Available online at www.sciencedirect.com

ScienceDirect

Vision Research 47 (2007) 3052–3060

Vision Research

www.elsevier.com/locate/visres

Observers are consistent when rating image conspicuity

Moran Cerf [a,*], Daniel R. Cleary [a], Robert J. Peters [b], Wolfgang Einhäuser [a,c], Christof Koch [a]

[a] *Computation and Neural Systems Program, California Institute of Technology, Caltech 216-76, Pasadena, CA 91125, USA*
[b] *Department of Computer Science, University of Southern California, Los Angeles, CA 90089, USA*
[c] *Institute of Computational Science, ETH Zurich, Switzerland*

Received 2 October 2006; received in revised form 21 March 2007

Abstract

Human perception of an image's conspicuity depends on the stimulus itself and the observer's semantic interpretation. We investigated the relative contribution of the former, sensory-driven, component. Participants viewed sequences of images from five different classes—fractals, overhead satellite imagery, grayscale and colored natural scenes, and magazine covers—and graded each numerically according to its perceived conspicuity. We found significant consistency in this rating within and between observers for all image categories. In a subsequent recognition memory test, performance was significantly above chance for all categories, with the weakest memory for satellite imagery, and reaching near ceiling for magazine covers. When repeating the experiment after one year, ratings remained consistent within each observer and category, despite the absence of explicit scene memory. Our findings suggest that the rating of image conspicuity is driven by image-immanent, sensory factors common to all observers.

Keywords: Conspicuity; Psychophysics; Human recognition; Memory

1. Introduction

Our natural visual environment is dynamic and requires rapid selection of relevant stimuli (James, 1890). Given the celerity of scene recognition, these processes can therefore be driven to a significant extent by stimulus-dependent factors rather than by top-down and higher-order cognitive factors (Biederman, 1981; Johnson, 2001; Kirchner & Thorpe, 2006; Li, VanRullen, Koch, & Perona, 2002; Oliva & Torralba, 2006; Potter, 1976; Potter & Levy, 1969; Potter, Staub, Rado, & O'Connor, 2002; Renninger & Malik, 2004; Rousselet, Fabre-Thorpe, & Thorpe, 2002; Thorpe, Fize, & Marlot, 1996). Computational studies suggest that object recognition, to some extent, can also be performed in such a sensory-driven ("bottom-up") manner (Riesenhuber & Poggio, 2000; Sun & Fisher, 2003; Thorpe, Delorme, & Van Rullen, 2001). Models of spatial attention often rely on a sensory-driven saliency metric to describe relevant subsets of a stimulus (Itti & Koch, 2001; Koch & Ullman, 1985). Such models predict certain aspects of observers' eye positions, change detection or pattern of attentional deployment (Deco & Schurmann, 2000; Navalpakkam & Itti, 2005; Parkhurst, Law, & Niebur, 2002; Peters, Iyer, Itti, & Koch, 2005; Sun & Fisher, 2003; Tsotsos, Culhane, Wai, Lai, Davis, & Nuflo, 1995; Wright, 2005). However, they typically measure saliency within a static image or video-frame, but do not address the question as to how conspicuous one image is relative to another. A prerequisite for such a measure to exist is that different individuals share common metrics of judging an image's conspicuity. When different observers judge conspicuity, do they form similar metrics and apply them consistently?

Here, we ask observers to assign a single measure of conspicuity to images within an image category (e.g., how conspicuous is one magazine cover image relative to other ones; how conspicuous is one outdoor photo relative to other outdoor photos). In addition, we test recognition memory

[*] Corresponding author. Fax: +1 626 796 8876.
E-mail address: moran@klab.caltech.edu (M. Cerf).

0042-6989/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2007.06.025

for these scenes and the effect of presentation duration. Using this experimental setting, we address three questions: First, are ratings of image conspicuity consistent across observers? Second, are the conspicuity ratings for the same image consistent across multiple presentations within the same observer? Third, is rating consistency primarily determined by recognition memory, or do low-level stimulus-driven factors play a decisive role? The extent to which these questions have positive answers enables us to construct sensory-driven (bottom-up) models of image conspicuity. Consequently, on a more abstract level, our study will provide an upper bound as to how far bottom-up models can capture seemingly subjective stimulus appeal.

