org/10.1523/JNEUROSCI.0559-09.2009

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology*, *15*(24), 2256–2262.