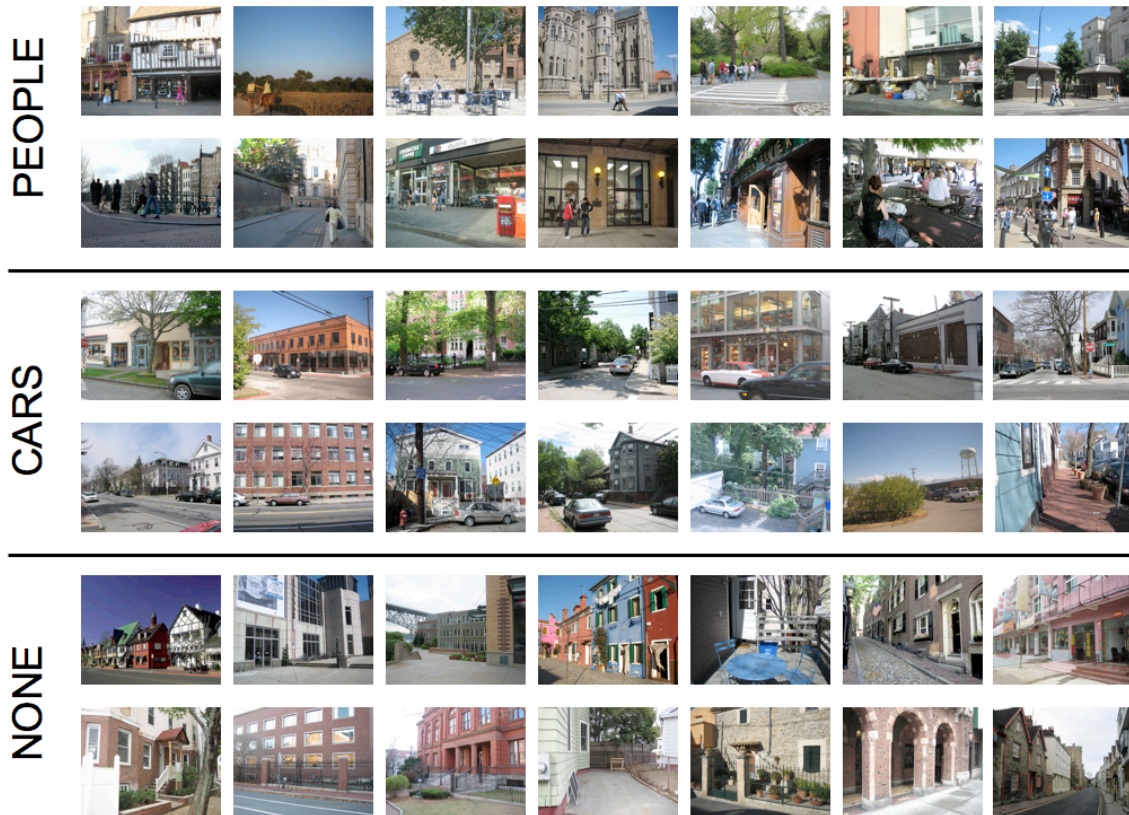
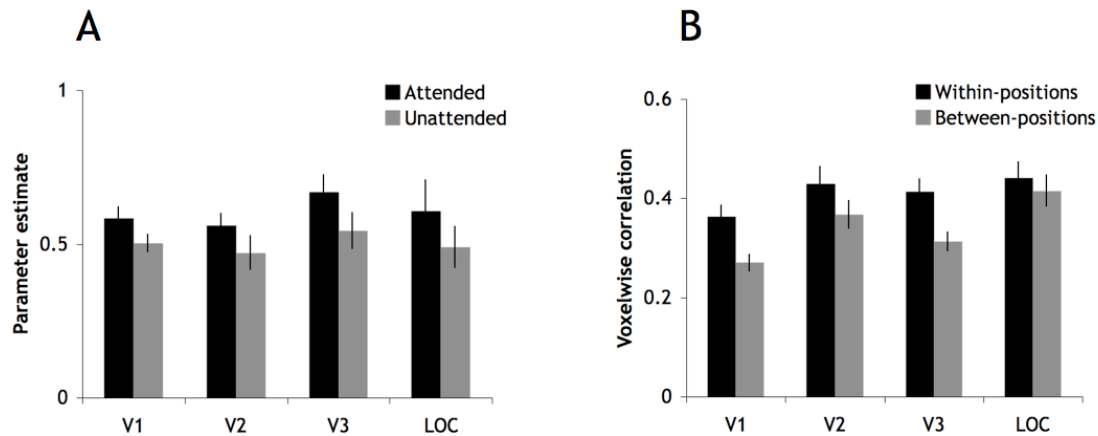


