Temporal and peripheral extraction of contextual cues from scenes during visual search

Kathryn Koehler

Miguel P. Eckstein

Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA, USA

Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA, USA



1

Scene context is known to facilitate object recognition and guide visual search, but little work has focused on isolating image-based cues and evaluating their contributions to eve movement guidance and search performance. Here, we explore three types of contextual cues (a co-occurring object, the configuration of other objects, and the superordinate category of background elements) and assess their joint contributions to search performance in the framework of cue-combination and the temporal unfolding of their extraction. We also assess whether observers' ability to extract each contextual cue in the visual periphery is a bottleneck that determines the utilization and contribution of each cue to search guidance and decision accuracy. We find that during the first four fixations of a visual search task observers first utilize the configuration of objects for coarse eye movement guidance and later use cooccurring object information for finer guidance. In the absence of contextual cues, observers were suboptimally biased to report the target object as being absent. The presence of the co-occurring object was the only contextual cue that had a significant effect in reducing decision bias. The early influence of object-based cues on eye movements is corroborated by a clear demonstration of observers' ability to extract object cues up to 16° into the visual periphery. The joint contributions of the cues to decision search accuracy approximates that expected from the combination of statistically independent cues and optimal cue combination. Finally, the lack of utilization and contribution of the background-based contextual cue to search guidance cannot be explained by the availability of the contextual cue in the visual periphery; instead it is related to background cues providing the least inherent information about the precise location of the target in the scene.

Introduction

Visual search is an important component of everyday life. Whether we are searching for the vitamin we dropped on the kitchen floor or the television remote in an unfamiliar living room, we have many visuocognitive mechanisms trained and ready to perform such tasks (Eckstein, 2011; Wolfe, 1994). If we know the basic features of the vitamin we dropped, we can use this information to facilitate locating the vitamin (Bravo & Farid, 2009; Burgess, 1985; Eckstein, Beutter, Pham, Shimozaki, & Stone, 2007; Malcolm & Henderson, 2009; Rao, Zelinsky, Hayhoe, & Ballard, 2002; Zelinsky, 2008). Similarly, we know that television remotes are generally on coffee tables, coffee tables are usually in front of couches, and we can easily identify the location of those things to define a small-relative to the entire visual field—region of space to search for a television remote in an unfamiliar living room. In realworld search tasks, we often employ our pre-existing knowledge about scenes and targets. The incorporation of contextual, often called top-down, information into models of human eye-movements has been shown to be much more important than intrinsic stimulus features (bottom-up information) for correctly predicting human fixations during a variety of visual search tasks (Birmingham, Bischof, & Kingstone, 2009; Chen & Zelinsky, 2006; Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Koehler & Eckstein, in press; Zelinsky, Zhang, Yu, Chen, & Samaras, 2005).

Much work has been done to quantify the contextual information contained within artificial and natural images. Contextual cues can range from familiar spatial layouts of objects in artificial scenes (typically known as contextual cueing; Chun, 2000; Chun & Jiang, 1998; Olson & Chun, 2002), to the identification of the category or gist of real scenes (Joubert, Rousselet, Fize,

Citation: Koehler, K., & Eckstein, M. P. (2017). Temporal and peripheral extraction of contextual cues from scenes during visual search. Journal of Vision, 17(2):6, 1–32, doi:10.1167/17.2.6.

doi: 10.1167/17.2.6

Received July 1, 2016; published February 28, 2017

ISSN 1534-7362 Copyright 2017 The Authors



& Fabre-Thorpe, 2007; Oliva, 2005; Torralba, Oliva, Castelhano, & Henderson, 2006), or to semantically related objects within a real scene (Hwang, Wang, & Pomplun, 2011; Moores, Laiti, & Chelazzi, 2003; Wu, Wick, & Pomplun, 2014). Contextual information was originally shown to facilitate object recognition (Biederman, 1972; Boyce & Pollatsek, 1992; Oliva & Torralba, 2007; Palmer, 1975) and, somewhat controversially (De Graef, Christiaens, & d'Ydewalle, 1990; Henderson, Weeks, & Hollingworth, 1999), guide initial scene exploration to informative or unexpected regions and objects (Antes, 1974; Loftus & Mackworth, 1978; Mackworth & Morandi, 1967). More recently, the effects of scene context on visual search have been explored, demonstrating that scene-based expectations can guide eye movements to expected target locations (Castelhano & Henderson, 2007; Eckstein, Drescher, & Shimozaki, 2006; Mack & Eckstein, 2011; Neider & Zelinsky, 2006; Torralba et al., 2006; Wu, Wang, & Pomplun, 2014).

Contextual cues have been dichotomized into local and global forms of context (Brockmole, Castelhano, & Henderson, 2006). Local cues are structural and spatial regularities immediately surrounding a visual target and have been shown to be the important factor in facilitating target localization (Olson & Chun, 2002) whereas global cues are comprised of elements in the overall display and have been shown to also improve observers' performance at target detection across repeated display epochs (Jiang & Wagner, 2004). Most commonly, researchers have explored the gist of a scene, loosely thought of as our overall impression of a scene and its content. Gist has been experimentally related to many different scene properties (Oliva, 2005; Koehler & Eckstein, in press), such as the basic-level category of a scene (Larson & Loschky, 2009), the background elements of a scene (Wu, Wang et al., 2014), and the perceptual content of an image, ranging from a description of low-level image properties to descriptions of objects and their key interactions within a scene (Fei-Fei, Iver, Koch, & Perona, 2007), to name just a few. Overall, current classifications of what constitutes contextual information are often broad or vary greatly across different studies.

In recent work (Koehler and Eckstein, in press), we have tried to partition scene context into separable image cues that can be independently manipulated. We investigated the influence of the scene background, the object that co-occurs most closely in space with the target (object co-occurrence) and the spatial configuration of the remaining objects in the scene (multiple object configuration) on search performance and eye movement guidance. Such work showed that objectbased cues guide and facilitate search more than the scene background. The current study investigates many remaining questions: the temporal dynamics of the cue extraction, the interaction of the scene contextual cues in contributing to search accuracy, and whether the utilization by the brain of each contextual cue to guide eye movements is related to the availability of the contextual cue in the visual periphery.

The first goal of the current study is to assess whether the contextual cues are all extracted with a similar time-course or whether observers rely on one contextual cue early in the search and then switch to a different cue as the other contextual cues become available to the visuo-oculomotor system. Interactions between the utilization of cues and time are expected given a large literature showing differential timecourses for the extraction of different information from scenes. For example, the basic-level category of a scene ("gist") can be processed with as little as 20 ms (Antes, Penland, & Metzger, 1981; Fei-Fei et al., 2007; Metzger & Antes, 1983; Potter, 1975; Thorpe et al., 1996). In contrast, estimates for the time to process objects based on behavioral and/or neural measures vary dramatically from as little as 14–40 ms for simple objects in isolation (Grill-Spector, Kushnir, Hendler, & Malach, 2000; Keysers, Xiao, Földiák, & Perrett, 2001) to 135–500 for object detection in complex natural scenes (Johnson & Olshausen, 2003; Thorpe et al., 1996). These results would suggest that the eye movement system might rely on different contextual cues (object based vs. background) as the search progresses from the first to later fixations during search.

A second goal of the study is to assess how the joint presence of each contextual cue contributes to increasing search accuracy. There is a well-established framework based on classic signal detection theory to predict the performance benefits from optimally combining statistically independent cues (Green & Swets, 1966). Perceptual accuracy is typically measured with each of the independent cues in isolation and subsequently with the combined cues. The benefits of multiple cues to accuracy (d') are compared to that expected from an optimal combination (typically with an assumption that the cues are statistically independent; Hillis, Watt, Landy, & Banks, 2004; Ernst, 2006; Landy, Maloney, Johnston, & Young, 1995; Shimozaki, Eckstein, & Abbey, 2003; Trommershauser, Kording, & Landy, 2011). Here, we use such framework to evaluate how the brain combines multiple scene context cues to benefit search accuracy.

The third goal of our investigation is to assess whether the utilization of a contextual cue to guide eye movements is intricately related to its availability in the visual periphery. This hypothesis would propose that the extent to which a contextual cue is utilized by the brain to guide eye movements is mostly determined by the degree to which the cue can be extracted in the visual periphery. There is a large literature suggesting that categorical and semantically descriptive scene information can be rapidly extracted prior to an eye movement from the visual periphery (Potter, 1975; Antes et al., 1981; Fei-Fei et al., 2007; Metzger & Antes, 1983; Calvo, Nummenmaa, & Hyönä, 2008; Li, VanRullen, Koch, & Perona, 2002) even at 70° into the visual periphery (Boucart, Moroni, Thibaut, Szaffarczyk, & Greene, 2013) and sometimes even better than in the fovea (Larson & Loschky, 2009; Loschky et al., 2015; see Strasburger, Rentschler, & Jüttner, 2011 for a review). On the other hand, previous studies have shown that object identification which comprises more spatially local information might degrade more abruptly in the visual periphery, particularly in the presence of other objects (known as crowding, Levi, 2008; Whitney & Levi, 2011). Utilizing object cooccurrence to guide search might require this precise object identification resulting in degradation of information from this cue in the visual periphery. Finally, multiple object configuration likely involves identifying whether objects are in a proper spatial arrangement (Jiang & Wagner, 2004; Olson & Chun, 2002), possibly only requiring observers to spatially resolve an arrangement of objects without having to identify the individual objects themselves. Previous studies have shown that coarse information of a scene's low spatial frequency information, sufficient to convey the approximate spatial layout of a scene, is used to categorize scenes during early-stage processing (Schyns & Oliva, 1994). Here, we evaluate the ability of observers to extract each of the contextual cues as a function of retinal eccentricity to understand whether the utilization of the contextual cues to guide search is related to observers' ability to extract the cue in the visual periphery.

Definition of scene context cues

Following previous work (Koehler & Eckstein, in press) we partitioned scene context cues into three distinct cues that can independently be manipulated in scenes: two object-based cues, the co-occurring object and multiple object configuration, and a backgroundbased cue. An element was considered to be part of the background if it would be regarded as nonmanipulable/ nonstructural from the point of view of an observer walking into a room or natural environment (as opposed to other spatial scales, e.g., a close-up view of items on a table where the items would be objects and the table-top a background; see Henderson & Hollingworth, 1999). See the top-left panel of Figure 1 for an example of each cue in a scene where the target was a pillow. The co-occurring object is an object that

typically co-occurs with the target object and is typically the closest spatially (among other objects in the scene) to the target (the bed in Figure 1). For the co-occurring object to facilitate search for the target it should be easier to detect than the target object itself. Multiple object configurations provide information about the location of the target through their spatial arrangement. The objects could be spatially distant from one another and the target, and unlike object cooccurrence, they individually do not provide spatially precise information about the target location. It is only the combination of all of the objects in a particular arrangement which provides target spatial location information. For example, bedrooms will almost always contain a bed with an adjacent nightstand and lamp, as well as a dresser and closet. Finally, the background category is comprised of all background elements of the scene (anything plausibly immovable or nonconfigurable in the scene, e.g., ceilings, floors, sky, ground, trees, doors, etc.) and portrays the superordinate level category of the background elements (either indoor, natural outdoor, or urban outdoor; Figure 1). By embedding these cues in a single set of computer rendered scenes, we can manipulate them in various combinations in order to carefully understand the individual, temporal, and peripheral characteristics of each cue and their interactions.