2. Methods

2.1. Stimuli

Five sets of images were used in the experiments: (i) colored fractals generated at gnofract4d (http://gnofract4d.sourceforge.net) and downloaded from the "Spanky Fractal Database" (http://spanky.triumf.ca/www/welcome1.html), (ii) grayscale outdoor photos of trees, shrubs, rivers, and other scenes that did not contain objects like cars, people or animals (van Hateren & van der Schaaf, 1998), (iii) overhead, grayscale, 10 m resolution satellite images from the NGA database (http://geoengine.nga.mil/), (iv) colored natural scenes that included landscapes, flowers, trees, and oceans, and (v) a set of colored contemporary magazine covers (e.g., Time, People Magazine) including text (Fig. 1). We chose the images such that—to the best of our possibilities—there are no obvious semantic differences within a category, i.e., images of the same category share about the same gist.

2.2. Participants

Ten volunteers (5 males, 5 females; ages 20–41) participated in experiment 1. Four of these 10 participants (2 males, 2 females) also participated in experiment 2. Six additional participants (ages: 21 and 22 males, 20 and 24 females) participated in experiment 3. All observers had uncorrected normal vision and were naïve with respect to the hypotheses tested. All experimental procedures were approved by Caltech's Institutional Review Board and were performed with the written informed consent of all participants.

2.3. Presentation

Participants viewed sets of images on a computer 19" CRT-monitor in a distraction-free, darkened and isolated environment. Participants were instructed to be well-rested for the experiment. To keep a constant distance of 50 cm from the screen, we encouraged participants to use a chin-rest. Using the Groovx framework (http://ilab.usc.edu/rjpeters/groovx), we developed an extended Tcl/Tk program that displayed the images at a uniform resolution of 1200 by 900 pixels at 60 Hz.

2.4. Conspicuity rating

In the main experiments observers were instructed to rate the conspicuity of images. In the written instructions given at the start of each experimental day, we defined this as "a measure of how "salient[1]" or noticeable an image is relative to its surroundings." (Literal quote from the instruction). We further instructed observers that they "will be asked to determine how salient you find each image relative to the images previously seen in the current set of images." The instructions furthermore directed observers only to make comparisons within the current image set (and not between sets) and to distribute their responses equally between 1 and 9 (1 being not conspicuous and 9 being very conspicuous). Instructions furthermore discouraged observers from questions on the purpose of the experiment prior to experiment conclusion, but encouraged them to ask any questions needed to clarify experimental procedures.

Experiment 1—Conspicuity rating. Experiment 1 consisted of three separate sessions conducted on different days. In the first two sessions, two blocks of one category each were tested (block design), the third session consisted of a single block. Between the blocks observers took a five-minute break. With the exception of the "magazine cover" category block, which all observers performed in the third session, the order of blocks was randomized across observes (Table 1). To ensure that this relative positioning of the "magazine cover" category had no effect on the results, we tested two additional participants, who had not participated in any of the other experiments, on the "magazine covers" category alone. All data of these two observers were well within the range of the original observers. This result, together with the randomization of the first four blocks, renders it unlikely that any of the observed effects is contingent on the order of category presentations.

Each block consisted of 2 phases: a conspicuity rating phase, and a memory phase. In the first phase observers rated conspicuity following the instructions as described above. At the start of each session, observers saw a screen with a reminder to use the values 1–9 for their responses. The experiment began when observers pressed the space bar. For each trial, the image was displayed for 600 ms on the entire screen (42 × 32 degrees of visual angle), followed by a request for a response and a count of the number of images remaining in the session. This request remained on-screen until the observer responded.

At the start of each block, observers saw 35 training images drawn from the same category so they could develop a consistent internal metric for judging images (Fig. 2). The training images did not reappear in the further course of the experiment and were excluded from analysis. Subsequently, observers viewed and responded to 300 images in three repetitions of 100 unique images from one of the five sets. There was no break between the training images and the entire 300 images sequence. Observers merely saw a sequence of 335 images, of which some repeated thrice (the 100 "test" images). While the same training and test set was used in all observers, the order of images within a set was randomized individually. Observers were not told that images could appear more than once, and were in particular never explicitly instructed or encouraged to respond consistently.

In each block, the conspicuity rating phase was followed immediately by a brief memory testing phase. Observers were presented 100 images, about half of which were taken from the previously viewed images, and the remainder novel (the "color natural" and "magazine cover" categories had a 48–52 split between novel and familiar, the remainder was split 50–50). Observers responded if they had previously seen this image or not. Images in the memory task were shuffled randomly, and observers did not know what fraction of the images they had previously seen. The first memory test came as a surprise to the observers, since they had not been specifically instructed to remember the images. This allowed us to test the results of the first block as compared to other blocks in terms of recognition memory and to control for observers' effort to explicitly remember their responses.

After finishing each session, observers were interviewed about their experience—how interesting they found the different image classes and how they thought they judged conspicuity, along with questions that were aimed at verifying their attentiveness during the experiment and their ability to follow the instructions.