# SUPPLEMENTARY INFORMATION



**Supplementary Figure 1. Sample pictures.** Examples of the natural scene pictures used in the main experiment.



**Supplementary Figure 2. Results of attention control analysis.** **a)** Response in areas V1, V2, V3, and LOC to spatially attended (black bars) and ignored (grey bars) pictures. Responses to attended and unattended conditions were calculated by localizing (in a separate experiment) parts of retinotopic cortex that responded significantly stronger to either the horizontal or vertical stimulus locations used in the main experiment. See Supplementary Methods for further details. Attended pictures gave stronger activation in corresponding locations in retinotopic cortex than ignored pictures (V1:  $p < 0.01$ , V2:  $p < 0.05$ , V3:  $p < 0.005$ , LOC:  $p < 0.05$ ). Error bars indicate  $\pm$  s.e.m. **b)** MVPA results in areas V1, V2, V3, and LOC, showing stronger within-positions (black bars) than between-positions (grey bars) voxelwise correlations (V1, V2, V3:  $p < 0.001$ ; LOC:  $p < 0.005$ ). Responses to the two conditions in the position localizer were correlated with the two spatial attention conditions in the main experiment. See Supplementary Methods for further details. Error bars indicate  $\pm$  s.e.m.

**Supplementary Table 1. Behavioral data.** Percent correct and reaction times from the main fMRI experiment, separately for trials where the target was present in both the task-relevant and the task-irrelevant pictures (present\_present), where the target was present in the task-relevant but not in the task-irrelevant pictures (present\_absent), where the target was absent in the task-relevant and present in the task-irrelevant pictures (absent\_present), and for trials where the target was absent in both the task-relevant and the task-irrelevant pictures (absent\_absent). This analysis allowed for a test of interference effects from the task-irrelevant pictures (indicated by the difference scores in the second and fourth data column). Across the two tasks, subjects were ~18 ms slower to respond to absent\_present than absent\_absent trials ( $p < 0.05$ ), indicating a mild interference effect. No other interference effects were significant. There were no overall differences in percent correct or RT between the body and car tasks ( $p > 0.1$ , for both tests).

		% Correct	Difference	RT (msec)	Difference
Body task	Present_Present	74.1	-0.3	711.2	6.8
	Present_Absent	74.4	$p=0.85$	704.4	$p=0.31$
	Absent_Present	80.9	-2.5	788.9	14.3
	Absent_Absent	83.4	$p=0.14$	774.6	$p=0.06$
Car task	Present_Present	70.5	0.8	712.1	-9.9
	Present_Absent	69.8	$p=0.62$	722.0	$p=0.35$
	Absent_Present	80.9	-0.3	815.5	22.3
	Absent_Absent	81.2	$p=0.79$	793.2	$p=0.05$
Across tasks	Present_Present	72.3	0.2	711.7	-1.5
	Present_Absent	72.1	$p=0.84$	713.2	$p=0.81$
	Absent_Present	80.9	-1.4	802.2	18.3
	Absent_Absent	82.3	$p=0.19$	783.9	$p=0.03$

## SUPPLEMENTARY METHODS

**Visual Display.** Visual displays were generated on a Macintosh G5 computer (Apple Computer; Cupertino, CA) using MATLAB software (The MathWorks; Natick, MA) and Psychophysics Toolbox functions<sup>1</sup>. A PowerLite 7250 liquid crystal display projector (Epson; Long Beach, CA) outside the scanner room displayed the stimuli onto a translucent screen located at the end of the scanner bore. Subjects viewed the screen at a total path length of 60 cm through a mirror attached to the head coil. The screen subtended 30° of visual angle in the horizontal dimension and 26° in the vertical dimension. A trigger pulse from the scanner synchronized the onset of stimulus presentation to the beginning of the image acquisition.

