

Rapid scene perception with tragic consequences: observers miss perceiving vulnerable road users, especially in crowded traffic scenes

Thomas Sanocki · Mohammed Islam ·
Jonathan K Doyon · Chanyoung Lee

Published online: 14 March 2015
© The Psychonomic Society, Inc. 2015

Abstract How does scene complexity influence the detection of expected and appropriate objects within the scene? Traffic research has indicated that vulnerable road users (VRUs: pedestrians, bicyclists, and motorcyclists) are sometimes not perceived, despite being expected. Models of scene perception emphasize competition for limited neural resources in early perception, predicting that an object can be missed during quick glances because other objects win the competition to be individuated and consciously perceived. We used pictures of traffic scenes and manipulated complexity by inserting or removing vehicles near a to-be-detected VRU (crowding). The observers' sole task was to detect a VRU in the laterally presented pictures. Strong bias effects occurred, especially when the VRU was crowded by other nearby vehicles: Observers failed to detect the VRU (high miss rates), while making relatively few false alarm errors. Miss rates were as high as 65% for pedestrians. The results indicated that scene context can interfere with the perception of expected objects when scene complexity is high. Because urbanization has greatly increased scene complexity, these results have important implications for public safety.

Keywords Scene perception · Object recognition · Attention · Traffic crashes

T. Sanocki (✉) · M. Islam · J. K. Doyon
Department of Psychology, University of South Florida, PCD 4118,
Tampa, Florida 33620-8200, USA
e-mail: Sanocki@usf.edu

C. Lee
Center for Urban Transportation Research, University of South
Florida, Tampa, Florida 33624-8200, USA

A central domain for perceptual science is the perception of everyday scenes and the objects within them. Research to date has established that the meaning of novel but typical scenes can be perceived remarkably quickly (e.g., Greene & Oliva, 2009; Kirchner & Thorpe, 2006; Potter, 1976) and can facilitate the perception of scene-congruent objects (e.g., Biederman, 1981; Boyce, Pollatsek, & Rayner, 1989; Davenport & Potter, 2004). However, a different relation between scene context and objects has been observed in research on the perception of traffic scenes and crashes. For many years, traffic researchers have documented that drivers sometimes fail to consciously perceive vulnerable road users (VRUs—pedestrians, bicyclists, and motorcyclists), with tragic consequences (e.g., Ernst, 2011; Forman, Watchko, & Seguí-Gómez, 2011; Frumkin, 2002; Wulf, Hancock, & Rahimi, 1989). In lab situations in which the explicit task is VRU detection, such misperceptions are misses to expected target categories (e.g., Borowsky, Oron-Gilad, Meir, & Parmet, 2012; Gershon & Shinar, 2013; Pinto, Cavallo, & Saint-Pierre, 2014). We suggest that this is an effect of scene complexity during rapid scene perception, stemming from limitations in the ability of early vision to individuate multiple objects (e.g., Franconeri, Alvarez, & Cavanagh, 2013; Sanocki, 1991; Xu & Chun, 2009).

Complex scene perception theory

Computational studies of scene perception reveal its high complexity (e.g., Granlund, 1999; Tsotsos, 1990, 2001). Complexity is high because stimulus information is locally ambiguous: Any given subarea of a scene (e.g., a receptive field or larger localized region) almost always has many interpretations in terms of possible objects and regions. Combining possibilities across a scene causes combinatorial explosion. Processing, time, and structural constraints are necessary to

disambiguate the scene information by reducing possible interpretations (Tsotsos, 1990; see also, e.g., Witkin & Tenenbaum, 1983).

Complexity is especially problematic in early vision, when the scene is parsed into meaningful objects and surfaces. Neurocomputational resources must be allocated wisely, and this involves favoring a few regions. The consequence is significant perceptual limitations and, in particular, a limitation on perceiving multiple objects during the early stages of scene perception (e.g., Franconeri et al., 2013; Sanocki, 1991; Xu & Chun, 2009). Thus, early vision can be viewed as a competition between candidate regions for individuation as an object (e.g., Itti & Koch, 2000; Xu & Chun, 2009; Yanulevskaya, Uijlings, Geusebroek, Sebe, & Smeulders, 2013). As we elaborate below, larger and more salient regions have an advantage in this competition, because they activate larger neural populations. VRUs, on the other hand, tend to be smaller (and slower) and have a disadvantage in the competition. Thus, complex scene perception theory predicts that VRUs are less likely to reach conscious awareness when competing objects are present.

We examined this idea by presenting pictures of traffic scenes briefly and to either side of fixation. The noncentral presentations represent the brief glances that drivers use to scan a traffic scene for hazards (Chapman & Underwood, 1998; Henderson & Hollingworth, 1998), and they place the critical stimulus in nonfoveal regions, where neural competition for individuation is strongest (e.g., Whitney & Levi, 2011). We manipulated scene complexity by adding automobiles near the VRU location. We used a signal detection approach in which failures to perceive VRUs were quantified as misses.

Research on driver perceptions of hazards

VRUs are one class of road hazard that must be detected and responded to quickly enough to enable appropriate action (e.g., Horswill & McKenna, 2004; Lee, 2008; Olson & Sivak, 1986; Shinar, 2007). Although a primary task for drivers is to stay on the road, they also need to scan locations on and near the road for hazards, by moving their eyes quickly around the traffic scene (e.g., Chapman & Underwood, 1998; Henderson & Hollingworth, 1998). Because drivers can only sample the visual field, VRUs will often not be foveated. Analyses of crash events (e.g., vehicles turning directly into motorcycles) support the idea that VRUs are often not consciously perceived by drivers (e.g., Borowsky et al., 2012; Gershon & Shinar, 2013; Pinto et al., 2014; Wulf et al., 1989). Traffic researchers have converged on the idea that VRUs are low in conspicuity—the ability to attract the observer’s attention (e.g., Langham & Moberly, 2003; Wulf et al., 1989). Recent research has examined methods for increasing conspicuity, such as bright outfits and unique lighting configurations, with

some positive results (e.g., Gershon & Shinar, 2013; Pinto et al., 2014).