Experiment 2—Control for the relevance of memory. To test the effect of memory on the conspicuity rating, we repeated experiment 1 with 4 out of the 10 participants about one year after experiment 1. We selected the 4 participants solely on the basis of logistic considerations, i.e., availability on campus, but not based on their results in experiment 1. Participants

[1] Although we used "salient" and "saliency" in the instructions, we refer to it as "conspicuous" and "conspicuity" throughout this paper to avoid confusion with different notions and definitions of "saliency" in the literature.

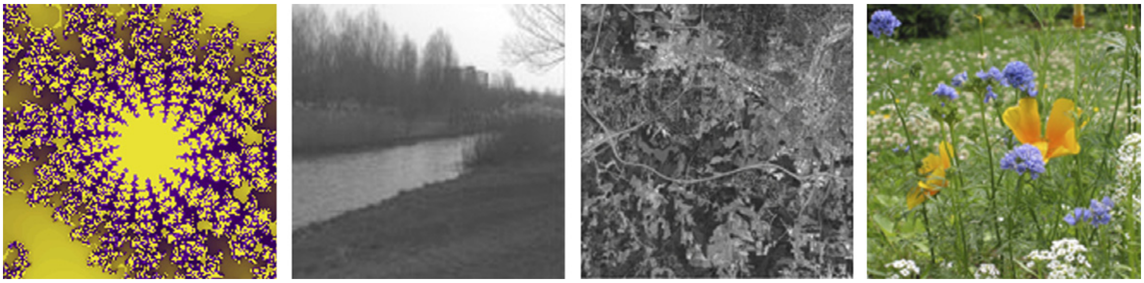

Fig. 1. Image classes used in this study. Observers judged the conspicuity of pictures from five different image classes, consisting of colored fractals, grayscale nature scenes, overhead satellite images, colored nature scenes, and contemporary magazine covers (not shown due to copyright reasons). Note that three of the five images classes (fractals, color nature and magazine covers) are in color, but are here printed in grayscale only. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

Table 1
Order of categories used in experiment 1

	Session 1		Session 2		Session 3
	Block 1	Block 2	Block 3	Block 4	Block 5
Subjects 1 and 8	Satellites	Color nature	Fractals	Grayscale nature	Magazines
Subjects 2 and 4	Grayscale nature	Fractals	Color nature	Satellites	Magazines
Subjects 3, 5, and 10	Fractals	Grayscale nature	Satellites	Color nature	Magazines
Subject 6	Fractals	Satellites	Grayscale nature	Color nature	Magazines
Subjects 7 and 9	Satellites	Color nature	Grayscale nature	Fractals	Magazines

repeated four of the five image classes using the very same images as in the year before (different order within each class) and otherwise repeated the paradigm of experiment 1. All 4 participants repeated the fractals, grayscale nature, and satellite images categories. Two of the participants also repeated the colored nature images, and the other two repeated the magazine covers. Thus, each participant judged the fractals, the grayscale nature, the satellite images, and one of the two remaining categories (magazine covers or color nature).

Experiment 3—Effect of presentation duration. In experiment 3, we tested the effect of viewing time, which was kept constant at 600 ms in the previous two experiments, on the conspicuity measure. Six additional observers viewed colored natural images in four blocks of different presentation durations (20 ms, 600 ms, 3 s, and again 20 ms). That is, each of the 100 images was seen 4 times, once in each block. Images were randomly shuffled within each block. Preceding the initial block, participant saw 35 color nature images that were not used in the remaining 100 for 600 ms to form a metric of conspicuity. Otherwise the same settings and instructions were used as in experiment 1.

3. Results

3.1. Experiment 1—Phase 1—Conspicuity rating

3.1.1. Intra-observer correlation

Fig. 3a depicts the time-course of one observer's (no 4—selected at random) judgments for fractal images. Although the responses look random at a first glance, the representation in Fig. 3b, in which the same data are sorted by image number shows high consistency across the three presentations of each image. We quantify this consistency by computing the correlation coefficient between the first and second set of responses, the second and third, and the third and first appearance of each individual image. For the example observer of Fig. 3, these values were $r = 0.83$, 0.90, and 0.83, respectively.

To investigate whether the same pattern holds for all individuals, we computed pair-wise correlations for the five image classes for all observers after transforming the conspicuity grading to z-scores (subtracting the mean of the entire set of answers for that observer and that set of responses, and dividing by the standard deviation) to make the responses of all observers comparable. For all but one observer (no. 10), all correlation coefficients exceed 0.4 (Fig. 4). These correlations are significantly different from 0 ($p < 10^{-4}$ for any pair-wise correlation). Note that significance prevails at a level of 0.005 even after a conservative Bonferroni correction for the 50 comparisons (as $10^{-4} = 0.005/50$). Observer 10 has notably less consistency than the other ones. His behavior during the experiment was idiosyncratic. It is likely that he was not really following the instructions; thus, his data were excluded from further conspicuity analysis, unless stated otherwise.