Koehler & Eckstein

Experiment 1: Explicit judgments about cue spatial informativeness and expected target locations

In the first experiment, we sought to make measurements about how informative the separate contextual cues were of the likely locations of the target objects in the scenes. We assessed this in two ways: by asking directly about the relative informativeness of scenes with variations of the contextual cues and also by requiring observers to select in the scenes where they would expect the target object to be located when viewing scenes containing only one type of cue or all three cues. These measurements serve as a validation of the experimentally manipulated cues. Because these measurements are made with unlimited time and foveation, they also serve as an upper bound of the inherent information provided by a contextual cue about the likely target location when observers had access to only a single cue. This upper bound can then be related to the guidance provided by the cue during search under brief time periods (prior to search saccades) and peripheral processing.



Figure 1. Example of stimuli presented to participants in the cue manipulation verification task for a sample scene. In this scene, the target was PILLOW. All stimuli in this task were target absent images. As labeled, observers in the object co-occurrence task (O task) viewed an image with all objects jumbled except the co-occurring object on a gray background, observers in the multiple object configuration task (M task) viewed images without the co-occurring object present with all objects ordered in a typical way or jumbled, observers in the background category condition (B task) viewed images with a matched or mismatched to the target background category. Observers' tasks were to select the object (O condition) or image (M and B conditions) that would provide the most information about where the target object would be located and to indicate where in the image they would expect the target object to be located.

Methods

Participants

A total of 360 individuals recruited from Amazon Mechanical Turk (AMT) reported having normal or corrected-to-normal vision and participated in the experiment. An additional 21 undergraduate students from University of California, Santa Barbara, who received course credit for participation and were tested to have normal or corrected-to-normal vision participated in the study. All observers provided informed consent to participate.

Design

Separate groups of 40 observers each viewed a group of 15 images such that no observer saw the same target

category twice. Each observer was assigned to the object co-occurrence (O), multiple object configuration (M), or background category (B) condition; therefore, a total of 120 observers viewed 45 images for each condition.

Stimuli

The stimuli comprised images depicting natural indoor and outdoor scenes with manipulations of three types of contextual cues. A base set of 45 scenes was constructed in Unity 3D (Unity Technologies, Bellevue, WA), a video game building and physics engine platform, each with a specified target object that would serve as the searched-for item in the visual search task. Each scene contained other objects that one might expect to find in a scene containing that object and a background that was consistent with the object. There were 15 unique target object types, each used three times in three different instantiations (e.g., the viewing angle, design, color, or size was varied across the three targets). Each scene has a base-version containing all three experimentally defined contextual cues. The base scenes were manipulated to form versions of the scene missing certain individual context cues. For the purpose of verifying the contextual cue manipulations, participants in this task viewed versions of the scenes with individually isolated contextual cues, different from Experiments 2 and 3 (see the relevant stimuli sections). Each scene was constructed such that the base scene contained a target with a frequently cooccurring object placed near to it (constituting the object co-occurrence cue), a number of other objects that would typically also be present in the scene arranged in a typical way (the multiple object configuration cue), within a background that exemplified the scene category and was consistent with the target and other objects (the background category cue). Other versions of the scenes were created to isolate the various contextual cues. Target absent versions of each scene type were created where everything in the scene remained the same except for the deletion of the target object. For this experiment, all participants viewed target absent versions of the scenes. Participants in the O condition viewed a version of the scene with all objects except the co-occurring object jumbled on a gray background. Participants in the M condition viewed a version of the scene with and without the cooccurring object on a gray background. Finally, participants in the B condition viewed versions of the scenes with all objects removed, i.e., just the backgrounds. Example stimuli are shown in Figure 1. There were two AMT quality assurance images included as well, described in the procedure.

Procedure

After consenting to participate in a psychological study and indicating that they had normal or correctedto-normal vision, participants were given a brief tutorial about how to use the experiment interface. Each participant performed two tasks. The first task varied by condition. Observers in the O condition were required to click on the object that was most informative of the target object's location. Condition M and B observers were required to select the image that they thought was most informative of the target object's location between a jumbled and nonjumbled version of the objects (without the co-occurring object on a gray background) or between an indoor or outdoor background (with no objects), respectively. The second task required participants to click with the computer mouse a location within the image (whichever they had previously selected in task 1 for M and B participants) where they would expect the target object to be located. The task instructions were designed to be uniform across conditions and to explicitly assess the informativeness of the cue manipulations. It is important to note for the first task that observers in the O condition were choosing to select one of the many objects (typically >10) in the scene whereas observers in the M and B conditions were choosing between only two possible options; therefore, selections made solely by chance would result in drastically different rates of selecting our experimental manipulation. Image order was randomly determined, and two quality assurance trials were randomly mixed into the experimental trials to verify that observers were capable of completing the tasks. The first quality assurance trial was to indirectly assess overall understanding of the task instructions and mastery of the English language by using a simplified version of the stimuli to which there was an obvious correct answer. The second trial was to ensure that click recording was calibrated correctly within the browser window and required participants to click at the center of a target. To summarize, on each trial, observers were prompted at the top of the screen with the task instruction, including which object they were to make assessments about. They had an unlimited amount of time to complete the first task, after which they immediately began the second task for the same object, again with unlimited time to make their assessment. Image order was randomized across participants. At the end of the experiment, participants filled out a short questionnaire indicating how well they felt they understood each tasks' instructions, their age, and their gender.

Given the differences in task 1 between the O, M, and B conditions, we opted to perform a follow-up secondary task for condition O, task 1, with a group of 21 separate undergraduate observers that more directly probed the basis of our object co-occurrence manipulation, but would have violated the uniformity of instructions and stimuli in the main task. These observers were asked to select the object that they would expect to be physically closest to the target object while viewing a scene with all contextual cues present.

Results

Verification of experimental contextual cue informativeness

Participants who reported understanding the tasks with a rating that was two standard deviations below the mean were discarded from analysis. The average reported level of understanding among remaining participants was 9.2 for both tasks 1 and 2 on a 10 point scale, with 10 being the highest level of



Figure 2. Part (a) of this figure depicts the proportion of observers that selected our chosen manipulation of a cue to be the most informative version of that cue for a target detection task for each of the 45 scenes. The O column corresponds to the object cooccurrence condition, the M to multiple object configuration, and B to background category information. The O2 column depicts the results from a follow-up task where we asked participants to select the object that they would expect to be physically closest to the target object. The color of the cell representing the proportion of times the participants selected the co-occurring object in that task. Part (b) shows the histogram of the proportions depicted in part (a).

understanding. After discarding an additional four observers who failed the MTurk quality assurance task criteria, we analyzed the data of 110 participants for the O condition (image group 1, n = 36; image group 2, n =36; image group 3, n = 38), 111 participants for the M condition (image group 1, n = 36; image group 2, n =37; image group 3, n = 38), and 107 participants for the B condition (image group 1, n = 35; image group 2, n =35; image group 3, n = 37). Shown in Figure 2 (left) is the proportion of participants who verified our manipulation of a particular cue for each image. A verification for the O task was taken to be an instance where the participant selected the experimentally defined co-occurring object as the most informative of the target object's location. We considered the manipulation of the M task to be verified when a participant

selected the experimentally defined nonjumbled version of the multiple objects. Finally we deemed that the manipulation of the B condition to be verified if the participant chose the experimental background as the most informative of the target object's location. The right side of Figure 2 shows a histogram of the proportion of agreement for each contextual cue.

There were many more instances of poor verification of the object co-occurrence (O) manipulation in the first task, likely because there were so many possible objects to choose from, justifying further exploration with our follow-up task. Figure 2, column O2, shows the proportion of times observers' selected our experimentally defined co-occurring object when instead asked to select the object they would expect to be closest to the target object (therefore most informative

	-	-
		,

			o-order <i>r</i>									
	М		В		OMB		β		pr ²		b	
	X	Y	X	Y	X	Y	X	Y	x	Y	X	Y
0	0.39**	0.50***	0.06	0.48***	0.88***	0.88***	0.84	0.84	0.86	0.82	1.07***	1.12***
Μ			0.14	0.52***	0.44**	0.49***	0.10	0.07	0.20	0.11	0.14	0.09
В					0.12	0.45**	0.06	0.01	0.13	0.02	0.09	0.01
									Int	ercept:	32.10	28.91

Table 1. Summary of results using the expected target locations collected from observers who viewed scenes with individual cues to predict the expected target location judgments of observers who viewed fully cued scenes. Notes: **: p < 0.01, ***: p < 0.001.

of the target object's location) in a fully cued scene. The distribution of those proportions is more similar to that of the M and B verifications.

Explicit judgments about expected locations with varying contextual cues

We assessed how scenes with individual cues contributed toward observers' explicit expectations about target location relative to scenes with all three cues.

To do so we evaluated the extent to which the target location expectations (collected in the second task) made by participants for scenes with the individual cues could predict the selections made by 60 participants from Koehler and Eckstein (in press) with all three cues in the scenes. The data from the 60 participants in Koehler and Eckstein (in press) will be used again in analyses for Experiments 2 and 3 here. For the data from Koehler and Eckstein (in press) as well as for the data collected in the second task in this work, we calculated the mode of observers' expected target locations. Specifically, for each image, we obtained the mode by creating a density map of selected locations within the image, blurring that map with a 2D Gaussian blob with standard deviation equal to one degree of visual angle, and then finding the x and ycoordinates of the maximum value in the blurred density map. Therefore, we obtained the mode of all participants' expected target locations for the scenes with only individual contextual cues (O, M, and B tasks in the MTurk experiment from this work) and the scenes with all three cues (OMB; taken from Koehler and Eckstein, in press). The O, M, and B coordinates were used as predictor variables of the OMB coordinates of the judgments in a linear regression model, summarized in Table 1. The x coordinate selections of the single cue scenes accounted for 79% of the xcoordinate location selections in the scenes with all three cues, F(3, 41) = 51.90, p < 0.001, $R^2 = 0.79$. The v coordinate selections for the single cue scenes accounted for 77% of variance of the selections in the scenes with all three cues-coordinate, F(3, 41) = 46.42, p $< 0.001, R^2 = 0.77$. Importantly, the only individual cue

that contributed significantly to predicting the multicue scenes' expected target location judgments was the object co-occurrence cue (O), for both the x and y coordinates. This serves as another useful verification of our manipulation. Because we selected the cooccurring object to be spatially close to the target object, to the extent that observers are utilizing this information and selecting target locations that are proximal to the co-occurring object when present, these two measures will be highly correlated (see zero-order rbetween O and OMB for both the x and y coordinates in Table 1) and predictive of one another (see the partial correlations and coefficients for O in Table 1). The other manipulations were not as tightly spatially coupled with the target object, so we would not expect observers' judgments of the target location in those tasks to necessarily be as predictive as object cooccurrence of observers' judgments in the scenes with all cues.

Experiment 2: Temporal dynamics of the utilization of contextual cues

Having assessed each cue's informativeness about target location in the scene, we then sought to explore how each cue was utilized during the first few fixations of a visual search task. We used a paradigm where scene viewing time was contingent upon the number of fixations made within a scene while also manipulating the cue information present within the scene. On each trial the display was randomly interrupted after one, two, or three fixations, or was presented for a full two s. We evaluated observers' task performance via the index of detectability (d'), their bias in reporting target presence, and the proximity of their eye movements to the target location. We related these measures to the explicit location expectation judgments measured for each cue in Experiment 1. To provide a more finegrained analysis of the utilization of each cue on a fixation-by-fixation basis, we used observers' eye movements and expected target location judgments



Figure 3. Sample scene images for a trial in which the participant searched for CORK. The top image shows the full cue scene, the middle left shows the scene with only the object co-occurrence cue (O), middle-right with only the multiple object configuration cue (M), bottom-left with only the background category cue (B), and the bottom right with no cues. The sample scenes contain the target. There were five additional complementary scenes with target object removed. Participants saw one of the 10 total scenes and their task was to determine if the target object was present, with a known 50% likelihood of target object presence.

when only a single cue was present as predictors of various performance metrics in images containing all cues in a general linear regression model.