Our proposal is that VRUs sometimes have such low conspicuity that they are not perceived at all in early vision; they are not individuated as objects, and thus never reach conscious awareness. This should be most likely with brief glances (early vision) and with complex scenes, because of the increased competition for object individuation. These ideas follow from models of the time course of scene perception.

Early scene perception

The perception of complex scenes unfolds over time. The overall meaning of a typical scene, termed gist (e.g., “I am in traffic,” “there is a mall”), is perceived quite rapidly, within a single glance (within about 150 ms; e.g., Greene & Oliva, 2009; Kirchner & Thorpe, 2006; Potter, 1976). However, further processing is necessary for information beyond gist, such as the identity and details of noncentral objects, relations between objects and surfaces, and atypical information. Often, such further processing does not get completed during brief glances, as has been indicated in a variety of paradigms (e.g., Biederman et al., 1988; Botros, Greene, & Fei-Fei, 2013; Fei-Fei, Iyer, Koch, & Perona, 2007; Franconeri, Scimeca, Roth, Helseth, & Kahn, 2012; Sanocki & Sulman, 2009; Treisman, 1988).

Recent models have described the time course of scene perception. Scene processing begins with a massive array of visual information that is sampled during eye fixations. The array is subdivided and grouped into potential objects and background structures during early and intermediate visual processing (e.g., Itti & Koch, 2000; Xu & Chun, 2009; Yanulevskaya et al., 2013). Determining object and region boundaries is complex, because stimulus information is locally ambiguous, as has been noted (e.g., Granlund, 1999; Tsotsos, 1990, 2001). The organizational processes are limited in capacity, and in early processing, four or fewer objects are correctly grouped and individuated within a scene (e.g., Franconeri et al., 2013; Xu & Chun, 2009). Conscious identification of an object requires individuation (e.g., Xu & Chun, 2009). Different regions of the processed scene compete with each other for individuation, and the likelihood that a region will win (be individuated and represented) increases with stimulus factors such as the region’s size and contrast from the background (e.g., Itti & Koch, 2000; Yanulevskaya et al., 2013). In regions away from the center of fixation (parafoveal and peripheral regions), the competition for individuation becomes especially high, because neural resources are increasingly scarce; as a result, objects crowd other objects and make them consciously imperceptible (e.g., Whitney & Levi, 2011). Because they recruit stronger neural responses, larger and more colorful objects dominate smaller objects, making the

smaller objects less likely to be consciously perceived (e.g., Itti & Koch, 2000; Sanocki, 1991). Instead, nondominant objects are replaced by neighboring objects or regularization of the larger surrounding region (e.g., Whitney & Levi, 2011).

VRU detection and scene complexity

Given that VRUs are generally smaller than automobiles, they should often fail to be individuated and perceived within a brief glance in complex stimulus conditions. Complexity was manipulated in the present experiments by using traffic scenes that were relatively empty, or moderately crowded, by automobile distractors (e.g., Fig. 1). The moderately crowded scenes were expected to impair VRU detection more than the uncrowded scenes, but the interesting issue was how perception would be impaired. Perception was measured and analyzed in terms of signal detection theory (SDT; D.M. Green & Swets, 1966), which separates *sensitivity* (here, the overall ability to discriminate VRU presence from absence) from *bias*, or changes in the criterion for interpreting the stimulus information (i.e., the amount of perceptual evidence needed for detecting the VRU). Crowding should decrease sensitivity and thereby increase errors. If crowding impaired perception in an *unbiased* manner, then it would increase errors of both types (missing VRUs, and mistakenly responding that VRUs are present, or false alarms). However, models of scene perception predict that crowding causes VRUs to not be individuated, and this should be manifested by a bias effect in which VRUs are often missed. False alarms are predicted to be

relatively low. Because the miss errors are caused by strained neural resources, the bias effect should occur even though observers are intentionally looking for VRUs.

The experimental paradigm is a visual search for an object (VRU) embedded in a natural scene with some nearby distractors (crowded condition) or no close distractors. Most search experiments measure reaction time, leaving the display on until the response (e.g., Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011). However, at least two studies have been reported with brief nonscene displays that probed early vision (Biederman, Blickle, Teitelbaum, & Klatsky, 1988; Cameroon, Tai, Eckstein, & Carrasco, 2004, Exp. 2). Neither of these studies obtained the bias effect with increasing distractors that we predict here. Cameroon et al. found that false alarms increased more than misses with display size. Biederman et al. found a high false alarm rate that increased with display size, and it was driven by their semantic manipulation. False alarms were high when the distractors were consistent with the searched-for target (from the same setting or scene type), but not when they were inconsistent. This bias effect is opposite to the increased misses predicted here. However, Biederman et al. used nonscene displays, in which the problems of individuation in early vision are greatly reduced. Our targets occurred on everyday streets that often had automobiles, and individuation was a perceptual challenge, especially in the crowded condition.

If VRUs are missed with brief glances, what conditions are necessary for them to be accurately perceived? Models of perception aver that if visual resources (attention and eye fixations) are drawn to the location of the VRU, then that region