Within observer correlation depended significantly on image class (ANOVA, $p = 0.002$, $F[44] = 5.02$), with the mean (over 9 observers) equal to 0.64 ± 0.05 (mean ± standard error of the mean) for the fractals, 0.70 ± 0.02 for the grayscale natural scenes, 0.67 ± 0.05 for the overhead satellite imagery, 0.76 ± 0.03 for the colored natural scenes, and 0.85 ± 0.01 for the magazines covers. Thus, observers were quite consistent in their conspicuity judgments for the same image, with their consistency highest for the magazine covers and the lowest for the fractals. While it seems surprising that magazine covers, which are of high semantic load, are most consistent, we want to stress that the present experiment was not designed for inter-category comparisons. The fact that we find an effect for any of the categories,

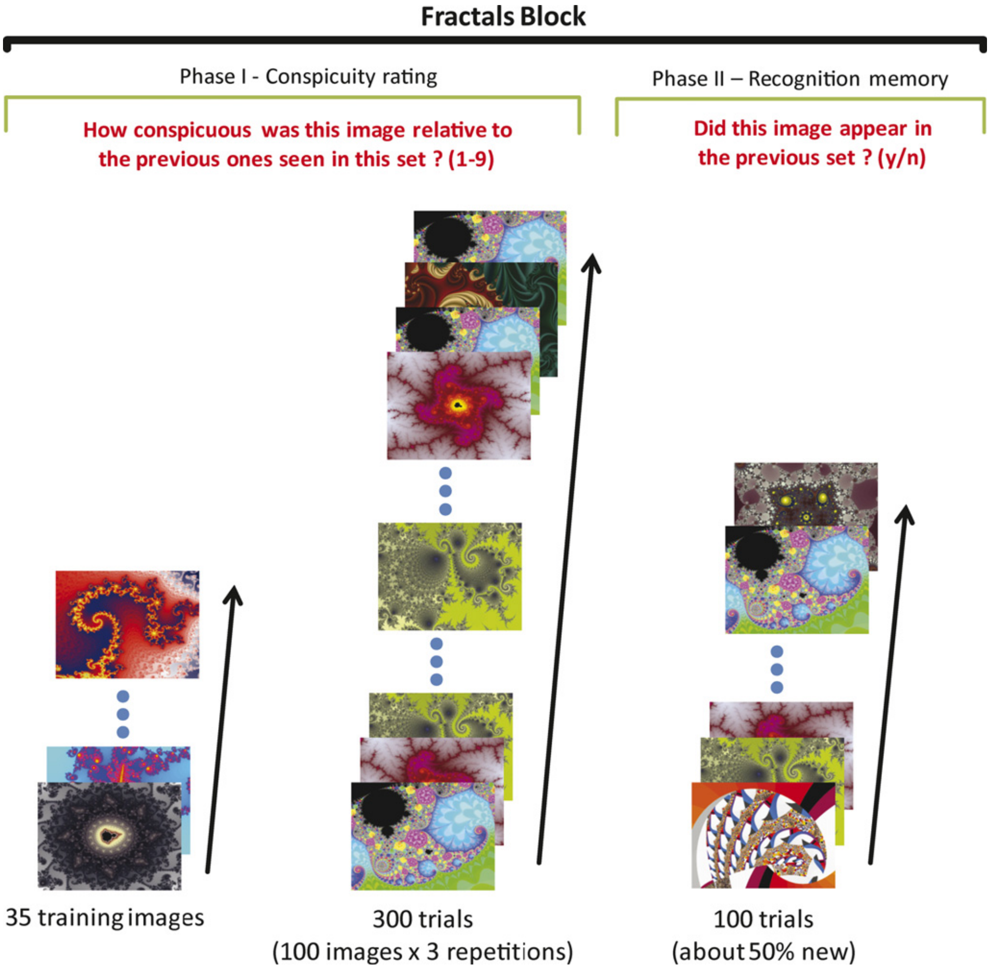
Fractals Block
Phase I - Conspicuity rating
How conspicuous was this image relative to the previous ones seen in this set ? (1-9)
Phase II – Recognition memory
Did this image appear in the previous set ? (y/n)
35 training images
300 trials
(100 images x 3 repetitions)
100 trials
(about 50% new)