Methods

Participants

A total of 300 undergraduate students at the University of California, Santa Barbara participated in the experiment in exchange for course credit. All participants provided informed written consent and were verified to have normal or corrected-to-normal vision.

Stimuli

The scenes as described above were used for this experiment with a few differences. In order to preserve

the overall difficulty (e.g., clutter) of the search task between conditions, instead of using a gray background in trials where the B cue was absent, we mismatched the background on such trials. Ten versions of each scene were created corresponding to the contextual information levels described in the experimental design. An example of each scene is shown in Figure 3. When background category information is present, the background depicts a hardwood floor, painted walls, and a window. When background category information is absent, the background is replaced with a snowy mountain background.

Design

We manipulated the type of contextual information present in the stimulus (five levels: None, O, M, B, and OMB) and the number of allowed saccades (three



Figure 4. Sample timeline of a single trial during Experiment 2. The trial initiated once the participant fixated a crosshair and pressed a button, after which they were cued with the target they were to search for. In Experiment 2, after participants made their first fixation within the image, they were then given either one, two, or three additional fixations to explore the scene. Once they exhausted their allowance, a response screen appeared where the participant indicated whether the target was present and how confident they were in their decision.

levels: one, two, or three) while completing the task. Each participant served in all of the conditions, resulting in a two-way (3×5) repeated-measures design. In order to determine which set of images a particular observer would see, we Latin-square counterbalanced the 45 images into groups of 3 across the 15 possible condition combinations. Observers were randomly assigned to one of the 15 image assignment groups and image presentation order was randomized. Target presence or absence was determined randomly (using a random number generator) on each trial in order to prevent participants counting or keeping track of the number of trial types to influence later decisions.

Apparatus

Stimuli were displayed on a Barco MDRC-1119 monitor with 1280×1024 pixel resolution. Participants positioned themselves on a chin and forehead rest 76 cm away from the monitor so that a single pixel subtended 0.022° of visual angle. Eye tracking data were recorded on an Eyelink 1000 (SR Research Ltd., Mississauga, Ontario, Canada) monitoring gaze position at 250 Hz using a nine-point grid calibration procedure. A velocity greater than 22° /s and acceleration greater than 4000° /s² classified an event as a saccade.

Procedure

Participants were instructed that they would be viewing a series of images on the computer monitor and determining whether or not various objects were present in those images. They were told that were was 50% likelihood that the target would be present in the images. The time course of a single trial is shown in Figure 4. At the beginning of each trial, participants were required to fixate a cross at the bottom-center portion of the display monitor outside of the to-bepresented scene. They initiated a trial by pressing the space bar, at which point the name of the object (e.g.,



Figure 5. A diagram depicting the criteria for terminating image presentation during stimulus presentation. Observers initially fixated a cross outside of the image on the display. The landing time of their first fixation is denoted as t_1 (see annotations for times t_2-t_7). The image was removed from the display after t_3 if only a single fixation was allowed, after t_5 if two fixations were allowed, or after t_7 if three fixations were allowed. This enabled us to analyze up to a total of four saccade endpoints within the image, and three fixation latencies.

TRASHCAN) they were to search for appeared. They were required to read the object name without moving their eyes; otherwise the trial would reset and they would be required to initiate the trial again with another press of the spacebar. After 500–1500 ms, the test image appeared. After the requisite number of fixations (one, two, or three) had been made within the image, it was removed. More precisely, the image was removed upon the detection of the end of the saccade (see Figure 5 for a diagram). After image termination, a response screen appeared where participants could indicate how confident they were that the target object

was present. Responses of 1–5 indicated the object was absent, 1 being highest confidence, whereas a response of 6–10 indicated the object was present, 10 being highest confidence. Participants' first saccade from the initial fixation location into the image was not counted as part of their saccade allowance. Observers were not explicitly told that the image display time was dependent on their eye movement behavior. Instead, they were informed that the image would appear for a variable amount of time on each trial. No participant reported knowledge or discovery of the display timing criterion dependency on the eye movements.



Figure 6. The average sensitivity index of detection as a function of fixation allowance within the image for each contextual cue condition. Error bars represent an estimate of the standard error of the mean, as calculated from the sensitivity indexes delineating the inner 68.29% of the distribution of sensitivity indexes from 10,000 bootstrap resampled samples.

Statistical analyses

In order to quantify observers' performance on the visual search task, we estimated their index of sensitivity (d') from each recorded hit rate and false alarm rate after collapsing their confidence rankings into binary yes/no decisions about target presence (Green & Swets, 1966; Macmillan & Creelman, 2004). Because some observers had perfect false alarm or hit rates for a session (precluding calculation of their individual d' and estimation of standard error of the mean across observers), we utilized bootstrap resampling methods (Efron, 1992; see methods for details) to estimate the variability of d' across observers and perform statistical analyses of differences between the experimental conditions. For example, to assess the main effect of contextual condition on d', we resampled the data from all 2,700 trials of data (recorded from all observers) for each contextual condition with replacement 10,000 times. For each 10,000 resamples of 2,700 trials, we calculated a d' score, then computed the difference between each of the 10,000 d' scores for each pair of contextual conditions. We assessed the proportion of those differences in the tail above or below zero to generate a *p* value.

We also analyzed the guidance of observers' eye movements toward the target location using the recorded eye-tracking data. We assessed the distance of the observers' closest fixation to the target location on each trial for target present trials or to the expected target location on target absent trials. The expected target locations was the mode of selections made by 60 separate observers who freely viewed target absent scenes with all cues present and chose the most likely target location in the scene (see previous Experiment 1 section for details). Therefore, the target present and absent data were analyzed separately, each using a twoway repeated measures ANOVA controlling for the false discovery rate in post-hoc comparisons.

Results

Target detectability

We first assessed the observers' ability to detect the target object as a function of the number of fixations within the image manipulated using the saccadecontingent display termination. Figure 6 depicts the sensitivity (index of detectability, d') across each fixation allowance condition for each of the single and multiple contextual cue conditions. First, exploring the main effect of contextual information, we found a significant increase in observers' sensitivity when the multiple object configuration cue was present, or when all cues were present compared to when no cues were present across all fixation allowance conditions (None vs. M, Mean difference = 0.27, p < 0.001; None vs. OMB, Mean difference = 0.46, p < 0.001). After controlling the false discovery rate (Benjamini & Hochberg, 1995) to correct for multiple comparisons, there was no significant difference in the index of sensitivity for observers searching with the object cooccurrence cue, the background category cue, or no cue at all (None vs. O, Mean difference = 0.11, p = 0.06; None vs. B, Mean difference = 0.07, p = 0.18). Although this result may suggest that the object cooccurrence and background category cues do not influence target detection task performance within the first three fixations overall, it is important to note that there is a significant increase in observers' index of sensitivity when the object co-occurrence and back-



Figure 7. Average bias for each cue condition and fixation allowance. Zero corresponds to optimal (maximizing proportion correct) criterion placement for 50% target present/absent paradigms. A positive bias indicates a greater tendency to make a target absent judgment. Error bars represent an estimate of the standard error of the mean, as calculated from the biases delineating the inner 68.29% of the distribution of sensitivity indexes from 10,000 bootstrap re-sampled samples.

ground category cues are added to multiple object configuration (M vs. OMB, Mean difference = 0.19, p = 0.007). This demonstrates that, whereas in isolation, neither cue's effect on sensitivity reached statistical significance; when combined, they significantly increased task performance. This result is further supported by our assessment of the additivity of the effects of each cue on the index of sensitivity. Furthermore, we replicate the finding from Koehler and Eckstein (in press) that object information facilitates target of detectability more than background information (object vs background contrast difference = 0.25, p = 0.02 overall).

To rule out any fixation duration/accuracy trade-offs across conditions we compared the average fixation latencies in each condition using a repeated-measures ANOVA. We found that there was no significant effect of contextual condition on fixation latency, F (4, 2392) = 2.12, p = 0.08. The mean fixation latencies (\pm the standard error) for each condition were: None = 180.87 \pm 1.85 ms, O = 178.57 \pm 1.98 ms, M = 180.66 \pm 1.95 ms, B = 179.76 \pm 1.90 ms, OMB = 176.26 \pm 1.90 ms.

The more interesting analysis probes the interaction between contextual information and fixation allowance to assess whether a particular cue type is utilized to varying degrees on different fixations. We assessed the increase in the index of sensitivity across fixation allowance conditions for each type of contextual cue. The increase in d' between the nth— (n - 1)th fixation allowance condition was significant in all cases except between the 1st and 2nd M and 2nd and 3rd O fixation allowances (using FDR correction). Overall, participants' index of sensitivity increases as they are given more time to explore the image. To assess whether performance improvement varies across fixations depending on the type of contextual information present, we looked at the distribution of the difference in performance between pairs of contextual information types across two fixations. More specifically, we calculated for example the difference between O and B in the third fixation (OB_3) , the difference between O and B in the second fixation (OB_2) , and then assessed the distribution of $OB_3 - OB_2$ across all 10,000 bootstrap resampled indexes of sensitivity.¹ We did this for each of the ten contextual information pairs (e.g., OB, OM, MB, O, None, etc.) and for both the 1st/2nd and 2nd/3rd fixation changes. In total, we therefore had twenty distributions, each comprising 10,000 differences. We pooled all of the differences and failed to show significant evidence of an interaction effect (p =(0.37). This conclusion is supported by running a twoway, repeated measures ANOVA on the PC (proportion of trials correctly classified as target present/ absent) data, F(8, 2392) = 0.744, p = 0.65. Therefore, while there are clear differences in utilization of contextual information across all fixations, the increase in sensitivity at detecting the target is similar for each cue as scene exploration unfolds.

Bias

We also explored the change in a participant's bias to make a target present judgment given that the index of sensitivity varied across conditions. Figure 7 portrays our measurement of bias, which indicates how far from optimal (d'/2) the average observer criterion was for making target present and absent judgments. A bias value of 0 corresponds to the optimal (maximizing proportion correct) criterion placement (d'/2) for trials with equal probability of target presence and equal payoffs for hits and false positives (Green & Swets, 1966). Again using the bootstrap resampling methods described earlier, for each contextual condition, we resampled the data 10,000 times with replacement from the 2.700 recorded trials (900 trials for each fixation allowance) from all observers and calculated bias from each of those 10,000 samples. To compare conditions, we then plotted the distribution of difference scores between each of the 10,000 bias scores for each pair of contextual conditions. We found that observers were biased to report the target absent when there were no cues or only the multiple object configuration or background present in the scene and the co-occurring object was the important cue for reducing bias toward optimality (average overall bias reduction: 0.23, None vs. O; 0.25, M vs. O; 0.24, B vs. O; p < 0.001). Of note is the increase in bias, corresponding to a tendency to decrease the propensity to report the target as present in the fully cued condition as search progresses from the first and second to third fixations (p = 0.007 and 0.008, respectively; not significant after FDR correction). This could be the result of participants initially perceiving contextually intact scenes, consistent with the target object, and thus being likely to assume the target was present when having very few exploratory fixations, but then becoming more confident in rejecting target presence upon further exploration of the scene.