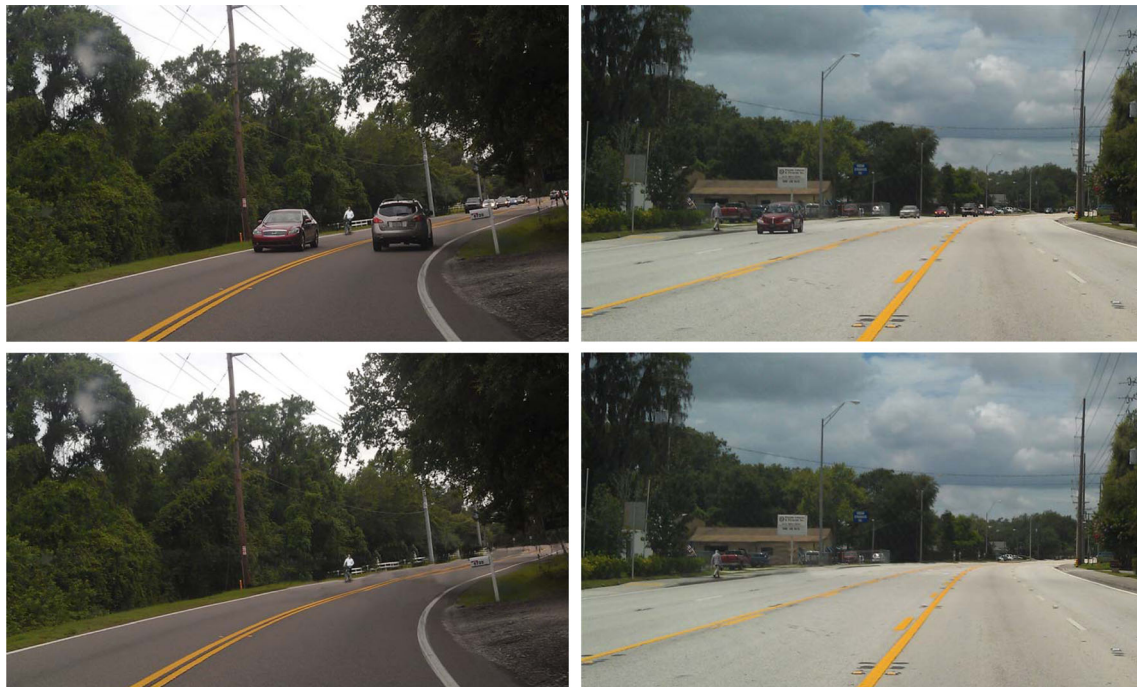


Fig. 1 Two traffic scenes in crowded and uncrowded versions; vulnerable road users are present in both

will receive fairly complete high-resolution processing, including individuation of its components. Studies of object search in natural scenes with long-duration displays have indicated that errors are low in such situations (e.g., Wolfe et al., 2011). Experiment 1B was a control experiment in which we examined the present stimuli under such conditions.

Experiments 1A and 1B

In the main experiment (Exp. 1A), the displays were designed to represent brief glances around a traffic scene for a VRU. Three types of VRUs were included, in a variety of typical (and legal) locations—a motorcyclist or bicyclist on the road, or a single pedestrian on the sidewalk or crosswalk.

The control experiment (Exp. 1B) utilized the same stimuli and design, but the trials were expanded to cause attention and the eyes to shift to the VRU's location. After the central fixation mark and before a particular scene, a cue (a red dot) was presented at the location of the VRU in the scene, for 250 ms. The scene followed immediately and remained on the screen until the response was made. Reaction time and overall accuracy were the dependent measures in this experiment. These conditions provided ample time for attention and eye fixations to the VRU region, and models of scene perception predicted high accuracy within reasonably fast reaction times.

Method

Overview The observers in Experiment 1A detected VRUs from brief glances (250 ms) of largish street-scene pictures, presented to the left or right of fixation. The scenes were *uncrowded* (no vehicles near or in front of VRU) or *moderately crowded* (one or more automobiles near VRU and others behind VRU; mean number of foreground automobiles = 2.1, range = 0 to 6). Figure 1 shows examples of crowded and uncrowded scene versions (VRU *present*). A single VRU appeared in a *present* scene; it could be in a variety of legal locations throughout the scene. Half of the scenes were *present* scenes, and the VRU was equally likely to be a pedestrian, bicyclist, or motorcyclist. The *absent* scenes were identical, except that the VRU was not present (scene background remained). Observers responded “yes” or “no” regarding the presence of a VRU in the scene, and received training with representative stimuli before testing. Experiment 1B had the same design, but the trial structure was expanded to bring attention and eye fixations to the location of the VRU.

Participants Thirty-four undergraduates who reported good or corrected-to-normal vision participated and received extra credit in Experiment 1A. The university is a large, pedestrian-friendly region, with reasonable numbers of bicyclists and motorcyclists. The data for two participants were not analyzed

because of low performance (<60% correct), leaving 12 males and 20 females. In Experiment 1B, a new sample of 12 undergraduates participated (eight females, four males).

Stimuli The test scenes were derived from 24 photographs of streets and intersections, taken from a viewpoint behind the steering wheel of an automobile (e.g., Fig. 1). The photos were all taken in daytime, with weather varying from sunny to rainy. Each of the 24 base scenes was modified to create four versions, defined by VRU presence or absence and crowding or noncrowding. Figure 1 shows the two *present* versions for two scenes. Vehicles were added or removed to produce the uncrowded and crowded versions. A different VRU was used with each of the 24 base scenes, and was identical between the crowded and uncrowded versions. The 24 different VRUs were extracted from the photographs and from Internet sources. Adobe Photoshop and ImageReady were used to remove or add VRUs and vehicles. The objects were added together with shadows and in an appropriate size and location. VRUs were located in legal positions on the road (motorcycles), the road or bike lane (bicyclists), or near the road on a sidewalk or crosswalk (pedestrians). Each image was saved as 740 × 443 pixels (18 × 11 deg).

Each block of trials consisted of the 24 scenes, with each scene appearing in one of the four versions. There were eight blocks of 24 trials each (two repetitions of each of the 96 versions). In Experiment 1A, trials began with a fixation cross in the center of the screen (500 ms), followed by a stimulus image (250 ms), with the image edge beginning 2 deg to the left or right of fixation. A blank screen then replaced the stimulus, and participants were instructed to indicate whether a VRU was present in the scene by using keyboard keys. A beep followed the response, but no further feedback was given. Testing was preceded by 24 trials of practice, generated from six additional scenes with four versions of each.

In Experiment 1B, the procedure was the same except for two additions to each trial. After fixation, a red dot appeared in the location of the VRU, to attract attention (250-ms duration, visual angle of approximately 0.49 deg, varying somewhat with position within the scene). In the VRU-absent scenes, the red dot appeared where the VRU was in the present version of the scene. The dot display was then replaced with the scene, which remained on until the response was made.