**Practice session and presentation time.** Each subject participated in a practice session outside the scanner, consisting of two or four runs of the main experiment. During this practice session, the presentation time (PT) of the pictures was adjusted using a staircase procedure to arrive at a performance of approximately 80% correct on average. The mean PT of the last practice run was used as indication for the PT of the first scanning session. PT of the second scanning session was adjusted based on the performance of the first scanning session (again with the aim to arrive at ~80% correct). PT was held constant within a scanning session and was the same for all conditions. The average PT was 130 ms for both the first and second scanning session and ranged from 90 ms to 150 ms across subjects (see Supplementary Fig. 3 for behavioral performance during the scanning sessions).

**Retinotopic mapping.** Retinotopic visual field representations were determined using a rotating wedge stimulus and standard phase-encoding analysis techniques to localize visual areas V1, V2, and V3<sup>2-4</sup>. The wedge stimulus was a flickering chromatic radial checkerboard (4 Hz flicker frequency) with both luminance and chromatic contrast<sup>5</sup>. Each run was composed of 15 cycles of 32 seconds each, while subjects performed a detection task at central fixation. Two runs were performed. The mean size, in number of voxels (SD), of the ROIs was: V1=412 (81), V2=372 (48), V3=308 (65).

**Category pattern localizer.** Category-selective patterns of activation were established using pictures of cars and human bodies. Stimuli were presented centrally, and had a size of 12°x12°. They showed isolated objects presented on a white background. The experiment consisted of four conditions: human bodies, cars, outdoor scenes, and faces. The data from the scene and face conditions were not used in the present study. One scanning run consisted of 21 blocks of 14 s each. Blocks 1, 6, 11, 16, and 21 were fixation-only baseline epochs. In each of the remaining blocks, 20 different stimuli from one category were presented. Each stimulus appeared for 350 ms, followed by a blank screen for 350 ms. Twice during each block, the same picture was presented two times in succession. Subjects were required to detect these repetitions and report them with a button press (1-back task). Each participant was tested with two different versions of the experiment that counterbalanced for the order of the blocks. In both versions, assignment of category to block was counterbalanced, so that the mean serial position in the scan of

each condition was equated. Subjects performed two runs during both main scanning sessions.

**Object-selective cortex localizer.** The design of the object-selective cortex localizer was identical to that of the category pattern localizer described above, except that pictures of intact and scrambled objects were presented. Subjects performed two runs during both main scanning sessions. Object-selective cortex was defined for each subject by contrasting responses evoked by intact objects with responses evoked by scrambled objects, at  $p < 0.05$  (uncorrected). Left and right hemisphere ROIs were combined. The mean size, in number of voxels, of the ROI was 379 (SD=121).

**Position localizer.** In order to study the responses evoked by pictures presented in the horizontal and vertical positions separately, a randomly selected subset of the natural scene pictures used in the main experiment were presented in either the two horizontal or the two vertical positions. The experiment consisted of 21 blocks of 14 s each. Blocks 1, 6, 11, 16, and 21 were fixation-only baseline epochs. In each of the remaining blocks, 28 different pairs of pictures were presented, either in the horizontal or the vertical positions used in the main experiment. Stimulus pairs appeared for 250 ms, followed by a blank screen for 250 ms. Twice during each block the two pictures of the pair were identical and subjects had to detect these target trials. One run was performed.

**Attention control analysis.** We measured responses evoked by the stimuli presented in the horizontal and vertical positions using the position localizer experiment. The analysis was performed in V1, V2, V3, and LOC. For each of these ROIs, we defined the subset of voxels that responded differentially (at  $p < 0.05$ ) to the horizontal versus the vertical conditions in the position localizer. This resulted in two sets of voxels for each ROI, one primarily representing the horizontal stimulus positions, and one primarily representing the vertical stimulus positions. Left and right hemisphere ROIs were combined. The mean sizes (SD), in number of voxels, of the ROIs were: V1\_horizontal=131 (31), V1\_vertical=80 (25), V2\_horizontal=73 (20), V2\_vertical=98 (34), V3\_horizontal=96 (19), V3\_vertical=75 (34), LOC\_horizontal=44 (22), LOC\_vertical=27 (18).