Contextual cue combination

A classic test when many visual cues are available evaluates whether a human's combination of multiple cues is consistent with an optimal combination (Green & Swets, 1966; Landy, Maloney, Johnston, & Young, 1995; Trommershauser, Kording, & Landy, 2011). It is common to first assume that the cues are statistically independent and that they elicit a Gaussian distributed internal response within the observer. Under these assumptions an optimal combination of cues reduces to a weighted average of the internal responses for each cue and makes a specific prediction about how performance (measured by the index of detectability; d') when all cues are present relates to the d's associated with each individual cue (see Appendix). Typically, the investigator makes a measurement of the accuracy (d') with each individual cue and then compares the accuracy prediction from the optimal cue combination to an empirical measurement of accuracy (d') with all cues.

Because search accuracy in our single contextual cue tasks has contributions from the presence of the contextual cue but also the physical presence of the target, a first step is to isolate the performance benefit arising from the presence of just the contextual cue. We first calculated the isolated effect (d' - cue) of each contextual cue relative to the condition where no cues were present and accuracy is only mediated by the

presence of the target (Equation 1; see the Appendix for the derivation of this equation). This isolates the contribution to search accuracy of each individual cue over that provided by the presence of the target. We then used Equation 2 (also derived in the Appendix) to calculate the predicted d' from the joint presence of the contextual cues, assuming that the cues are statistically independent and combined optimally (i.e., linearly combined with weights set optimally; see Appendix, Green & Swets, 1966). We compared this value to the empirically observed effect on d' with all cues present. Equation 3 therefore shows an example calculation of the predicted d'_{OMB} effect.

Koehler & Eckstein

$$d'_{\text{cue}_{i}} = \sqrt{d'_{\text{cue}_{i},\text{target}}^{2} - d'_{\text{target}}^{2}} \quad (1)$$

$$d'_{\text{predicted}} = \sqrt{d'_{\text{cue}_{1}}^{2} + d'_{\text{cue}_{2}}^{2} + d'_{\text{cue}_{3}}^{2}} \quad (2)$$

$$d'_{\text{predicted}} = \sqrt{d'_{\text{cue}_{1}}^{2} + d'_{\text{cue}_{2}}^{2} + d'_{\text{cue}_{3}}^{2}} \quad (2)$$

$$d'_{\text{OMB,predicted}} = \sqrt{d'_{O,\text{target}}^2 + d'_{M,\text{target}}^2 + d'_{B,\text{target}}^2 - 2d'_{\text{target}}^2}$$
(3)

Figure 8 displays the average predicted d' of the combination of the individual cues (from Equation 2) in comparison to the observed experimental result using the d' for each fixation allowance condition and also the average d' across all fixation allowances. To evaluate the consistency of our results across data sets, Figure 8 also shows a similar comparison of predicted d' for optimal cue combination versus empirical d' for the data in Koehler and Eckstein (in press), where observers were given 1500 ms to search for the target in a group of 48 scenes. That experiment also included additional contextual conditions combining two cues (MB, OB, and OM). The points lie generally along the identity line, suggesting that observer benefits with multiple contextual cues are consistent with that expected from optimal integration of independent cues. We calculated individual slopes for the 10,000 bootstrap sample point sets while forcing the intercept to be zero. The average slope was 0.994 with 45.67% of the slopes greater than one; therefore, we fail to reject the hypothesis that the cue combinations are consistent from that expected from optimal integration of independent cues. The finding seems to generalize across data sets, irrespective of shorter presentations in the current paper and longer presentations and additional conditions in Koehler and Eckstein (in press).

Eye movement guidance

In order to assess the extent of guidance for subsequent saccades offered by each cue on the visual



Figure 8. Observed target detectability (d') for various cue conditions versus the predicted d's of multiple contextual cues based on optimal combination of independent cues. These calculations were made using the average d' for each fixation allowance condition (labeled as one, two, and three fixations in the legend) and by averaging across the fixation allowance conditions (labeled as "all fixation allowances"). The error bars represent the inner 68.29% of the distribution of 10,000 bootstrap resampled average derived and observed d' values. Points in the legend with the symbol * were calculated identically, but correspond to additional data taken from Koehler and Eckstein (in press), where observers had 1500 ms to search for the target in 48 scenes and there were an additional three contextual combination conditions comprised of two cues (MB, OB, and OM).

search task, we computed the average distance of the closest saccade endpoint to the target (on target present trials) or expected target (on target absent trials) location for each fixation allowance and contextual cue condition. In a trial where observers were allowed to make one fixation, we recorded two saccade endpoints (the endpoint of the saccade corresponding to the fixation and the endpoint of the subsequent saccade). Therefore, in determining the closest fixation to the target location, we have two, three, and four saccade endpoints to analyze for the one, two, and three fixation allowance conditions (refer again to Figure 5). First, we will consider the results for target present trials, shown in Figure 9. The minimum distance of the closest fixation to the target location was analyzed using a two-way ANOVA.²

The effect of contextual condition was significant, F(4, 3931) = 27.97, p < 0.001, as was the effect of fixation allowance, F(2, 3931) = 345.05, p < 0.001. The interaction between fixation allowance and contextual



Figure 9. Average distance of an observers' closest fixation to the target location as a function of fixation allowance for each contextual cue condition. Target present trials only are included in this analysis. Error bars represent standard of the mean.



Figure 10. Average distance of an observers' closest fixation to the expected target location as a function of fixation allowance for each contextual cue condition. Target absent trials only are included in this analysis; therefore this data illustrates participants' behavior in the absence of target feature information guidance. Expected target location was calculated as the mode of the location where a separate group of observers expected the target to be located for a given scene. Error bars represent *SEM*.

condition was not significant, F(8, 3931) = 1.003, p =0.43. In order to understand the overall benefits of the various contextual cues to eve movement guidance. irrespective of fixation allowance, we performed posthoc comparisons controlling for the false discovery rate between the fully cued and no-context conditions as well as between each of the singly-cued conditions and the no-context condition. Eye movements were significantly closer to the target location when all contextual cues were present than when none were present (mean difference = 1.01° , p < 0.001). Compared to when no cues were available, eye movements were also significantly closer overall when multiple object configuration (mean difference = 0.41° , p < 0.001) and object co-occurrence (mean difference = 0.58° , p <0.001) information was present, but not when background category information was present (mean difference = 0.08° , p > 0.25).

Next, we turn to the results of target absent trials shown in Figure 10, where target feature information is removed, isolating the contribution of contextual information to eye movements guidance. For these trials, we analyzed the distance of fixations from the mode of the expected target location, as reported by a group of 60 separate observers (discussed in Experiment 1). The minimum distance of the closest fixation to the expected target location was analyzed using a two-way, repeated measures ANOVA. The effect of contextual condition was significant, F(4, 1188) = 60.93, p < 0.001, was the effect of fixation allowance, F(2,594) = 194.25, p < 0.001. The interaction between fixation allowance and contextual condition was also significant, F(8, 2376) = 2.033, p = 0.039. We performed posthoc comparisons while controlling for the false positive rate between the condition with all three cues and no-contextual cue conditions as well as between each of the single cue conditions and the no-contextual cue condition. Eye movements were significantly closer to the target location when all contextual cues were present than when none were present (mean difference = 1.26° , p < 0.001). Compared to when no cues were available, eye movements were also significantly closer overall when multiple object configuration (mean difference = 0.44° , p = 0.004) and object co-occurrence (mean difference = 1.01° , p <0.001) information was present, but not when background category information was present (mean difference = -0.06° , p > 0.25).

We were interested in assessing the time course of contextual guidance of each contextual cue, but also in interpreting the significant interaction between contextual cue and fixation allowance. Again, for all fixation allowance conditions, background category fails to have a significant effect on eye movement guidance (p > 0.15 in all cases). In contrast to the results for target present trials, the facilitative effect of the object cooccurrence is present throughout all fixation allowances, (None vs. O: First fixation, mean difference = 0.60° , p = 0.01; second fixation, mean difference = 1.34, p < 0.001; and third fixation, mean difference = 1.07, p < 0.001), whereas the multiple object configuration cue does not have a significant influence on eye movement guidance until after the second fixation within the image (None vs. M: first fixation, mean difference: 0.24° , p > 0.20; second fixation, mean difference = 0.62, p = 0.001; third fixation, mean difference = 0.60, p =0.001).



Figure 11. The squared partial correlations of the fixation mode locations (separately for the x and y coordinates) for each individual cue with the fixation mode locations of the scenes with all three cues. Error bars represent the inner 68.29% of the distribution of partial correlations for each cue from 10,000 bootstrap resampled linear regression models.

Relating eye movements in scenes with a single contextual cue to eye movement behavior in scenes with all contextual cues

In order to further quantify the relative degree to which each contextual cue was being utilized for fixation guidance across fixations, we explored how well the x and y coordinates of observer fixations to scenes with single cues could predict the x and y coordinates of the fixations in scenes with all three cues. This will help us understand the individual contributions of each cue to eye movement guidance (relative to the guidance demonstrated when all cues were present) fixation-by-fixation.

For each fixation allowance condition, and contextual cue condition, we calculated the mode of the xcoordinate of all observers' closest fixations to the target or expected target location (for target present and absent trials, respectively) for each image. We did the same to obtain a *v* coordinate mode for each fixation allowance and contextual cue condition across images. In this way, we used the x coordinate fixation modes for the O, M, and B conditions for each image to predict the x coordinate fixation modes for each image in the OMB condition in a general linear regression model (and similarly for the y coordinate) for each fixation allowance. Our expectation was that the amount of information contributed to eve movement guidance during a particular fixation allowance for a given cue type will be captured by its squared partial correlation, i.e., its proportional contribution to the variance in fully cued fixations with the other cue contributions removed. Note that it is likely the case that eye movements between conditions may be collinear, so the model specified here may be underpowered, but this should not affect our interpretations of the partial correlations.

The Appendix shows a full table of zero-order correlations, partial correlations, standardized, and unstandardized model coefficients for each x/y coordinate and each fixation allowance condition. The proportion of variance accounted for in the fixations in scenes with all cues by the fixations in scenes with single cues (the coefficient of determination) was significant for both the x and y coordinates for all fixation allowances: one fixation, x: F(3, 41) = 12.60, $R^2 = 0.48, p < 0.001$; one fixation, y: F(3, 41) = 16.22, $R^2 = 0.54, p < 0.001$; two fixation, x: F(3, 41) = 44.80, $R^2 = 0.77, p < 0.001$; two fixations, y: F(3, 41) = 11.62, $R^2 = 0.46, p < 0.001$; three fixations, x: F(3, 41) = 58.03, $R^2 = 0.71$, p < 0.001; and three fixations, y: F(3, $(41) = 28.27, R^2 = 0.64, p < 0.001$. Plotted in Figure 11 are the squared partial correlations of each individual cue with x and y coordinates of fixations in scenes containing all cues. Error bars represent the inner 68.29% of the distribution of squared partial correlations for each cue from 10,000 bootstrap resampled linear regression models. Of note is the overall lack of explanatory power along the vertical dimension (y coordinate) by the object co-occurrence cue during the first fixation and by the background category across all fixations.

We performed a contrast-like analysis using the bootstrapped squared partial correlation distributions to assess the differences between the correlations for each condition. We calculated the difference of the summed x and y cue correlations between cues for each fixation allowance condition (or across fixation allowance conditions) and assessed the proportion of differences above or below zero (depending on the direction of the difference). The results demonstrate



Figure 12. The squared correlations of observers' expected target locations when cued with one type of contextual information with the expected target locations of observers viewing images containing all contextual cues (*x* and *y* coordinates considered separately). Error bars represent the inner 68.29% of the distribution of squared correlations for each cue from 10,000 bootstrap resampled correlations.

that the proportion of variance in eye movements within images containing all cues associated with the multiple object configuration cue is significantly greater than that associated with object co-occurrence and background category on the first fixation (M vs. B, p < 0.001; M vs. O, p = 0.046). Across all fixations, the multiple object configuration and object cooccurrence cues uniquely accounted for a greater proportion of the fully cued eye movement variability than the background category cue (M vs. B, p = 0.009; O vs. B = 0.001). Therefore, the multiple object configuration cue accounts for the most variance as compared to other cues in the scenes with all cues during the first fixation, and is the only cue to show diminishing explanatory power overall across fixations. The other cues generally plateau or increase in explanatory power across fixations, suggesting a differential utilization of cue information as time progresses.