Fig. 2. Experimental design. A single experimental block as used in experiments 1 and 2. Only one category was tested per block (here: fractals). The first 35 images did not reappear ("training images", *left*) during the remainder of the block. The next 300 trials (*middle*) consisted of 100 images that were presented thrice in random order (with the restriction that at least one different image occurred between two presentations of the same image). For all 335 trials, observers rated subjective conspicuity. Immediately following this "rating" phase, there was a "memory" test phase (*right*), in which 100 images were presented. About half of the images were repetitions of the rating phase, half were novel images of the same class. Observers responded as to whether or not they had seen the image before. In both phases, images were presented for 600 ms. In experiment 1, each observer performed 5 of these blocks distributed over three sessions (Table 1), in experiment 2, four blocks were tested, split across two sessions.

irrespective of semantic load, is nonetheless striking and is to be tested in further research.

All of the 9 observers have a higher correlation between their 2nd and 3rd responses as compared to the correlation between their 1st and 2nd responses (Fig. 5), with all but 6 points (out of 45) falling above the diagonal. A sign-test[2] reveals that this bias is significant ($p = 5 \times 10^{-7}$).

As an additional measure beyond linear correlation, we counted the number of images to which a given observer responded identically thrice (e.g., rating image no. 29 a 7–7–7). The example observer (no. 4) responded to 22 fractals with the same rating on all three trials. The average number across the 9 analyzed observers is 12.78 ± 7.17 (mean ± SD) for fractals, 15.65 ± 8.30 for grayscale nature, 18.44 ± 6.54 for satellite imagery, 22.33 ± 7.50 for color nature and 32.11 ± 11.99 for magazine covers. This is far above the chance level of random occurrence of $100/9^2 = 1.23$ for each image class, and further supports the high intra-observer consistency reflected in the correlation analysis.

3.1.2. Inter-observer correlation

Analyzing the responses for the 100 images across 9 observers revealed significant correlations of the conspicuity judgments across all observers and categories, though the correlations were lower than the intra-observer correlations (Fig. 6). Correlations range from $r = 0.09$ for fractals to $r = 0.51$ for magazine covers. These findings demonstrate that observers are consistent across categories. While we do not deny a category-dependency of this effect (ANOVA, $p < 10^{-19}$, $F[179] = 23.55$), it is intriguing that

[2] The sign-test tests against the null hypothesis that both correlation values are drawn from the same, arbitrary (but continuous) distribution and just compares whether r_{12} is larger or smaller than r_{13} irrespective of their absolute values. This is the most conservative estimate; additional assumptions on the distribution would yield lower p-values.