Results

Experiment 1A As desired, overall performance was in the moderate range, and responding was more accurate for uncrowded scenes (80%) than for crowded scenes (71%). Table 1 shows percentages correct (top matrix) and a recoding of the results as errors (bottom). The most important result was a bias

Table 1 Percentages correct (top) and error rates (bottom) in each condition of Experiment 1

	No VRU	VRU
Percentage Correct		
Uncrowded	84.0 (correct rejections)	76.5 (hits)
Crowded	83.5 (correct rejections)	58.5 (hits)
Error Rate (%)		
Uncrowded	16.0 (false alarms)	23.5 (misses)
Crowded	16.5 (false alarms)	41.5 (misses)

VRU, vulnerable road user

effect—an unequal distribution of the two possible types of errors. Across crowded and uncrowded scenes, observers often failed to detect VRUs (33% miss rate) while making fewer false alarms (16% mistaken reports of VRUs present). SDT provides a valid measure of sensitivity and the recommended estimate of bias (C ; Macmillan & Creelman, 1990). The overall bias estimate was $C = 0.31$ (standard error = 0.06), indicating a shift in decision criteria away from neutrality ($C = 0$) of about third of a standard unit, in the direction of requiring more evidence to detect a VRU. The overall bias was reliably greater than zero, $t(31) = 4.88$, $p < .001$, 95% CI [0.42, 0.19].

As predicted, the bias of failing to detect VRUs was stronger with crowded scenes (42% miss rate). There was also a smaller but reliable bias effect with uncrowded scenes (24% miss rate). The false alarm rates remained 16% in both conditions. The bias was reliable for uncrowded and crowded scenes considered separately {uncrowded $C = 0.19$, $t(31) = 2.60$, $p < .02$, 95% CI [0.33, 0.05], and crowded $C = 0.43$, $t(31) = 7.24$, $p < .001$, 95% CI [0.54, 0.31], respectively}. The bias was greater for crowded than for uncrowded scenes, $t(31) = 6.00$, $p < .001$. The bias effects are statistically independent of sensitivity, which was higher for uncrowded scenes ($d' = 1.90$) than for crowded scenes ($d' = 1.30$), $t(31) = 8.29$, $p < .001$.

A second recommended measure of bias is the overall percentage of *Yes* responses (Macmillan & Creelman, 1990), and this measure also showed strong bias in the crowded conditions, as well as some bias in the uncrowded condition (Table 2). Nonparametric estimates of bias (B'') and sensitivity (A') yielded the same conclusions, and also appear in Table 2. These results support the robustness of the conclusions across measures.

Table 2 Alternative measures of bias and sensitivity for Experiment 1

	Uncrowded	Crowded
<i>Yes</i> rate (bias)	46.3%	37.5%
B'' (bias)	0.199	0.339
A' (sensitivity)	0.880	0.805

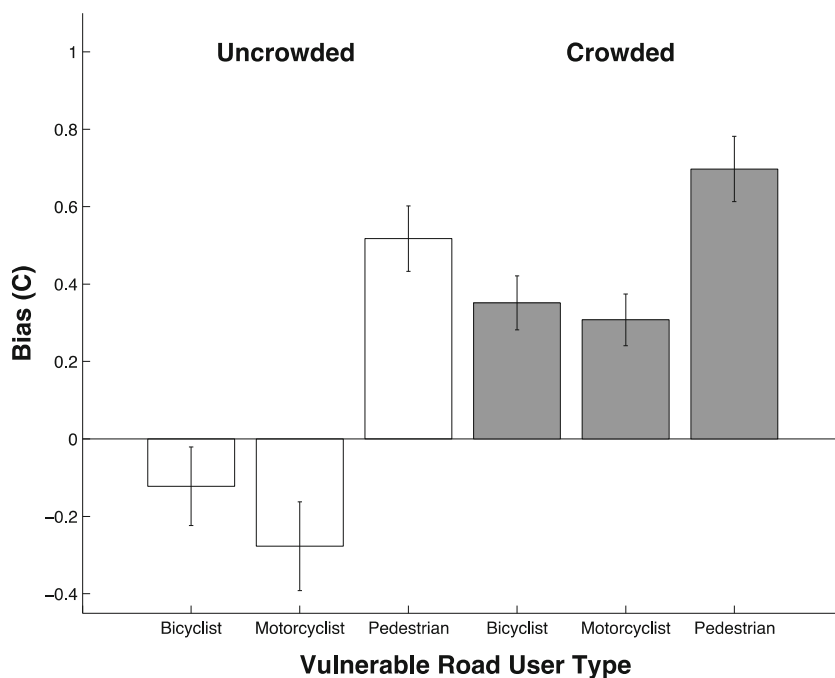
How consistent was the bias effect across the three different types of VRUs? Figure 2 shows bias estimates for each type of VRU. (This analysis used the miss rates for each VRU, considered against the common false alarm rate for all three types of VRUs.) Analysis of variance confirmed a strong interaction between VRU type and crowding, $F(2, 62) = 10.48$, $p < .001$. In crowded conditions, the bias effects were strong and reliable for all three types of VRUs. (We conducted t tests comparing the bias values to zero, and p values are reported in the figure caption.) In contrast, in the uncrowded conditions the bias effect occurred for pedestrians but not for bicyclists or motorcyclists. In fact, the bias for motorcyclists was reliably negative (more false alarms) in this experiment.

Note that differences between VRU types should be interpreted with some caution, because the exact differences between VRU types were not controlled; they might have differed overall on variables such as location, size, and contrast. Also, the crowding manipulation itself might vary somewhat with VRU type, because there was off-road clutter near the pedestrians. Nevertheless, we suggest two conclusions. First, the consistency of bias across VRUs in the crowded condition supports the idea that scene complexity contributes to the missing of VRUs; the effect is fairly general. Second, the bias goes away with empty roads for two VRU types because of reduced competition (reduced complexity) in early vision. Some bias remains with an empty road and a pedestrian VRU, but there may be several reasons for this. For instance, off-road locations may be lower in priority for attention than on-road locations, and, as we noted, additional objects in the off-road area might have competed with pedestrians (see, e.g., the second scene in Fig. 1). Perhaps both factors combine to cause bias in this condition.

Sensitivity and bias for the main conditions are visualized in terms of SDT in Fig. 3. In this figure, perceptual effects (the VRU signals) increase to the right along the axis. Bicyclists and motorcyclists have stronger perceptual effects, producing distributions more rightward from the noise (absent) distribution. The criterion (vertical line) is assumed to remain constant across VRU types but to vary with crowding. With uncrowded scenes, the perceptual effects are fairly strong and the criterion is close to neutrality (0 on the axis), producing bias only for the weaker pedestrian signals. With crowded scenes, the perceptual effects are less separated from noise because of scene complexity. The criterion shifts rightward, producing bias for all VRUs. The weakest signals in crowded scenes (pedestrians) are missed on a majority of trials.