Relating eye movement behavior in scenes with all contextual cues to a single cue's upper bound of target location informativeness

The previous section investigated the relationship between the eye movements in the condition with all cues to the eye movements with individual cues. Here we assess the temporal dynamics of the acquisition of information from each cue relative to the upper bound of information about expected target location available for each cue. From Experiment 1, we can calculate the mode of the expected target location of images containing a single cue, taken from observers who had unlimited time to study the image, foveate all regions, and make a selection. We can use this result as an upper bound of the information provided by each cue concerning the expected target location. We then calculated the correlation between the mode of observers' closest fixations to the target location for each fixation allowance with the mode of observers' expected target locations when viewing images containing a single cue (in both cases).

Figure 12 presents the squared correlations of each individual cue fixation mode to the individual cue expected target locations in the *x* and *y* coordinate space. Error bars represent the inner 68.29% of the distribution of bootstrapped squared correlations. The results indicate that object information is the only cue information to be increasingly extracted across fixations relative to the upper bound of information available (difference in r^2 between third and first fixations: *x* coordinate = 0.48, *z* = 3.17, *p* < 0.001; *y* coordinate = 0.36, *z* = 2.21, *p* = 0.01). All other cues were not significantly differentially utilized across fixations.

Discussion

Our analyses in this experiment were concerned with quantifying the time-course of the influence of three types of contextual cues on target detection performance and eye movement guidance to a target during a visual search task. We were also interested in assessing how each cue was combined using a signal detection theory framework.

To assess target detection performance we analyzed the index of sensitivity and bias across fixation allowance conditions. We found that across all

In addition, we have used three main analyses to understand the time-course of the influence of three types of contextual cues on the guidance of the first few fixations made during visual search for a target: (a) the change in the average distance of the closest fixation to the target for each fixation allowance condition, (b) the observed power of fixations made in scenes with a single cue to predict fixations made in scenes with multiple cues for each fixation allowance, and (c) the observed power of the locations where observers would expect a target to be located in scenes containing a single cue (the upper-bound of informativeness of a cue) to predict fixations made in scenes containing all contextual cues. In combination these analyses have shown us that overall, across fixations when only a single type of contextual information is available, guidance toward the target location is improved more by object information than by background information. When all cues are present, as in natural search, object configuration information is most utilized initially to guide eye movements to a proximate region where object co-occurrence information is then utilized for a finer search.

Finally, we have demonstrated that the combination of each cue's contribution to target detection performance as measured by the index of sensitivity is done consistently with what would be mathematically predicted by a linear combination of the cue information.

Experiment 3: Extractability of contextual cues in the periphery

Our final motivation was to understand how able observers are to extract each cue in the visual periphery. This may help clarify why certain cues appear to have less of an effect on visual search task performance. For example, there are two reasons why a cue may not be found to provide visual search guidance: (a) it simply does not provide information that is useful in facilitating visual search performance or (b) the cue could provide useful information, but is not easily extractable in the periphery, and is therefore never utilized as an information-providing source. We assessed each cue's extractability across the visual field by displaying images with or without each contextual cue present at various eccentricities from a fixation cross at which observers maintained their gaze. The observer's task was to indicate whether a particular cue was present in the image. To assess the possible

interaction of multiple cues, we also manipulated whether observers detected the presence of the cue of interest within an image containing no other additional cues or all other cues. We measured observers' cue detection performance as a function of eccentricity and whether the image contained no or all other cues.

Methods

Koehler & Eckstein

Participants

Undergraduate and graduate students (n = 360) at the University of California, Santa Barbara, participated in the experiment in exchange for course credit or cash payment. All participants provided informed written consent and were verified to have normal or corrected-to-normal vision.

Stimuli

The same base-set of scenes used in the definition verification task and in Experiment 2 were used for Experiment 3. Different stimulus sets were used for each between-subjects condition. In the condition where no other cues were present besides the cue that it was the participants' task to detect, O images were taken from scenes where the background was mismatched and all objects except the co-occurring object were jumbled. M images contained scenes with mismatched background and no co-occurring object, and B images were scenes with jumbled objects and no co-occurring object. When all other cues were present, all conditions used the scenes with all the cues. Due to the fact that participants were performing a task requiring them to judge the presence of a particular type of context, for each image, there was a complementary cue-absent version. To remove the O cue, we deleted the co-occurring object, to remove the M cue, we jumbled the objects, and to remove the B cue, we modified the background of the image to correspond to the mismatched (either the indoor or outdoor) category. The images were circularly cropped to a 700 pixel (11.9°) diameter and the targets from Experiment 2 were never present in the images. Each participant viewed a set of 45 images total. An example of each image type is shown below in Figure 13.

Design

We manipulated the type of contextual cue that participants were instructed to detect (three levels: O, M, or B, between-subjects), the eccentricity of the center of the image from the observers' fixation point (five levels: 0° , 4° , 8° , 12° , or 16° , within-subjects; see Figure 14 for a visualization), and whether all or no other types of contextual cues were present in the



Figure 13. Task instructions and sample stimuli for Experiment 3. The first two columns indicate the condition corresponding to the stimuli in the rightward columns and the specific task that participants performed in that condition. Overlaid on the possible stimuli are the correct responses to the task question. As indicated by the tasks, only one of the two images for each condition appeared on screen, chosen randomly with equal probability. Note the difference between stimuli for when all other cues are present alongside the cue that defines the observers' condition versus when no other cues are present alongside the cue relevant to the condition.





stimuli (two levels, between-subjects), resulting in a three-way mixed design. Different groups of sixty participants were randomly assigned to each of the experimental conditions. The eccentricity of the scene was again Latin-squares counterbalanced across levels using groups of nine images. Image presentation order was randomized.

Apparatus

Stimuli were displayed on a 3440×1440 pixel resolution LG 34UM95 LED monitor. Participants used a chin and forehead rest to stabilize their heads 77 cm from the monitor, resulting in a single pixel subtending 0.017° of visual angle. The same eyetracking equipment, settings, and procedures as Experiment 1 were again used to ensure that participants did not initiate any eye movements during the experimental trials.

Procedure

Participants were instructed that they would be viewing a series of images and determining certain properties about those images without moving their eyes. Every task required observers to make a yes/no judgment about whether a particular cue was present in the images, and they were informed that there was 50% likelihood that it would be present. Observers in the O condition had to determine whether the co-occurring object was present in the image. Observers in the M condition determined whether the objects in the image were jumbled. Observers in the B condition determined whether the background of the images matched a category cue.

The trial structure was similar to that depicted in Figure 4 for Experiment 3. Participants' initiated a trial by fixating a cross and pressing the space bar. To manipulate the distance of the fixation from the image, the initial fixation cross appeared in one of five locations. If participants were assigned to the O or B condition, after pressing the space bar, the cross would be replaced with the name of the object (e.g., BENCH) they were to search for or the category cue (e.g., INDOOR); they were to determine if the images matched, respectively. Participants in the M condition simply maintained fixation on the cross. After 500-1500 ms, if the fixation located was not within the image boundaries, the cross would reappear for 200 ms before the image appeared. If the fixation was located within the image boundaries, the fixation cross and cue would disappear 200 ms prior to the image appearance to eliminate masking effects. The image was displayed for 500 ms, after which a response screen appeared identical to that of Experiment 2 where participants indicated their confidence as to the presence or absence of the contextual cue. If observers broke fixation either because their fixation drifted outside a 3.5° radius around the indicated fixation location or if a saccade was detected, the trial was restarted. On trials where the observers broke fixation after the stimulus was presented, they were allowed to repeat the trial in an effort to give them more opportunities to learn to avoid broken fixation errors, but data from those trials was not included in analysis.

Statistical analyses

We analyzed how performance at the yes/no task changed as a function of image eccentricity between contextual cue conditions by measuring the proportion



Figure 15. Sensitivity index as a function of image eccentricity from fixation for each of the contextual cue conditions. Error bars represent an estimate of the standard error of the mean, as calculated from the sensitivity indexes delineating the inner 68.29% of the distribution of sensitivity indexes from 10,000 bootstrap resampled samples.

of trials in which participants correctly detected the presence of the contextual cue (PC). We conducted a three-way, mixed ANOVA and used false discovery rate controlled posthoc comparisons to assess pairwise differences and interpret significant interactions.

Results

Target detectability as a function of image center eccentricity

To test for differences between levels of eccentricity, type of contextual information, and the contextual manipulation being presented alongside all or none of the other contextual cues, we assessed the index of sensitivity and the proportion of trials in which the participant correctly detected the target. Figure 15 depicts the index of sensitivity. Similar to Experiment 1, it is difficult to assess interactions with our d' data because some observers had perfect hit or false alarm rates. Because there is clearer evidence of a significant interaction with these results, and the PC results (shown in Figure 16) are very similar to the index of sensitivity, we have focused our analyses on the PC results where we could statistically assess the interaction. We performed a three-way, mixed ANOVA with eccentricity as a within-subjects factor and context type and other context presence as between subject factors. All three main effects were significant: eccentricity, F(4,1416 = 49.47, p < 0.001; context type, F(2, 354) =200.63, p < 0.001; other context presence, F(1, 354) =6.22, p = 0.013. There was also a significant interaction between context type and other context presence, F(2,(354) = 5.29, p = 0.005, as well as between eccentricity and context type, F(8, 1416) = 2.76, p = 0.005. All other interactions were not significant.



Figure 16. Average proportion of trials where the participants correctly determined the presence/absence of the target as a function of image eccentricity from fixation for each contextual cue condition. Error bars represent the SEM.



Figure 17. Average proportion of correct judgments about target presence as a function of image eccentricity from fixation for each contextual cue condition, irrespective of the presence of other cue information, i.e., an illustration of the interaction between eccentricity and contextual cue type.

We performed posthoc comparisons while controlling the false discovery rate to interpret the main effects in light of the significant interactions (Figures 17 and 18). First, we wanted to understand whether performance at detecting each contextual cue changed as a function of eccentricity for each type of contextual cue (Figure 17). For all three cues, detection performance was significantly better for the nearest eccentricities than for the farthest (B, mean difference = 0.09, p <0.001; M, mean difference = 0.14, p < 0.001; O, mean difference = 0.17, p < 0.001). The slope of the drop-off for background category was shallower than that of multiple object configuration and object co-occurrence information. Second, we wanted to better understand the effect of the presence of other contextual cues on observers' performance (Figure 18). When no other contextual cues were present, participants performed

best at determining the presence of the background category cue, second best at detecting the co-occurring object, and third best at detecting multiple object configuration information (B vs. O, mean difference = 0.14, p < 0.001; B vs. M, mean difference = 0.23, p <0.001; O vs. M, mean difference = 0.09, p < 0.001). The pattern of results is identical when all other contextual cues were present as well (B vs. O, mean difference = 0.11, p < 0.001; B vs. M, mean difference = 0.17, p <0.001; O vs. M, mean difference = 0.06, p < 0.001). However, critically, the only contextual cue that was significantly affected by the presence or absence of other cues was multiple object configuration information (all vs. no other cues present, mean difference = 0.06, p < 0.001).