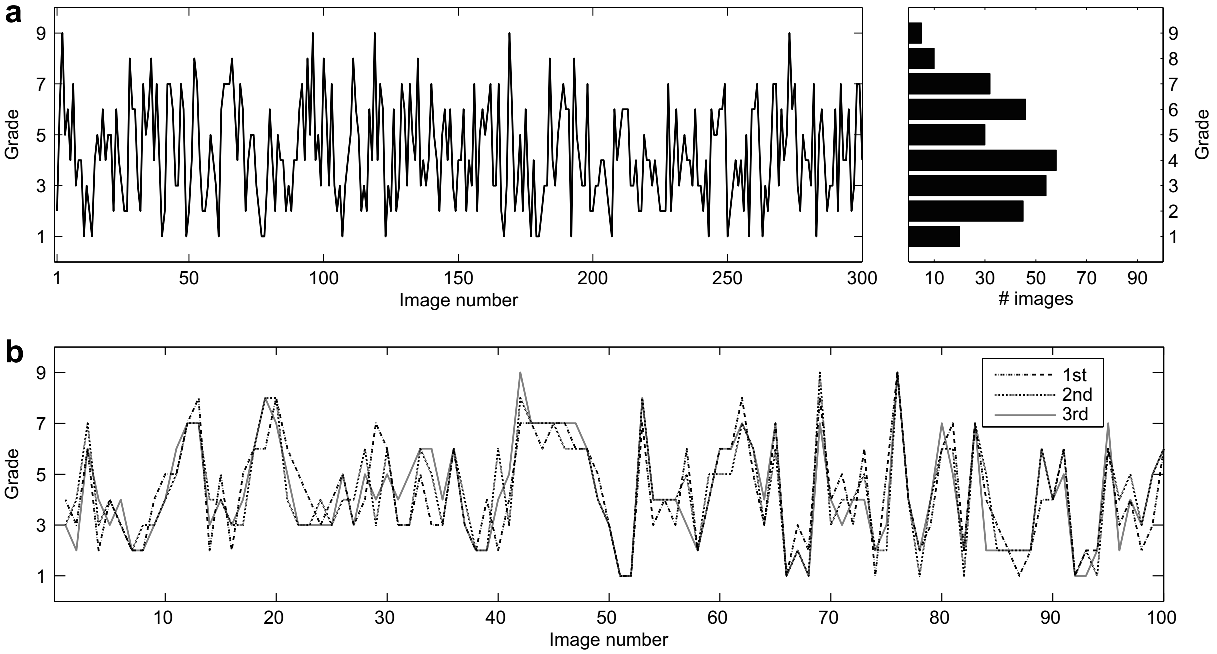

Fig. 3. Observer 4's responses to fractal images. (a) The conspicuity judgment of observer 4 in the order in which the observer viewed the 300 fractals images. On the right is a histogram of the distribution of the 300 responses. (b) When re-arranged by image identity, the consistency of the observer in assigning the same (or similar) conspicuity values to the same image becomes apparent. We found similar trends in 9 out of 10 observers and across all image classes.

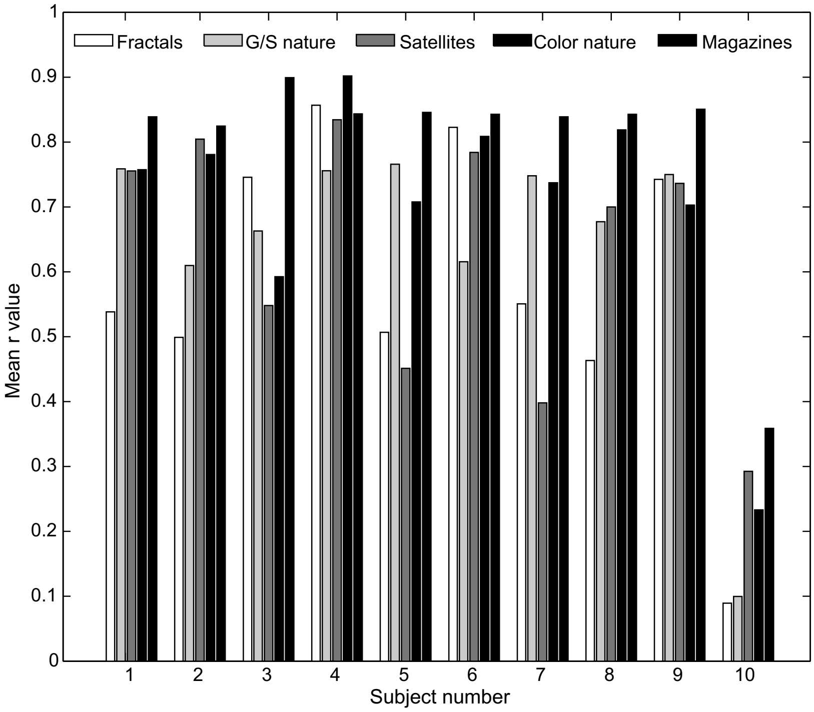

Fig. 4. Intra-observer correlation. Mean pair-wise correlation between 1st and 2nd answers, 2nd and 3rd, and 1st and 3rd answers. Observers, with one exception, are consistent when judging the conspicuity of the same image three times. The *r* correlation values are given as a function of individual and image class.

the correlation is significant for a wide variety of stimuli, ranging from fractals, which have no obvious semantic content, to magazine covers, which seem intuitively largely dominated by content.

3.3. Experiment 1—Phase 2—Memory

How much of an observer's internal consistency can be explained by a sensory-driven conspicuity module that

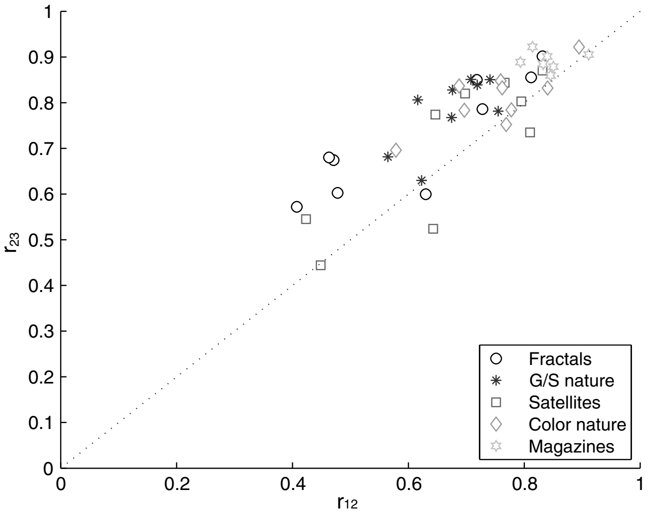

Fig. 5. Observers become more consistent after repeated exposure. Scatter diagram of the *r* correlation values in the conspicuity judgments between the 1st and 2nd showing of an image (*x*-axis) and its 2nd and the 3rd repetition (*y*-axis) for each image class for the 9 observers with high consistency.