Experiment 1B When visual resources were drawn to the location of the VRU, detection was accurate, exceeding 96% in all of the main conditions, as is shown in Table 3. The high accuracy levels indicate that the VRUs can be perceived if they are attended and if sufficient time is allocated for processing. Because accuracy was near ceiling, the accuracy data

Fig. 2 Bias effect in each condition of Experiment 1. Error bars are standard errors of the effect. In *t*tests against zero, the bias effects were reliable in five conditions, but not for uncrowded bicyclists ($p > .20$; for uncrowded motorcyclists, $p < .05$; for uncrowded pedestrians and all crowded conditions, $ps < .001$)



were not analyzed further. Reaction times (also in Table 3) indicate that VRU detection was faster than no-VRU responses, $F(1, 11) = 6.65, p < .01$, and that uncrowded responses tended to be faster than crowded responses, $F(1, 11) = 3.49, p = .09$. There was no interaction, $F(1, 11) = 1.45, p > .20$. The speed of VRU detection responses was close to that for brake light reaction times (e.g., M. Green, 2000; Shinar, 2007), supporting the representativeness of the present experimental situation.

Discussion

The results of Experiment 1A provide strong support for the prediction of a bias effect, in which VRUs are often missed when they appear among other vehicles (crowded conditions). The miss rates were high for all three types of VRUs, and especially for pedestrians. Crowding was predicted to cause misses by competing with the VRU for representation in early vision.

Fig. 3 Signal detection theory interpretation of perceptual effects on trials with no vulnerable road users (VRUs, dashed lines) and with VRUs (solid lines), in Experiment 1. The criterion is shown as the vertical lines

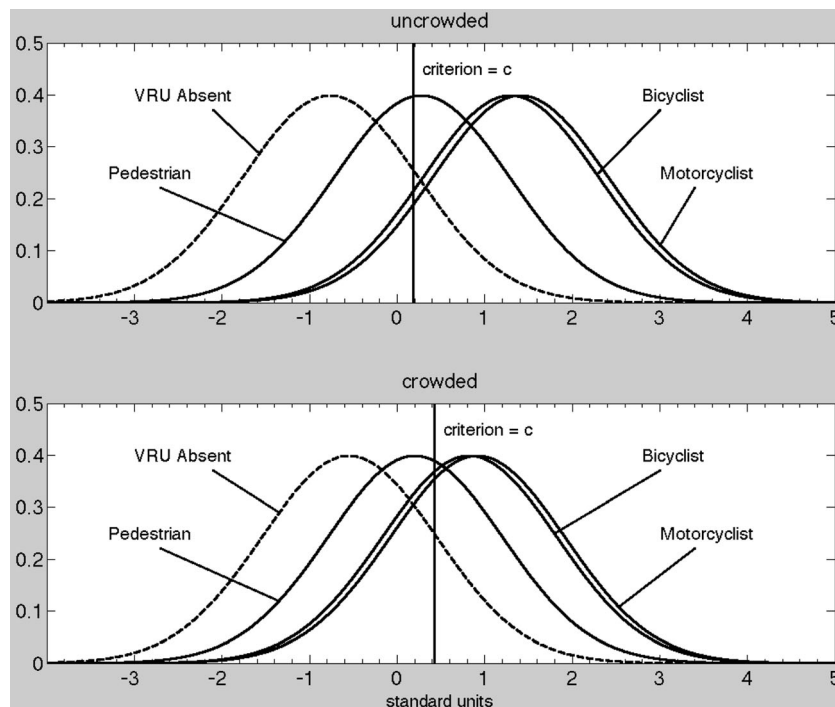


Table 3 Accuracy, in percentages correct, and reaction times (seconds, in parentheses) when observers could complete processing of the vulnerable road user (VRU; Exp. 1B)

	No VRU	VRU
Percentage Correct		
Uncrowded	96.7 (1.36)	96.5 (0.79)
Crowded	97.0 (1.40)	96.4 (0.91)

The results were different in the uncrowded condition. For bicyclists and motorcyclists on an empty road, there was a tendency *toward* VRU detection (and more false alarms). For pedestrians, however, there was still some bias against perception. Pedestrians may have been crowded by off-road objects, or the road may have received higher perceptual priority. Because differences between VRU types and their scenes were not carefully controlled, further research will be necessary to make precise conclusions about VRU type and other factors, such as size and location.

In contrast to the brief-exposure results, Experiment 1B indicated that, regardless of crowding, VRUs can be perceived accurately if attention and processing time are increased. The results suggest that the missing of VRUs can be corrected when resources are directed to VRU locations.

Experiment 2

The main purpose of Experiment 2 was to attempt to replicate Experiment 1A, while exploring pedestrian perception a little further. We examined whether a minor change in instructions would influence detection of pedestrians. Observer scanning behavior is known to vary with both expertise (driving experience) and type of traffic scene (e.g., Borowsky, Shinar, & Oron-Gilad, 2010). Could instructions influence how the somewhat ambiguous off-road areas are attended? As in Experiment 1, the observers were instructed to indicate whether the scene contained a VRU (defined as a pedestrian, bicyclist, or motorcyclist). However, in Experiment 1, the definition of VRUs noted that VRUs are encountered “on *or near* the road while driving.” This could have extended the observer’s attention to include off-road pedestrian locations. In Experiment 2, we used a slightly more narrow wording that may be representative of some drivers’ assumptions: VRUs are encountered “on the road while driving.” This could deemphasize pedestrians and further increase the bias against perceiving them.

Method

The stimuli, design, and procedure were the same as in Experiment 1A, except for the deletion of the two words from the instructions (“or near”). As noted, observers were again

instructed to respond to whether there was a VRU in the image scene. A new sample of 30 undergraduates who reported good or corrected vision participated and received extra credit. The data for three participants were not analyzed because their performance was near chance, leaving 19 females and eight males.