Figure 18. Average proportion of correct judgments about target presence as a function of contextual cue type depending on the presence of other cue information, irrespective of image eccentricity from fixation, i.e., an illustration of the interaction between the manipulated cue type and the presence of other information.

The need for validation of contextual cue manipulations

The majority of historical work looking at the influences of scene context on search and recognition has manipulated scene components based on intuitions by the investigators about what constitutes scene context. Here, we partitioned scene context cues into three specific components based on discussions by previous authors (Henderson & Hollingworth, 1999): object co-occurrence, multiple object configuration, and background. Our division of cues is clearly not unique or unequivocal but is a starting point to define reproducible scene context cues. Other researchers partition "scene context" differently: grouping background and structural object elements in a scene in one category and movable object information in another (Pereira & Castelhano, 2014).

Irrespective of the partition scene context cues, there is a clear need to independently assess context manipulations utilizing separate observers and tasks and obtain quantitative measures of the inherent information provided by a contextual cue about likely target locations. In recent years investigators have increasingly used separate judgments to measure how scene context contributes to observers' expectations of target locations (Droll & Eckstein, 2010; Preston, Guo, Das, Giesbrecht, & Eckstein, 2013; Torralba, Murphy, & Freeman, 2010). In this respect, the current work follows these studies by utilizing explicit judgments under unlimited time and foveal inspection to quantify the maximum informativeness that independent contextual cues can provide about likely target locations. Such independent measures help inform why a contextual cue might or not guide eye movements during search.

Our results also highlight some complexities: Different explicit judgments might not result in unanimous findings. For example, there was some disagreement between independent observers' ratings of the informativeness of the co-occurring object depending on a subtle difference in instructions. One group of observers, when asked to select an object in the scene that provides the most information about the location of a target object, selected different objects than did a group of observers who were asked to select the object in the scene that they would expect to be closest to a target object. Our assumption was that the former task instructions would indirectly assess participants' spatial expectations about object location relations, but that was not the case. Furthermore, from a second task where observers were instructed to select the location in images containing a single cue where they would expect a target object to be located, we were able to predict the fixations of observers searching for the same targets in an image containing all cues.

In addition, it is important for researchers to not only obtain observers' explicit judgments measuring the informativeness of a scene context cue but also to assess the reliability of the cue by making measurements of statistical relations between the targets and context cues in scenes (Greene, 2013, 2016). As scene databases expand, and the accurate labeling of those scenes becomes more feasible using microtask work forces such as Amazon Mechanical Turk, it is crucial to document and assess natural scene statistics and scene manipulations. One effort along these lines has been to quantify object-scene relations in two separate scene image databases (Greene, 2013), where it was also demonstrated that humans are prone to overestimate the frequency of a particular object being present in those scene images (Greene, 2016). The latter result could reflect a divide between real-world and imagebased object-scene relations, and over estimations of image statistics could arise from accurate estimations of real-world statistics. In summary, improved operationalization of context sources, measurements of observers' explicit judgments about expected target locations, and image analysis to document statistical relationships between targets and contextual cues in large datasets of scenes are all necessary for the scene context field.

Contextual cue combination

One objective of the current work was to isolate separable contextual cues and assess their contribution to search performance. One goal of the current work was to assess how the cues combine and compare to the classic model of statistically independent cue-combination that has been applied to spatial (Graham, 1989; Shimozaki et al., 2003), depth (Landy et al., 1995), and cross-modality cues (Ernst & Banks, 2002). We assessed whether the performance (d') benefit measured in humans was consistent with an optimal linear combination of each cues' independent effect on behavior. We found evidence that the observed performance of humans who viewed full cue images was consistent to that which would be expected by an optimal combination of single cues. This finding was consistent with the current data as well as for data from past work (Koehler & Eckstein, in press).

It is clear that combining two cues is not simply supplying redundant information to observers, and the data show additional benefits as additional contextual cues are added to the scenes. Yet, the statistically independent cue framework is certainly an oversimplification of the process by which each cue guides eye movements and covert attention toward likely target locations and increases the probability of detecting the target (Eckstein et al., 2006; Kanan, Tong, Zhang, & Cottrell, 2009; Torralba et al., 2006). In this sense, the consistency with optimal integration of independent cues is rather surprising. Literal interpretation of our result could suggest a modularity of organization in the brain for dealing with various scene and object-based contextual cues. Aside from the debate about whether the cues are statistically independent or not, the present analysis provides a benchmark about how multiple scene context cues benefit performance relative to a normative model that can be used as a reference to compare to future results assessing cue combination of scene context cues.

Contributions of individual contextual cues to target detectability and decision bias

We also sought to compare the contributions of separate well-defined cues to facilitating the search for targets in scenes and how it progressed through time. Past work that used contextual information in conjunction with saliency information to predict human fixations in a series of images during a target-search task demonstrated that scene context (as extracted from global image features) was influential in determining a region along the vertical dimension of an image (that spanned the entire width of the image) in which a target was likely to be located (Torralba et al., 2006). Pereira and Castelhano (2014) found that background information and object information extracted from the periphery interact such that background information provide coarse guidance of eye movements to regions likely to contain a target, whereas object information provides more localized information about where to search. Additional studies have also highlighted an interaction between background and object information (Davenport & Potter, 2004; Joubert et al., 2007; Võ & Schneider, 2010). Crucially, many past studies that have investigated the interaction between "scene"-or "global"-level information and "local"-or "object"-level information have utilized a flash-preview, moving-window paradigm where observers received a quick preview of a scene and subsequently searched the scene with access only to the information within a small window surrounding their gaze location (Monica S. Castelhano & Heaven, 2011; M. S. Castelhano & Henderson, 2010; Võ & Henderson, 2011; Võ & Schneider, 2010). It has been shown that the information gleaned from a flashpreview of a scene has no effect eye movements past the second fixation when observers have subsequent full visual access to scene information (as they would naturally during search; Hillstrom, Scholey, Liversedge, & Benson, 2012). Here we systematically evaluated the contributions of three separate contextual cues to target detectability, decision bias, and eye movement guidance during a more naturalistic search task.

In general terms we find that object-related contextual cues (multiple of configuration and object cooccurrence) contributed the most to eve movement guidance and behavioral decisions. For target detectability, we found that object-related cues contributed the most to increasing target detectability (see Figure 6) although only the increase due to multiple object configuration reached statistical significance. Observer decision bias to report the target as absent was reduced the most by the presence of the object cooccurrence cue and approached the decision criterion utilized in scenes with all cues, both of which were closer to the optimal criterion for 50% target prevalence. Eye movement guidance as measured by the fixation distance to the target location was also smallest for scenes containing multiple object configuration or object co-occurrence information. In the absence of the target in the scene, it was the presence of the co-occurring object that guided eye movements to be closest to what observers considered the most probable location of the target as judged in scenes with all contextual cues (and no target) under unlimited time. The correlation of fixation locations from scenes with individual cues and those from scenes with all cues (as well as the correlation with explicit judgments of expected locations) also suggested the importance of object-based cues in guiding eve movements.

Across all our measures, the results demonstrate that background category information alone provides the least guidance in observers' fixations and contributions to behavioral decision performance. Of course, the extent of this conclusion might vary with the type of scenes. For example, you would expect background information to be much more useful in helping an observer localize an airplane, which will typically be found in easily identifiable sky regions, than a pencil, which will be easier to localize relative to other objects with which it frequently co-occurs. Our findings suggesting little contribution of background information is consistent with the guidance of attention between semantically related objects on an image memory task (Wu, Wang et al., 2014). The results might seem to be at odds with the Pereira and Castelhano study (2014), which found that fewer required fixations and shorter search times overall to localize a target when observers had access to background information (referred to in the study as "scene context") than when they had access to object information (referred to as "object content"). One important distinction is that Pereira and Castelhano

manipulated the presence of object information, but not the relative configuration of objects, nor the inclusion of objects that frequently co-occur with search targets in real scenes. Thus, the objects in their study might have not provided as much contextual information about the location of the target as the objects utilized in the present study.

Temporal dynamics of the influences of contextual cues

An important objective of the current study was to assess the temporal dynamics of how different cues influenced search utilizing a saccade contingent display. A study by Spotorno, Malcolm, and Tatler (2014) investigated how target template specificity and the consistency of an object with the contextual information provided by other scene elements affected eve movements across three temporal epochs of visual search: initiation, scanning, and verification. They found that the visual system employs contextual information early during the initiation of visual search. Our work builds upon theirs by decomposing scene context into multiple contextual cues and assessing their influence at a finer scale during search initiation and scanning (the first few eye movements of visual search) using a restricted fixation allowance paradigm. We found that the performance benefits (d') from each cue across fixations increased similarly for each cue (i.e., there was no interaction), and that object-based cues (multiple object configuration and object cooccurrence) provided greater facilitation of search and perceptual performance behavior overall. We did not identify a significant interaction between the individual cues and fixation number on our measure of object detection performance and eye movement guidance on target present trials (index of detectability and distance of closest fixation to target location, respectively). This is perhaps not surprising because the presence of the target can become a major component in the eye movement guidance (Eckstein et al., 2007; Eckstein, Beutter, & Stone, 2001; Findlay, 1997; Malcolm & Henderson, 2009).

For target absent trials for which guidance is solely based on contextual cues, we did discover a significant interaction and significant differences between each cues' usefulness to guide eye movements toward locations judged likely to contain the target. The interaction between cue type and fixation allowance on the distance of an observers closest fixation to the expected target location revealed that observers fixated significantly closer to the target region across all fixations when provided with the object co-occurrence cue than when provided with no cue. Not until the second fixation did observers show similar benefits when provided with the multiple-object configuration cue. However, we also observed that second fixations within scenes containing only multiple-object configuration information account for the most variance of the second fixation locations (x and y coordinates) in scenes with all cues.

However, multiple object configuration cued fixations' explanatory power (of fixations with all cues) decreases as scene exploration unfolds, with object cooccurrence information providing the most explanatory power overall by the third fixation. We take this as evidence that the spatial configuration of objects is initially perceived and utilized by the visual system to guide eye movements to likely general target locations, at which point information about specific objects can be fully utilized to further localize the target. However, even if object co-occurrence is not fully utilized by in the second fixations, it still guides search because it is the contextual cue providing the most information about the likely target location.

The inability to fully utilize all available object cooccurrence information to guide the second saccade might seem at odds with the finding that the decision criterion is significantly more liberal upon the first fixation when object co-occurrence information is present in the scene. However, such apparent discrepancy can be explained by the fact that the second saccade is programmed based only on the processing of visual information up to 100 ms prior to the execution of the saccade (Caspi, Beutter, & Eckstein, 2004) whereas the perceptual decision made after the interruption of the display in our experiment is based on information from the additional 100 ms of visual processing that occurred after the command to execute the saccade. The longer processing time for the perceptual decision might explain the larger effect of object co-occurrence on decision bias relative to its influence on the second saccade endpoint. The average distance of the co-occurring object from the initial forced fixation location was 9.5°, and Experiment 3 demonstrated that observers are able to detect the cooccurring objects above chance up to 16° into the visual periphery. Thus, our results suggest that observers are detecting the co-occurring object in the periphery prior to the first fixation and using that information to optimally adjust their bias, but failing to fully utilize that information for target localization until the third saccade endpoint.