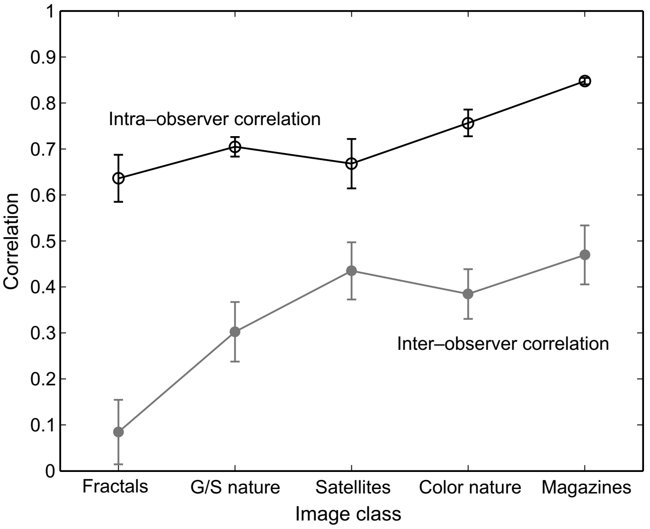

Fig. 6. Inter-observer correlation. Mean pair-wise *r*-values and standard error of the mean among the 9 consistent observers when judging the conspicuity of identical images as a function of image class. Each *r*-value is significantly different than 0 at $p < 0.05$. The upper line corresponds to the mean and standard error of the within-observer correlation (Fig. 3). Across-observer correlation was calculated for each participant and then averaged across participants.

responds in a stereotypical manner to repetitions of the same image and how much has to be attributed to memory for individual images? After each block of 300 images, we asked observers to view an additional set of 100 images, about half of which had been presented during the preceding conspicuity grading phase. All observers performed the memory test above chance (50%) for fractals (69.9 ± 11.5%; mean ± SD over observers; Fig. 7), grayscale images (90.2 ± 6.7%), color images (85.3 ± 7.5%), and magazine covers (96.3 ± 5.4%). Across observers these numbers are significantly different from chance ($p = 3.9 \times 10^{-4}$, $p = 1.5 \times 10^{-8}$, $p = 1.2 \times 10^{-7}$, $p = 5.7 \times 10^{-10}$, *t*-test, respectively). In contrast, performance was close to—but still significantly above—chance for satellite images (55.0 ± 6.5%, $p = 0.04$). This indicates that observers have good recognition memory for briefly presented images of a variety of classes. If memory were the primary factor for consistency we would expect a comparably low intra-observer consistency for the image class that is remembered worst. However, the consistency for satellite images is well within the range of the other classes, which are remembered better. This argues against the notion that rating consistency is primarily memory-driven.

Does the rating an observer gives for an image's conspicuity relate to recognition memory? In particular are highly conspicuous images remembered better? For each observer we selected those images that they rated consistently with the highest score (e.g., responses of "8–8–8", if 8 were the highest value the observer gave for that category). Across the 9 observers, for whom we had obtained consistent conspicuity, we obtain thus 58 "high conspicuity" images (out of 4500 = 9 observers × 100 images × 5 categories), of which 26 were probed in the memory tests. Of these 26, all but one (96.2%) were remembered correctly. Similarly we selected the images with the consistent lowest scores (e.g., 1–1–1). This yielded 163 "low conspicuity" images of which 70 were probed in the memory test. Of those 70, 68 (97.1%) were correctly recalled. In comparison, of all 2286 images ($9 \times [3 \times 50+2 \times 52]$) that were targets in the memory test (i.e., were rated in the preceding phase), only 1790 (78.3%) were correctly identified as repetition. A χ^2-test rejects the null-hypothesis that the factors "conspicuity" (low, medium, high) and "memory" (hit/miss) are independent ($p = 1.3 \times 10^{-4}$; $\chi^2 = 17.95$; 2 d.o.f.). This suggests that repetitions of images with consistently extreme (high or low) conspicuity are more likely to be reported than repetitions of images with intermediate ratings. We have no means of assessing the conspicuity of images that have not been presented before. We therefore cannot know whether images of extreme conspicuity that were previously not shown would have high false alarm rates for recall. Hence, it remains open whether increased recall for extreme conspicuity is due to better memory or lower response threshold during retrieval. In either case, images of extreme conspicuity are qualitatively different with respect to memory, be it in memorization or retrieval.