Results

Overall, accuracy was higher for uncrowded scenes (79%) than for crowded scenes (72%). Table 4 shows percentages correct and the results recoded as errors. The overall bias effect was somewhat larger in this experiment; observers often failed to detect VRUs (37% miss rate), while making fewer false alarms (11% mistaken reports of VRUs present). The overall bias estimate was 0.50 (standard error = 0.09), indicating a shift in decision criteria away from neutrality of about half a standard unit; $t(26) = 6.14$, $p < .001$, 95% CI [0.65, 0.35]. As predicted, the bias of failing to detect VRUs was strongest with crowded scenes (45% miss rate). However, there was also a fairly strong bias effect with uncrowded scenes (29% miss rate). The bias effects were reliable for both crowded scenes ($C = 0.64$), $t(26) = 7.61$, $p < .001$, 95% CI [0.81, 0.48], and uncrowded scenes ($C = 0.40$), $t(26) = 4.07$, $p < .001$, 95% CI [0.60, 0.21]. The bias was greater for crowded than for uncrowded scenes, $t(26) = 3.49$, $p < .002$, and sensitivity was higher for uncrowded ($d' = 1.99$) than for crowded ($d' = 1.56$) scenes, $t(26) = 3.34$, $p = .002$. These effects were also reliable with the alternative measures, reported in Table 5.

The bias effects are broken down by the three types of VRUs in Fig. 4. There was again a striking interaction of crowding and VRU type, $F(2, 52) = 10.56$, $p < .001$. In crowded conditions, there was a strong bias effect for each type of VRU, supporting the pervasiveness of the crowding effect. In uncrowded conditions, the bias effect was very strong for pedestrians but not different from zero for bicyclists or motorcyclists. We found no negative bias in the present experiment. Averaging across Experiments 1A and 2, the bias

Table 4 Percentages correct (top) and error rates (bottom) in each condition of Experiment 2

	No VRU	VRU
Percentage Correct		
Uncrowded	87.9 (correct rejections)	70.8 (hits)
Crowded	89.1 (correct rejections)	55.2 (hits)
Error Rate (%)		
Uncrowded	12.1 (false alarms)	29.2 (misses)
Crowded	10.9 (false alarms)	44.8 (misses)

VRU, vulnerable road user

Table 5 Alternative measures of bias and sensitivity for Experiment 2

	Uncrowded	Crowded
Yes rate (bias)	41.4%	33.1%
B'' (bias)	0.354	0.481
A' (sensitivity)	0.878	0.827

effect for motorcycles and bicycles on uncrowded streets was -0.07 ($SE = 0.06$).

How did the bias effects for pedestrians compare to those in Experiment 1A, where the instructions noted that VRUs could be near the road? The present bias effect was greater in magnitude [$t(57) = 1.80, p = .04$, one-tailed, for the comparison of bias effects between experiments]. The changed instructions may have further reduced the priority of off-road VRUs, and thereby increased misses. Even when the roads were empty in the uncrowded condition, the miss rate for pedestrians was 55% in this experiment. In the crowded condition, the miss rate for pedestrians reached 65%.

General discussion

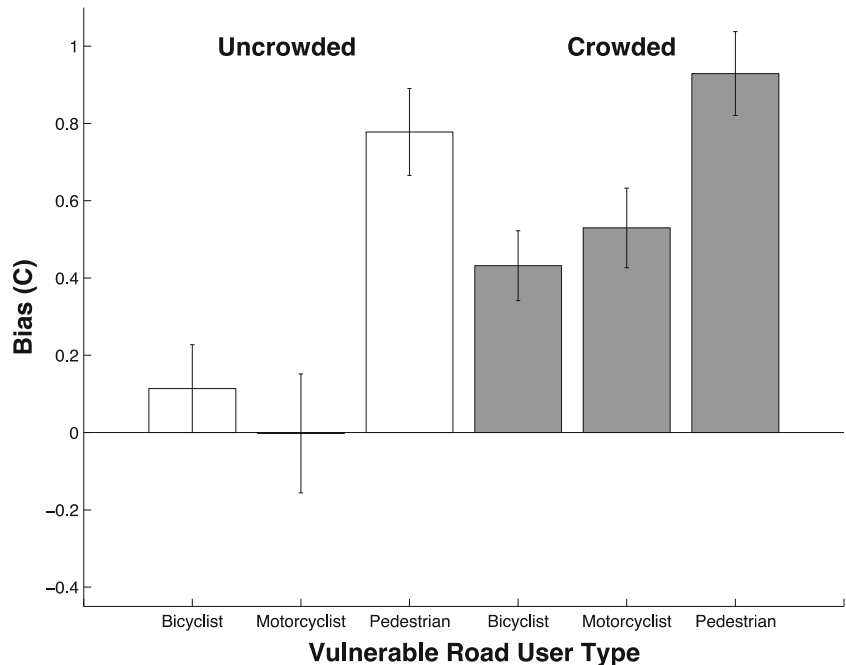
The results document an important limitation in human scene perception—an effect of scene complexity that causes expected objects not to be perceived. The findings are an ecologically significant new example of the limitations in everyday scene perception, and they inform the debate over the perceptual costs of everyday scene perception (e.g., Cohen, Alvarez,

& Nakayama, 2011; Kirchner & Thorpe, 2006) by documenting a critical cost of scene complexity.

The searched-for objects (VRUs) were always appropriate for the scenes, and they were also expected, because the observers’ sole task was to look for a VRU. Nevertheless, observers missed many of the VRUs. With complex (crowded) scenes, there were strong biases against perceiving all three types of VRUs, both on-road and off-road. As a result, the VRU was often missing from conscious perception, whereas the other error type (false alarms) was much less frequent. The result is consistent with the ideas that during rapid scene perception, observers have difficulty individuating multiple objects, and that VRUs often lose out in perception to other objects, such as vehicles. The other objects dominate conscious perception, and observers appear to think that no VRU is present. These results are consistent with a body of traffic research documenting the difficulty of detecting VRUs in pictures and in videos of traffic (e.g., Borowsky et al., 2012; Gershon & Shinar, 2013; Pinto et al., 2014). These biases against perception contrast with studies that have manipulated display size with brief nonscene displays (Biederman et al., 1988; Cameron et al., 2004). Perhaps the critical factor in producing the bias effect here is structural complexity near the target—other objects that compete for individuation.