Extraction of contextual cues in the periphery as a possible bottleneck to eye movement guidance

There are number of possible explanations for the greater contribution of object-cues than scene back-

ground to the guidance of eye movements toward likely target locations. The first explanation is related to the inherent information provided by each contextual cue. Observers' explicit judgments of expected target locations support the idea that object-based contextual cues provide more information about the likely target location. However, an alternative explanation is that information about scene background is less available in the visual periphery relative to object based cues, thus representing a bottleneck for the eye movement system to guide eye movements based on scene background information. To assess whether inability to access information about a contextual cue in the visual periphery represents a bottleneck for utilization of the cue, we separately evaluated observer performance at detecting co-occurring objects, the configuration of objects, or background scene category as a function of retinal eccentricity. We found that observers are able to extract each type of contextual cue above chance from images presented at least up to sixteen degrees into the visual periphery. Note, for reference, that in the search experiment the average distance to the target location was approximately 9.5 degrees from the initial fixation and 6 degrees from the first fixation into the image. Critically, the background category was most easily extracted in the periphery. This finding suggests that peripheral extraction of background category information is not the bottleneck and explanation for its failure to facilitate visual search and its lesser contribution to guiding search. Instead, the results reinforce the conclusion that background category information is less informative for visual search guidance. A surprising result is that observers are more able to detect single objects than they are able to detect whether multiple object configurations have been jumbled in the periphery. This might seem to be contrary to the result from the eye movements in the search task that indicated that multiple object configuration was more available to guide the second saccade than object cooccurrence. It is possible that the utilization of multiple object configurations to guide search might not specifically rely on the prototypical arrangement of the objects. For example, center of mass saccades might be guided by objects that are in a prototypical arrangement or jumbled. In addition, observers in Experiment 3 (assessing ability to access contextual cues in the periphery) had 500 ms to view the scene stimuli. The aggregate time before the second saccade is approximately 400 ms (300 ms of processing available to the second saccade) which is somewhat shorter than the display time in Experiment 3. Thus, it is likely that the ability to access the cues in the search task is less than that measured in the assessment of peripheral extraction of contextual cues (Figures 16 and 17) until the second fixation within the image.

We also found that multiple object configuration is the only cue that is differentially detectable depending on whether the two other cues were also present in the image, suggesting that gleaning a perceptual sense of the structure of a scene is highly dependent on multiple cue types being present and also highlighting some of the limitations of the independent contextual cue framework. The results do indicate a possible useful role for background information. Multiple object configuration information was easier to detect when paired with background category and co-occurring object information. Given that the presence of a single additional object (the co-occurring object) among other jumbled objects likely does very little in helping observers determine whether the remaining objects are jumbled, it is likely that background information is contributing more to improved detection of the multiple-object configuration cue. This suggests that background information might facilitate the extraction of other cues. Although background information in isolation does not provide as much localization or perceptual performance benefit as object-based information, it certainly facilitates observers' ability to interpret the spatial arrangement of objects (when it is their explicit task to do so) and presumably then utilize object information for eye movement guidance.

Summary of conclusions

The general goal of the present work was to try to understand why and how different scene contextual cues are utilized by the brain to guide and facilitate visual search. Our experiments suggest the following time course of these influences during search: When the initial fixation is outside the scene (about 0.84 degrees from the edge of the scene), the first saccade is directed toward the center of the image. The second eye movement (after a first fixation of about 180 ms and based on information from only up to 100 ms of visual processing, i.e., 80 ms prior to the execution of the saccade; Caspi et al., 2004) is directed toward likely target locations and is more driven by configuration of multiple objects than by the presence of objects that co-occur spatially with the target. Later eye movements are guided by the cooccurring object. In general, object based contextual cues contributed more to the guidance of eye movements and target detectability than the scene background. The dissociation in the contributions of object-based versus background-based cues is not related to an inability to extract the cues in the visual periphery, but is a consequence of the inherent information about target location provided by the cues. When contextual cues are absent in the scene, observers tend to adjust their decision criterion more conservatively, reporting that the target is absent more frequently. Of the contextual cues,

the presence of the co-occurring object influenced the decision criterion the most and as early as after the first fixation within the image. Finally, an analysis within the theoretical framework of cue-combination of the joint contributions of the scene contextual cues to decision accuracy was quantitatively consistent with the optimal combination of statistically independent cues. Yet, experimental findings also suggest that there are interactions among the cues such as the presence of the background facilitating the extraction of information about the spatial configuration of objects. Together, our results contribute to the understanding of how the brain might utilize multiple sources of contextual information in scenes to guide eye movements and shape search decisions.

Keywords: scene context, visual search, scene perception, peripheral processing, contextual guidance

Acknowledgments

This work was supported by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office and by the Naval Air Warfare Center AD under Prime Grant No N68335-16-C-0028 and Mayachitra Incorporated. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

Commercial relationships: none.

Corresponding authors: Kathryn Koehler; Miguel P. Eckstein.

Email: koehler@umail.ucsb.edu; miguel.eckstein@ psych.ucsb.edu.

Address: Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA, USA.

Footnotes

¹We utilized re-sampling methods because some observers did not show errors preventing us from calculating d' and running an ANOVA. The resampling method should be more robust to departures from parametric assumptions although less powerful.

²Although the experimental design was repeatedmeasures, we have elected to analyze the design as if it were between subjects, sacrificing some experimental power, because there were many instances where for a given context type and fixation allowance there were either no target present or target absent trials (resulting in many empty cells in the repeated-measures design).

References

- Antes, J. R. (1974). The time course of picture viewing. Journal of Experimental Psychology, 103(1), 62–70, doi.org/10.1037/h0036799.
- Antes, J. R., Penland, J. G., & Metzger, R. L. (1981). Processing global information in briefly presented pictures. *Psychological Research*, 43(3), 277–292.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Biederman, I. (1972). Perceiving real-world scenes. *Science*, 177(4043), 77–80, doi.org/10.1126/science. 177.4043.77.
- Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research*, 49(24), 2992–3000.
- Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research*, 86, 35–42, doi.org/10.1016/j.visres.2013.04.006.
- 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.org/10.1037/0278-7393.18.3.531.
- Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision*, *9*(1):34, 1–9, doi:10.1167/9.1.34. [PubMed] [Article]
- 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.org/10.1037/0278-7393.32.4. 699.
- Burgess, A. (1985). Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation. JOSA A, 2(9), 1498–1507.
- Calvo, M. G., Nummenmaa, L., & Hyönä, J. (2008). Emotional scenes in peripheral vision: Selective orienting and gist processing, but not content identification. *Emotion*, 8(1), 68–80.
- Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences, USA*, 101(35), 13086– 13090, doi.org/10.1073/pnas.0305329101.
- Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements

guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, *18*(5), 890–896, doi. org/10.3758/s13423-011-0107-8.

- Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 753–763.
- Castelhano, M. S., & Henderson, J. M. (2010). Flashing scenes and moving windows: An effect of initial scene gist on eye movements. *Journal of Vision*, 3(9): 67, doi:10.1167/3.9.67. [Abstract]
- Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, *46*(24), 4118–4133, doi.org/10.1016/j. visres.2006.08.008.
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4(5), 170–178.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71.
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15(8), 559–564.
- De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*(4), 317– 329.
- Droll, J., & Eckstein, M. (2010). Expected object position of two hundred fifty observers predicts first fixations of seventy seven separate observers during search. *Journal of Vision*, 8(6):320, doi:10. 1167/8.6.320. [Abstract]
- Eckstein, M. P. (2011). Visual search: A retrospective. Journal of Vision, 11(5):14, 1–36, doi:10.1167/11.5.14. [PubMed] [Article]
- Eckstein, M. P., Beutter, B. R., Pham, B. T., Shimozaki, S. S., & Stone, L. S. (2007). Similar neural representations of the target for saccades and perception during search. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(6), 1266–1270, doi.org/10.1523/ JNEUROSCI.3975-06.2007.
- Eckstein, M. P., Beutter, B. R., & Stone, L. S. (2001). Quantifying the performance limits of human saccadic targeting during visual search. *Perception*, *30*(11), 1389–1401.
- Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Sci*-

ence, *17*(11), 973–980, doi.org/10.1111/j.1467-9280. 2006.01815.x.

- Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics* (pp. 569–593). New York: Springer.
- Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, 17(6–7), 945–978, doi.org/10. 1080/13506280902834720.
- Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. *Human Body Perception from the Inside Out*, 131, 105–131.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, 7(1):10, 1–29, doi:10.1167/ 7.1.10. [PubMed] [Article]
- Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, *37*(5), 617–631.
- Graham, N. V. S. (1989). Visual pattern analyzers. Retrieved from https://books.google.com/books? hl=en&lr=&id=O6FypXJk9e4C&oi=fnd&pg= PR11&dq=graham+1989+visual+pattern+ analyzers&ots=NYI8O2ZTDH&sig=OaZRLFhh_ DRVe452oiKSejnTzgs
- Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics (Vol. 1974). Retrieved from http://andrei.gorea.free.fr/Teaching_fichiers/ SDT%20and%20Psytchophysics.pdf
- Greene, M. R. (2013). Statistics of high-level scene context. *Frontiers in Psychology*, *4*, 777. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/ PMC3810604/
- Greene, M. R. (2016). Estimations of object frequency are frequently overestimated. *Cognition*, 149, 6–10.
- Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, 3(8), 837–843, doi. org/10.1038/77754.
- Henderson, J. M., & Hollingworth, A. (1999). Highlevel scene perception. *Annual Review of Psychology*, 50(1), 243–271.
- 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(1), 210–228, doi.org/ 10.1037/0096-1523.25.1.210.

- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues:
 Optimal cue combination. *Journal of Vision*, 4(12):
 1, 967–992, doi:10.1167/4.12.1. [PubMed] [Article]
- 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, doi.org/ 10.3758/s13423-011-0205-7.
- Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, 51(10), 1192–1205.
- Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cuing— configuration or individual locations? *Perception & Psychophysics*, 66(3), 454–463, doi.org/10.3758/BF03194893.
- Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal* of Vision, 3(7):4, 499–512, doi:10.1167/3.7.4. [PubMed] [Article]
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, 47(26), 3286–3297.
- Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, 17(6–7), 979–1003.
- Keysers, C., Xiao, D., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, 13(1), 90–101.
- Koehler, K., & Eckstein, M. P. (in press). Beyond scene gist: Objects guide search more than scene background. *Journal of Experimental Psychology: Human Perception and Performance*, in press, doi:10. 1037/xhp0000363.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35(3), 389–412.
- Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, 9(10):6, 1– 16, doi:10.1167/9.10.6. [PubMed] [Article]
- Levi, D. M. (2008). Crowding—an essential bottleneck for object recognition: A mini-review. *Vision Research*, 48(5), 635–654, doi.org/10.1016/j.visres. 2007.12.009.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, USA, 99(14), 9596– 9601, doi.org/10.1073/pnas.092277599.

- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. Journal of Experimental Psychology: Human Perception and Performance, 4(4), 565–572, doi.org/10.1037/0096-1523.4.4.565.
- Loschky, L., Boucart, M., Szaffarczyk, S., Beugnet, C., Johnson, A., & Tang, J. L. (2015). The contributions of central and peripheral vision to scene gist recognition with a 180° visual field. *Journal of Vision*, 15(12):570, doi:10.1167/15.12.570. [Abstract]
- Mack, S. C., & Eckstein, M. P. (2011). Object cooccurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision*, 11(9):9, 1–16, doi:10. 1167/11.9.9. [PubMed] [Article]
- Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics*, 2(11), 547–552, doi.org/10. 3758/BF03210264.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, 9(11):8, 1–13, doi:10.1167/9.11.8. [PubMed] [Abstract]
- Metzger, R. L., & Antes, J. R. (1983). The nature of processing early in picture perception. *Psychological Research*, 45(3), 267–274.
- Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, 6(2), 182–189, doi. org/10.1038/nn996.
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, 46(5), 614–621.
- Oliva, A. (2005). Gist of the scene. *Neurobiology of Attention*, 696(64), 251–258.
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527.
- Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, *9*(3), 273–302.
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3, 519–526.
- Pereira, E. J., & Castelhano, M. S. (2014). Peripheral guidance in scenes: The interaction of scene context

and object content. Retrieved from http://psycnet. apa.org/psycinfo/2014-31982-001/

- Potter, M. C. (1975). Meaning in visual search. *Science*, *187*(4180), 965–966.
- 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*, 33(18), 7846– 7855.
- Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, 42(11), 1447–1463, doi.org/10.1016/S0042-6989(02)00040-8.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatialscale-dependent scene recognition. *Psychological Science*, 5(4), 195–200.
- Shimozaki, S. S., Eckstein, M. P., & Abbey, C. K. (2003). An ideal observer with channels versus feature-independent processing of spatial frequency and orientation in visual search performance. *Journal of the Optical Society of America A*, 20(12), 2197–2215, doi.org/10.1364/JOSAA.20.002197.
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in realworld scenes. *Journal of Vision*, 14(2):7, 1–21, doi: 10.1167/14.2.7. [PubMed] [Article]
- Strasburger, H., Rentschler, I., & Jüttner, M. (2011).
 Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5):13, 1–82, doi:10.1167/11.5.
 13. [PubMed] [Article]
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.
- Torralba, A., Murphy, K. P., & Freeman, W. T. (2010). Using the forest to see the trees: Exploiting context for visual object detection and localization. *Communications of the ACM*, 53(3), 107–114, doi.org/ 10.1145/1666420.1666446.
- 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.
- Trommershauser, J., Kording, K., & Landy, M. S. (2011). Sensory cue integration. Retrieved from https://books.google.com/books?hl=en&lr=&id= M41pAgAAQBAJ&oi=fnd&pg=PP1&dq= landy+cue+combination&ots=afosw5upGZ&sig= hHK_BCkrtTnxeSENN3IFdvtsQbI
- Võ, M. L.-H., & Henderson, J. M. (2011). Object-scene

inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics, 73*(6), 1742– 1753.

- Võ, M. L.-H., & Schneider, W. X. (2010). A glimpse is not a glimpse: Differential processing of flashed scene previews leads to differential target search benefits. *Visual Cognition*, 18(2), 171–200, doi.org/ 10.1080/13506280802547901.
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15(4), 160–168, doi.org/10.1016/j.tics.2011.02.005.
- Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238.
- Wu, C.-C., Wang, H.-C., & Pomplun, M. (2014). The roles of scene gist and spatial dependency among objects in the semantic guidance of attention in real-world scenes. *Vision Research*, 105, 10–20. https://doi.org/10.1016/j.visres.2014.08.019
- Wu, C.-C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, 5. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/ articles/PMC3915098
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*(4), 787–835.
- Zelinsky, G. J., Zhang, W., Yu, B., Chen, X., & Samaras, D. (2005). The role of top-down and bottom-up processes in guiding eye movements during visual search. *In Advances in neural information processing systems* (pp. 1569–1576). Retrieved from http://machinelearning.wustl.edu/ mlpapers/paper files/NIPS2005 727.pdf

Appendix

Combination of statistically independent cues

There is a well-established literature of the expected performance benefits of combining statistically independent visual cues based on signal detection theory (Green & Swets, 1966; Landy et al., 1995; Shimozaki et al., 2003; Trommershauser et al., 2011). For thoroughness we outline the derivation of the relationships between single cue and multiple cue performance for our assessment of the combination of contextual cue information. We assume that each scene will elicit an internal response in the observer, x_{cue} , the mean of

which is dependent upon whether the target is present in the image and subject to internal noise. The internal response of an observer to a target absent (noise) and target present (signal) scene can be represented by Gaussian probability distributions. The observers' ability to discriminate between target-present and target-absent images is represented by the index of sensitivity, d', and is equivalent to the difference in mean internal responses to the signal and noise response distributions, divided by the standard deviation of the of the internal response, assumed to be equal for the signal and noise distributions (Equation A1).

$$d'_{x} = \frac{\langle x_{s} \rangle - \langle x_{n} \rangle}{\sigma_{x}}$$
 (A1)

For simplicity, we assume that the mean internal response to a target absent image containing a given cue, $\langle x_{cue}, n \rangle$ is zero. We assume that the mean of the internal response distribution to a target varies with the presence of different cues (due to covert or overt attention mechanisms) and is given by, $\langle x_{cue}, s \rangle$, and an associated target detectability, d'_{cue} . When multiple cues are present in the image, we assume that the internal responses from various cues are combined optimally assuming statistical independence. In such case the optimal linear integration of internal responses to the individual cues (Equation A2).

$$y = w_1 x_{\text{cue}_1} + w_2 x_{\text{cue}_2} + w_3 x_{\text{cue}_3}$$
 (A2)

Using Equation A1, we can derive the index of detectability for the combine cues, d'_y , given the mean internal responses to the signal and noise only $\langle y_s \rangle$, $\langle y_n \rangle$, and its standard deviation, σy . We can calculate the expected value of the signal and noise distributions of y using Equations A3 and A4.

$$\langle y_n \rangle = w_1 \langle x_{\text{cuel},n} \rangle + w_2 \langle x_{\text{cue},n} \rangle + w_3 \langle x_{\text{cue},n} \rangle = 0$$
(A3)

$$\langle y_s \rangle = w_1 \langle x_{\text{cuel},s} \rangle + w_2 \langle x_{\text{cue},s} \rangle + w_3 \langle x_{\text{cue},s} \rangle \quad (A4)$$

We assume that $\langle x_{cue}, n \rangle = 0$, therefore $\langle y_n \rangle = 0$. For the statistically independent cues with unit variance, the optimal weighting is known to be $w_{cue} = d'_{cue}$.(proportional to the information provided by those cues represented by the index of sensitivity for that cue (Green & Swets, 1966). Again, assuming unit variance, $\langle x_{cue}, s \rangle = d'_{cue}$. Replacing w_{cue} and $\langle x_{cue}, s \rangle$ in Equation A4. results in

$$\langle y_s \rangle = d_{cue_1}^{\prime 2} + d_{cue_2}^{\prime 2} + d_{cue_3}^{\prime 2}$$
 (A5)

Finally, the standard deviation of y is derived using error propagation and partial derivatives in Equation

A6 and solved in A7, noting that we assume unit variance.

Koehler & Eckstein

$$\sigma_{y}^{2} = \left[\frac{\partial y}{\partial x_{\text{cue}_{1}}}\right]^{2} \sigma_{x_{\text{cue}_{1}}}^{2} + \left[\frac{\partial y}{\partial x_{\text{cue}_{2}}}\right]^{2} \sigma_{x_{\text{cue}_{2}}}^{2} + \left[\frac{\partial y}{\partial x_{\text{cue}_{2}}}\right]^{2} \sigma_{x_{\text{cue}_{2}}}^{2}$$
(A6)
$$\sigma_{y} = \sqrt{d_{\text{cue}_{1}}^{\prime 2} + d_{\text{cue}_{2}}^{\prime 2} + d_{\text{cue}_{2}}^{\prime 2}} \quad (A7)$$

Returning to Equation A1, using Equations A3, A4, and Equation A7, we can derive the optimal linear additive combination of d',

$$d'_{y} = \sqrt{d'_{\text{cue}_{1}}^{2} + d'_{\text{cue}_{2}}^{2} + d'_{\text{cue}_{3}}^{2}}$$
 (A8)

Equation A8 is the known relation between the index of detectability of multiple combined statistically independent cues and the individual d's of each cue. The d'of each cue is the contribution of that cue to target detectability. In our experiment, we measure the d'associated with scenes with a specific cue (object cooccurrence, multiple object configuration, or background category) and the presence of the target. To isolate the contribution of the cue to the target detectability, we assume that the presence of a target is another statistically independent source of information that is combined with the presence of a contextual cue. Thus, target detectability (d') in a scene with the target present and a contextual cue $(d'_{cue,none})$ is given by the combination of the d' in a scene with only the target and no contextual cue (d'_{none}) and the d' contribution of the cue (d'_{cue}) :

$$d'_{\rm cue,target} = \sqrt{d'_{\rm target}^2 + d'_{\rm cue}^2} \quad (A9)$$

Both $d'_{cue,none}$ (target detectability from scenes with one cue and the target) and d'_{none} (target detectability from scenes with the target but no cues) are obtained from the experiments and the contribution to d' from the contextual cue (d'_{cue}) can be estimated by solving Equation A9:

$$d'_{\rm cue} = \sqrt{d'^2_{\rm cue,target} - d'^2_{\rm target}}$$
 (A10)

A d'_{cue} can be estimated for each contextual cue from Equation A10. Predictions from the combination of multiple cues can be then calculated for each of the combined cue experimental conditions (OM, OB, MB, and OMB) including the contribution of the target (d'_{none}) . Equation A11 shows the prediction for the scenes with all cues. Similarly, one can predict the effective d' from the presence of two cues.

$$d_{\text{cue}_{1,2,3},\text{target}}^{\prime 2} = \sqrt{d_{\text{cue}_{1}}^{\prime 2} + d_{\text{cue}_{2}}^{\prime 2} + d_{\text{cue}_{3}}^{\prime 2} + d_{\text{target}}^{\prime 2}} \quad (A11)$$

Tables of linear regression results

The tables below present the results of using the mode of closest fixations to target and expected target locations x and y coordinates in the individually cued conditions to predict the same for the fully cued condition. See the main text for highlights of the results concerning the partial correlations, but refer below to assess the zero-order correlations and model coefficients.

*significant at the 0.05 level **significant at the 0.01 level ***significant at the 0.001 level

One fixation allowance, using singly cued fixations to predict fully cued fixations:

	. <u></u>		Zero	o-order <i>r</i>								
	M		В		OMB		β		pr ²		b	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
О М	0.25*	0.42**	0.62*** 0.14	0.54*** 0.30*	0.58*** 0.51***	0.35** 0.73***	0.48 0.39	0.01 0.70	0.46 0.46	0.01 0.68	0.50** 0.61**	0.01 0.73***
В				R ² :	0.36** 0.48***	0.31* 0.54***	0.01	0.10	0.01 Int	0.12 ercept:	0.01 -73.84	0.09 59.04

Two fixation allowance, using singly cued fixations to predict fully cued fixations:

		order <i>r</i>										
	M		В		OMB		β		pr ²		b	
	X	Y	x	Y	x	Y	X	Y	X	Y	X	Y
0 M	0.56***	0.38**	0.50*** 0.55***	0.62*** 0.43**	0.80*** 0.72***	0.66*** 0.39**	0.54 0.34	0.64 0.17	0.67 0.48	0.56 0.21	0.63*** 0.55***	0.50*** 0.17
В				R^2 :	0.60*** 0.77***	0.39** 0.46***	0.15	-0.08	0.24 In	-0.08 tercept:	0.27 —214.60	-0.07 141.85

Three fixation allowance, using singly cued fixations to predict fully cued fixations:

			Zero-o	order r								
	M		В		ОМВ		β		pr ²		b	
	X	Y	X	Y	x	Y	X	Y	X	Y	X	Y
0 M	0.84***	0.56***	0.64*** 0.70***	0.75*** 0.54***	0.86*** 0.85***	0.80*** 0.60***	0.44 0.36	0.66 0.23	0.48 0.39	0.59 0.31	0.52** 0.46*	0.65*** 0.24*
В				R ² :	0.71*** 0.81***	0.64*** 0.67***	0.18	0.02	0.28 Int	0.03 ercept:	0.22 	0.02 24.27