For these ROIs, average parameter estimates of picture trials (all conditions combined) were calculated during runs where subjects were performing the task on the horizontal pictures and during runs where subjects were performing the task on the vertical pictures. We then averaged the values of conditions for which the spatial preference of the voxels (horizontal or vertical) corresponded to the task-relevant location (horizontal or vertical). Similarly, values of conditions for which the spatial preference of the voxels was different from the task-relevant location were combined. This resulted in two values for each ROI, one for the condition where the spatial preference of the voxels matched the task-relevant locations, and one for the condition where it did not. Note that both these values came from the same set of voxels and the same set of trials. Therefore, the only difference between these values was due to the attentional manipulation directing attention either to the horizontal or to the vertical pictures in the main experiment.

Attention effects were further investigated by using MVPA. T-values were extracted for the horizontal and vertical conditions in the position localizer for each voxel

in V1, V2, V3, and LOC. Similarly, t-values were extracted for the trials in the main experiment when subjects performed the task on the horizontal pictures and for trials when subjects performed the task on the vertical pictures. Then, the t-values of the two conditions in the position localizer (horizontal and vertical) were correlated with the t-values of the two conditions in the main experiment (horizontal and vertical). These correlations were Fisher transformed. The two within-position correlations (e.g., horizontal-horizontal) were averaged, and the two between-position correlations (e.g., horizontal-vertical) were averaged. The difference between these two values was tested using a paired-samples t-test.

## SUPPLEMENTARY DISCUSSION

**What is attended in category-based attention?** Previous studies have shown that attention to an object (e.g., a face) increases responses in visual areas that are selective for that object category<sup>6,7</sup>. In these previous studies, however, it was not investigated whether featural, spatial, and/or categorical cues were used to select the attended object. In the present study we provide evidence for a neural mechanism of attentional selection that operates at a level that is specific to object category. Attention to a particular object category biased activation evoked by real-world scenes in favor of objects belonging to the target category, both within and outside the focus of spatial attention. This object-category-based mechanism appeared to be different from spatial attention mechanisms in that it operated across the visual field. Such global attention effects are reminiscent of results from studies investigating attention to low-level features, such as color or direction of motion<sup>8,9</sup>. It is important to note, however, that it is unlikely that our results reflected attention to such low-level features. During naturalistic situations, such as those tested here, the search target (e.g., a person) is defined by a complex combination of low-level features that vary from one situation to the next (e.g., due to different body postures, clothes, viewing angle, occlusion, etc.). Furthermore, in naturalistic situations, search targets are embedded in cluttered visual scenes with a large number of competing objects that heavily overlap with the search target in terms of their low-level features. Therefore, to efficiently detect the relevant objects, one would need to attend to more complex features than, for example, color or orientation. Importantly, these complex features have to be strongly correlated with the target object category, be invariant to viewpoint and size differences, and at the same time be relatively unique to this category. Thus, these features need to be object-category-specific<sup>10</sup>. Indeed, it is conceivable that subjects' attention was directed to an abstract and holistic representation of the attended object-category rather than to diagnostic features of intermediate complexity (such as a person's arm, or a car's tire). At a neural level there is evidence for both types of representations in higher-level visual cortex. For example, single-cell recordings have revealed neurons that respond to complex combinations of features that constitute object components<sup>11</sup>. Other studies have found selective responses that directly match object categories, such as neural responses selective for realistic and schematic faces and bodies<sup>12-16</sup>. These category-selective responses may not be restricted to biologically relevant object categories such as faces and bodies. Extensive experience with particular object categories (e.g., cars) may similarly result in the formation of sparse and selective representations for these categories<sup>17-19</sup>. To further investigate the precise characteristics of what exactly subjects attend to in real-world visual search situations, such as those investigated here, will be an important avenue for future research.