In uncrowded scenes, there was no bias at all for bicyclists and motorcyclists; in fact, miss errors tended to be less frequent than false alarms. Thus, the bias effect went away when complexity was reduced (for bicyclists or motorcyclists on an empty road). Bias remained for pedestrians, possibly because off-road regions are deprioritized or because of increased

Fig. 4 Bias effect in each condition of Experiment 2. Error bars are standard errors of the effect. In t tests against zero, the bias effects were reliable in four conditions ($p < .001$), but not for uncrowded bicyclists or motorcyclists ($p > .20$)



clutter in off-road regions. The bias against pedestrian VRUs became stronger when the instructions did not mention off-road locations (Exp. 2). These bias effects have important implications for public safety, because urbanization continues to increase the complexity of traffic situations (see below).

One additional result was that, whereas errors were frequent with the brief presentations, accuracy reached high levels when observers had time to shift their eyes and attention to the location of the VRU (Exp. 1B). VRUs can be perceived with sufficient perceptual resources and time, even in crowded conditions.

The present bias effect should be distinguished from semantic congruency effects on object perception, which depend on the meaning of the scene and its relation to the target object (e.g., Hollingworth & Henderson, 1998; Palmer, 1975). The present bias effects occurred with scene-appropriate objects. The present bias effect can also be distinguished from another important misperception in scene perception, *inattentional blindness*. Inattentional blindness occurs when an irrelevant object is missed because the observer's attention is focused on other, highly relevant objects or tasks (e.g., Most, Scholl, Clifford, & Simons, 2005; Simons & Chabris, 1999; for relations to traffic crashes, see Chabris & Simons, 2009; Talbot, Fagerlind, & Morris, 2013). In contrast, in the present experiments, the sole task was to look for VRUs. The bias effect appears to be a separate and also important perceptual limitation that is pertinent to road safety. Chabris and Simons argued that inattentional blindness is an error of attention—if I didn't see it, it wasn't there. The present results extend this principle to misperception in general.

Rapid scene perception

How do the present results relate to evidence for rapid perception of scenes and objects in them (e.g., Greene & Oliva, 2009; Kirchner & Thorpe, 2006; Potter, 1976)? We suggest that the results are consistent with the idea of rapid scene categorization, but that rapid scene categorization is restricted to single salient interpretations (Sanocki & Sulman, 2013). The single interpretation (or gist) is registered across the scene in parallel, perhaps as a kind of global averaging or feature registration process (e.g., Epstein & MacEvoy, 2011; Greene & Oliva, 2009). We suggest that categorization is rapid and accurate only when there is high evidence for each possible interpretation on its respective trial type (e.g., “animal present” or “no animal present”). In the present experiments, accuracy was highest when there was a strong “VRU present” interpretation (“street containing *only* a motorcycle or bicycle”) or a strong “no VRU” interpretation (“empty street,” in the uncrowded absent VRU condition).

Scene categorization becomes much less efficient when different objects are in the scene (Walker, Stafford, & Davis,

2008), and it becomes extremely inefficient when there are competing categorizations (Evans, Horowitz, & Wolfe, 2011). The present crowded conditions would be difficult, because the “street with vehicles” and “street with vehicles and VRU” interpretations are perceptually close and compete with each other. Critically for the bias effect when a VRU was present, the competition for representation among objects results in the VRU often being absent in the early scene interpretation, supporting the “street with vehicles but no VRU” response. We suggest that the uncrowded pedestrian scenes were difficult in part because of the strong evidence from the empty street supporting the “no VRU” interpretation. Competition from off-road objects would also compete with pedestrian perception in this condition.

Rapid scene perception and public safety

Applied and experimental research approaches can converge on the problems of traffic safety (e.g., Shinar, 2007). The present results suggest that the visual complexity created by urbanization is an important factor for public safety, because it contributes to VRU misperception. There are both direct and indirect consequences of VRU misperception, which are reaching epidemic proportions as urban density continues to increase (e.g., Forman et al., 2011). Direct consequences include crashes and deaths that can occur when VRUs are not perceived. The indirect consequences, however, may be further reaching, more frequent, and more costly. The possibility of misperception and crashes changes behavior, reducing healthy and inexpensive behaviors such as walking and bicycling, and increasing expensive, less sustainable behaviors such as driving. Indirect consequences include reduced general health and increased type 2 diabetes (e.g., Ernst, 2011; Forman et al., 2011; Frumkin, 2002). We surveyed a sample of students from our university and found that more than 92% had at least some safety concerns about walking or bicycling across the streets that bound their university. The streets are often crowded with traffic, and the present results indicate that concerns about safety are well founded.

Traffic designers have used a number of measures that make pedestrians and bicyclists easier to perceive and avoid, such as sidewalks, bike lanes, and marked crosswalks. However, the present results indicate that these are not fully effective when observers look quickly in complex scenes. In the present experiments, bicyclists were missed despite being in bike lanes at the sides of roads. Motorcyclists were missed despite being at normal locations within traffic lanes. Further research on the perception and misperception of VRUs will be needed to explore the effectiveness of alternative solutions (e.g., Gershon & Shinar, 2013; Pinto et al., 2014). Also potentially important is training of driver attention, because

perception while driving is an active and strategic process (e.g., Borowsky et al., 2010).