**Attention and perception in natural vs. artificial scenes.** In the present study, we used natural scenes to test for neural mechanisms related to the detection of object categories. Surprisingly little is known about neural mechanisms of natural scene perception, and even less is known about attentional effects on these. Natural scenes differ from artificial scenes in several important aspects. For example, natural scenes provide a rich context that shapes and facilitates the perception of objects within them<sup>20</sup>. Another unique aspect of natural scenes is that individual objects within the scenes are highly variable in their

viewpoint, size, shape, and location, and are often occluded by other objects. Importantly, these variations follow the variations typically encountered in daily life, and are surprisingly easy for the visual system to deal with, as demonstrated by the fast detection of object categories in briefly presented natural scenes<sup>21-24</sup>. The visual system has adapted to function optimally in a natural environment, and it is therefore likely that the organization of the visual system can best be revealed when approximating its daily-life input. This is emphasized by recent studies showing that response properties of visual neurons are different for naturalistic compared to artificial stimuli<sup>25-28</sup>. Indeed, the predictions for our study were derived from behavioral studies that used natural scene images, and could not have been derived from studies using artificial stimuli alone. More generally, we expect that the use of naturalistic stimuli will be of critical importance for our understanding of the visual system and its interaction with other processes such as attention, memory, and decision-making.

**Relation between category-selective activity patterns and behavior.** Our behavioral results (see Supplementary Table 1) showed a mild interference effect related to the target presence in the task-irrelevant pictures. Specifically, for those trials where the task-relevant pictures did not contain the target category, but the task-irrelevant pictures did, subjects were slower to respond than when the task-irrelevant pictures did not contain the target category. However, these effects were only observed on subjects' reaction times (interference effect = ~18 msec), and no such interference effects were observed on their accuracy scores. Thus, subjects could successfully select the relevant pictures and respond only to these pictures. By comparison, it has previously been shown that when the target location is unknown in advance, subjects have difficulties localizing object categories in natural scenes that were nonetheless detected<sup>10</sup>. This latter result fits nicely with the parallel and spatially-unspecific object-category mechanism observed in our present study. It also suggests that under conditions where the relevant locations are known in advance (as in our study), the parallel category-based mechanism requires other, space-based, selection mechanisms to prevent responses to the task-irrelevant pictures. Our results suggest that these additional selection mechanisms (e.g., the retinotopically-specific spatial attention effects observed in early visual cortex) operate independently from the object-category mechanism observed in LOC activity patterns. At some stage, however, the spatial selection and the category-based selection mechanisms must interact in order to successfully support the subject's behavior. How this interaction is implemented at the neural level will be an important topic for future research.

Recent evidence suggests that task-relevance *per se* may also increase neural responses independently of spatial attention (note that in our study task-relevance and spatial attention were not separately investigated – relevant pictures were also always spatially attended). For example, a recent study investigated the influence of task-relevance on fMRI responses to faces<sup>29</sup>. The main finding was that when subjects performed a gender-discrimination task on a spatially unattended peripheral face in addition to a central letter-discrimination task, fMRI responses in face-selective cortex were stronger than when this additional task was not performed and the unattended peripheral face was completely task-irrelevant. These results suggest that task-relevance can influence neural responses independent of spatial attention, although follow-up studies are needed that exclude the possibility that this effect was related to performing a



dual versus a single task, a manipulation that was confounded with the manipulation of task-relevance.

Another interesting issue is the degree to which attention to object category, and the subsequent biasing of activity patterns, is *necessary* for the successful detection of objects in rapidly presented natural scenes. The influence of prior knowledge regarding the target category on behavior will likely depend on specific task parameters such as presentation time and scene complexity, and may be particularly important when viewing-time is limited and in situations where many objects compete for representation in the visual system, as is typical in natural viewing conditions. Interestingly, the influence of top-down biasing mechanisms may also be apparent in situations where viewing-time is not restricted. For example, subjects are mostly unaware of large changes to (or even complete removal of) objects in natural scenes when these changes happen during a visual interruption between the two displays<sup>30,31</sup>, especially when the changed object is of marginal interest to the observer<sup>32</sup>. These results suggest that only some objects of a complex natural scene are accurately visually represented at any given time. The category-based biasing mechanism reported here will ultimately work in concert with other top-down (e.g., spatial and feature-based attention) and bottom-up (e.g., emotional and visual salience) biasing mechanisms, which together determine which objects are successful in the competition for visual representation and gain access to our conscious visual experience.