References

- Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–263). Hillsdale: Erlbaum.
- Biederman, I., Blicke, T. W., Teitelbaum, R. C., & Klatsky, G. J. (1988). Object search in nonscene displays. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 456–467. doi:10.1037/0278-7393.14.3.456
- Borowsky, A., Oron-Gilad, T., Meir, A., & Parmet, Y. (2012). Drivers' perception of vulnerable road users: A hazard perception approach. *Accident Analysis and Prevention*, *44*, 160–166.
- Borowsky, A., Shinar, D., & Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving. *Accident Analysis and Prevention*, *42*, 1240–1249.
- Botros, A., Greene, M., & Fei-Fei, L. (2013). Oddness at a glance: Unraveling the time course of typical and atypical scene perception [Abstract]. *Journal of Vision*, *13*(9), 1048. doi:10.1167/13.9.1048
- Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 556–566. doi:10.1037/0096-1523.15.3.556
- Cameron, E. L., Tai, J. C., Eckstein, M. P., & Carrasco, M. (2004). Signal detection theory applied to three visual search tasks—Identification, yes/no detection and localization. *Spatial Vision*, *17*, 295–325.
- Chabris, C., & Simons, D. (2009). *The invisible gorilla*. New York: Random House.
- Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception*, *27*, 951–964. doi:10.1068/p270951
- Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural scene perception requires attention. *Psychological Science*, *22*, 1165–1172. doi:10.1177/0956797611419168
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*, 559–564. doi:10.1111/j.0956-7976.2004.00719.x
- Epstein, R. A., & MacEvoy, S. P. (2011). Making a scene in the brain. In L. Harris & M. Jenkin (Eds.), *Vision in 3D environments*. Cambridge: Cambridge University Press.
- Emst, M. (2011). *Dangerous by design 2011: Solving the epidemic of preventable pedestrian deaths*. Washington, DC: Transportation for America. Available at <http://www.smartgrowthamerica.org/documents/dangerous-by-design-2011.pdf>
- Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). When categories collide: Accumulation of information about multiple categories in rapid scene perception. *Psychological Science*, *22*, 739–746. doi:10.1177/0956797611407930
- 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. doi:10.1167/7.1.10
- Forman, J. L., Watchko, A. Y., & Seguí-Gómez, M. (2011). Death and injury from automobile collisions: an overlooked epidemic. *Medical Anthropology*, *30*, 241–246.
- Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. C. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*, *17*, 134–141. doi:10.1016/j.tics.2013.01.010
- Franconeri, S. L., Scimeca, J. M., Roth, J. C., Helseth, S. A., & Kahn, L. E. (2012). Flexible visual processing of spatial relationships. *Cognition*, *112*, 210–227. doi:10.1016/j.cognition.2011.11.002
- Frumkin, H. (2002). Urban sprawl and public health. *Public Health Reports*, *117*, 201–217.
- Gershon, P., & Shinar, D. (2013). Increasing motorcycles attention and search conspicuity by using Alternating-Blinking Lights System (ABLS). *Accident Analysis and Prevention*, *50*, 801–810. doi:10.1016/j.aap.2012.07.005
- Granlund, G. H. (1999). The complexity of vision. *Signal Processing*, *74*, 101–126.
- Green, M. (2000). “How long does it take to stop?”: Methodological analyses of driver perception–brake times. *Transportation Human Factors*, *2*, 195–216.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *20*, 464–472. doi:10.1111/j.1467-9280.2009.02316.x
- Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–293). Oxford: Elsevier.
- Hollingworth, A., & Henderson, J.M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, *127*, 398–415. doi:10.1037/0096-3445.127.4.398
- Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness* (pp. 155–175). Aldershot: Ashgate.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506. doi:10.1016/S0042-6989(99)00163-7
- Kirchner, H., & Thorpe, S. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*, 1762–1776.
- Langham, M., & Moberly, N. (2003). Pedestrian conspicuity research: A review. *Ergonomics*, *46*, 345–363. doi:10.1080/0014013021000039574
- Lee, J. D. (2008). Fifty years of driving safety research. *Human Factors*, *50*, 521–528.
- Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and “nonparametric” indexes. *Psychological Bulletin*, *107*, 401–413. doi:10.1037/0033-2909.107.3.401
- Most, S. B., Scholl, B. J., Clifford, E. R., & Simons, D. J. (2005). What you see is what you set: Sustained inattention blindness and the capture of awareness. *Psychological Review*, *112*, 217–242. doi:10.1037/0033-295X.112.1.217
- Olson, P. L., & Sivak, M. (1986). Perception–response time to unexpected roadway hazards. *Human Factors*, *28*, 91–96.
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*, 519–526. doi:10.3758/BF03197524
- Pinto, M., Cavallo, V., & Saint-Pierre, G. (2014). Influence of front light configuration on the visual conspicuity of motorcycles. *Accident Analysis and Prevention*, *62*, 230–237.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 509–522. doi:10.1037/0096-1523.2.5.509
- Sanocki, T. (1991). Intra- and interpattern relations in letter recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 924–941. doi:10.1037/0096-1523.17.4.924
- Sanocki, T., & Sulman, N. (2009). Priming of simple and complex scene layout: Rapid function from the intermediate level. *Journal of*

- Experimental Psychology: Human Perception and Performance*, 35, 735–749.
- Sanocki, T., & Sulman, N. (2013). Complex, dynamic scene perception: Effects of attentional set on perceiving single and multiple event types. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 381–398. doi:10.1037/a0030718
- Shinar, D. (2007). *Traffic safety and human behavior*. Burlington: Emerald.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28, 1059–1074. doi:10.1068/p2952
- Talbot, R., Fagerlind, H., & Morris, A. (2013). Exploring inattention and distraction in the SafetyNet Accident Causation Database. *Accident Analysis and Prevention*, 60, 445–455. doi:10.1016/j.aap.2012.03.031
- Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, 40A, 201–237. doi:10.1080/02724988843000104
- Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, 13, 423–445.
- Tsotsos, J. K. (2001). Complexity, vision, and attention. In M. Jenkin & L. Harris (Eds.), *Vision and attention* (pp. 105–128). New York: Springer.
- Walker, S., Stafford, P., & Davis, G. (2008). Ultra-rapid categorization requires visual attention: Scenes with multiple foreground objects. *Journal of Vision*, 8(4), 21.1–12. doi:10.1167/8.4.21
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15, 160–168. doi:10.1016/j.tics.2011.02.005
- Witkin, A. P., & Tenenbaum, J. M. (1983). On the role of structure in vision. In J. Beck, B. Hope, A. Rosenfeld, & National Science Foundation (Eds.), *Human and machine vision* (pp. 481–543). New York: Academic Press.
- Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics*, 73, 1650–1671. doi:10.3758/s13414-011-0153-3
- Wulf, G., Hancock, P. A., & Rahimi, M. (1989). Motorcycle conspicuity: An evaluation and synthesis of influential factors. *Journal of Safety Research*, 20, 153–176.
- Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences*, 13, 167–174. doi:10.1016/j.tics.2009.01.008
- Yanulevskaya, V., Uijlings, J., Geusebroek, J.-M., Sebe, N., & Smeulders, A. (2013). A proto-object-based computational model for visual saliency. *Journal of Vision*, 13(13), 27:1–19. doi:10.1167/13.13.27