1. Brainard, D.H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433-6 (1997).
2. Engel, S.A., Glover, G.H. & Wandell, B.A. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex* **7**, 181-92 (1997).
3. Schneider, K.A., Richter, M.C. & Kastner, S. Retinotopic organization and functional subdivisions of the human lateral geniculate nucleus: a high-resolution functional magnetic resonance imaging study. *J Neurosci* **24**, 8975-85 (2004).
4. Sereno, M.I. et al. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889-93 (1995).
5. Swisher, J.D., Halko, M.A., Merabet, L.B., McMains, S.A. & Somers, D.C. Visual topography of human intraparietal sulcus. *J Neurosci* **27**, 5326-37 (2007).
6. O'Craven, K.M., Downing, P.E. & Kanwisher, N. fMRI evidence for objects as the units of attentional selection. *Nature* **401**, 584-7 (1999).
7. Wojciulik, E., Kanwisher, N. & Driver, J. Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *J Neurophysiol* **79**, 1574-8 (1998).
8. Motter, B.C. Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J Neurosci* **14**, 2178-89 (1994).
9. Saenz, M., Buracas, G.T. & Boynton, G.M. Global effects of feature-based attention in human visual cortex. *Nat Neurosci* **5**, 631-2 (2002).
10. Evans, K.K. & Treisman, A. Perception of objects in natural scenes: is it really attention free? *J Exp Psychol Hum Percept Perform* **31**, 1476-92 (2005).
11. Tanaka, K. Inferotemporal cortex and object vision. *Annu Rev Neurosci* **19**, 109-39 (1996).

12. Downing, P.E., Chan, A.W., Peelen, M.V., Dodds, C.M. & Kanwisher, N. Domain specificity in visual cortex. *Cereb Cortex* **16**, 1453-61 (2006).
13. Downing, P.E., Jiang, Y., Shuman, M. & Kanwisher, N. A cortical area selective for visual processing of the human body. *Science* **293**, 2470-3 (2001).
14. Gross, C.G., Rocha-Miranda, C.E. & Bender, D.B. Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol* **35**, 96-111 (1972).
15. Kanwisher, N., McDermott, J. & Chun, M.M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* **17**, 4302-11 (1997).
16. Peelen, M.V. & Downing, P.E. Selectivity for the human body in the fusiform gyrus. *J Neurophysiol* **93**, 603-8 (2005).
17. Booth, M.C. & Rolls, E.T. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex* **8**, 510-23 (1998).
18. Freedman, D.J., Riesenhuber, M., Poggio, T. & Miller, E.K. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb Cortex* **16**, 1631-44 (2006).
19. Logothetis, N.K., Pauls, J. & Poggio, T. Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* **5**, 552-63 (1995).
20. Bar, M. Visual objects in context. *Nat Rev Neurosci* **5**, 617-29 (2004).
21. Fei-Fei, L., Iyer, A., Koch, C. & Perona, P. What do we perceive in a glance of a real-world scene? *J Vis* **7**, 10 (2007).
22. Li, F.F., VanRullen, R., Koch, C. & Perona, P. Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci U S A* **99**, 9596-601 (2002).
23. Potter, M.C. Meaning in visual search. *Science* **187**, 965-6 (1975).
24. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *Nature* **381**, 520-2 (1996).
25. Felsen, G. & Dan, Y. A natural approach to studying vision. *Nat Neurosci* **8**, 1643-6 (2005).
26. Kayser, C., Kording, K.P. & Konig, P. Processing of complex stimuli and natural scenes in the visual cortex. *Curr Opin Neurobiol* **14**, 468-73 (2004).
27. Rolls, E.T., Aggelopoulos, N.C. & Zheng, F. The receptive fields of inferior temporal cortex neurons in natural scenes. *J Neurosci* **23**, 339-48 (2003).
28. Simoncelli, E.P. & Olshausen, B.A. Natural image statistics and neural representation. *Annu Rev Neurosci* **24**, 1193-216 (2001).
29. Reddy, L., Moradi, F. & Koch, C. Top-down biases win against focal attention in the fusiform face area. *Neuroimage* **38**, 730-9 (2007).
30. Rensink, R.A., O'Regan, J.K. & Clark, J.J. To see or not to see: the need for attention to perceive changes in scenes. *Psychol Sci* **8**, 368-373 (1997).
31. Simons, D.J. & Rensink, R.A. Change blindness: past, present, and future. *Trends Cogn Sci* **9**, 16-20 (2005).
32. O'Regan, J.K., Rensink, R.A. & Clark, J.J. Change-blindness as a result of 'mudsplashes'. *Nature* **398**, 34 (1999).