3.4. Experiment 2—Long-term longitudinal consistency

A year after experiment 1, 4 of the 10 original observers (nos. 2, 4, 7, and 9) took the experiment for the second time (experiment 2). All four observers reported in debriefing after experiment 2 that they had not recalled any of their conspicuity ratings from the previous year. The mean change for the intra-observer correlation over the image categories was −0.033 for the 4 observers. For logistic rea-

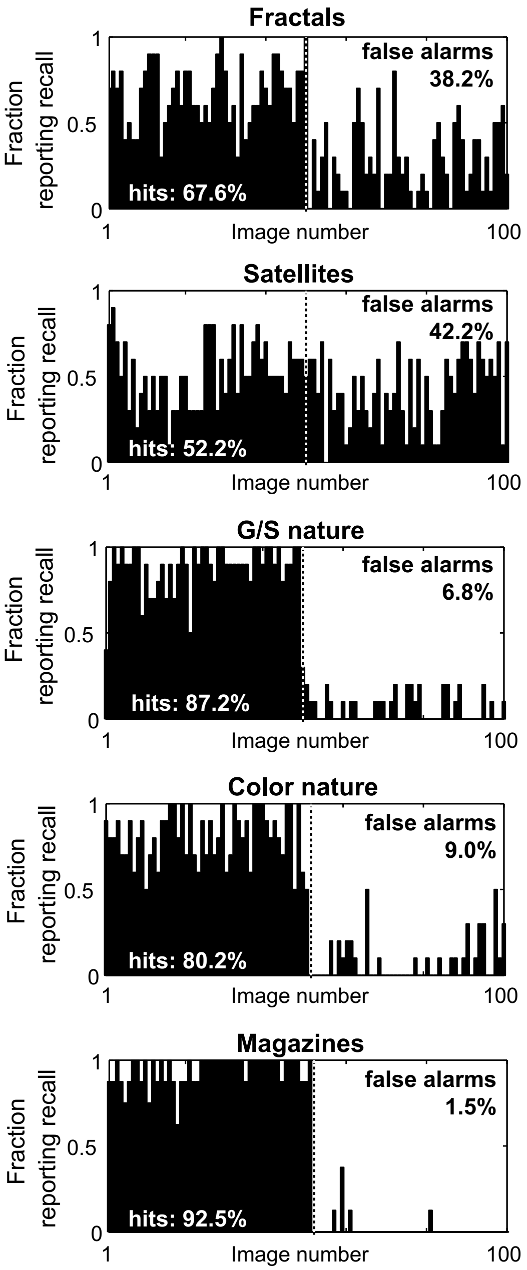

Fig. 7. Recognition memory. Normalized performance on the recognition test for all 10 participants for the 50 previously seen images versus the 50 (52 and 48, respectively, for the color and magazine covers) novel images drawn from the same class. The data are sorted so that recognition performance with perfect memory would be 1 for images 1–50 and 0 for images 51–100 (1–52, and 53–100, respectively, for the color nature and magazine covers). Each figure includes the hits and false alarms percentages. Observers have almost perfect memory for magazine covers but perform much worse (yet still above chance) for fractal. The satellite images performance is at chance. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

sons, only 3 categories (fractals, grayscale nature, and satellites) were taken by all observers, whereas a remaining category was taken by only two observers each (Table 2). Out of 16 (4 observers × 4 categories) correlation coefficients, 14 were larger for the correlation between 2nd and 3rd presentation than between the 2nd and 1st presentation. This trend is significant ($p = 0.004$, sign-test) and consistent with the data of experiment 1. Comparing the mean r-values from experiment 1 for each image class with those of experiment 2 for the 3 categories taken by all observers yields only small changes (change of mean r: fractals, +0.05; grayscale nature, −0.07; satellite images, −0.04). Recognition memory also remains virtually unchanged: the fractals miss percentage in year two is 32% (comparing to 32.40% in year one for 10 observers), grayscale nature is 9% (12.8% in year one), and overhead satellite imagery have 29% misses (47.8% in year one). To further quantify any behavioral change, we compared the correlation (z-score) between the 3rd viewing of the images in year one with the 1st viewing in year two. The mean intra-observer correlation between the years was 0.51 (with $p < 0.01$ for all individual correlations). Finally, the mean inter-observer correlations for the three images classes taken by all four observers remain unchanged. While the mean pair-wise inter-observer correlations for those 4 observers alone in the previous year were 0.15 ± 0.13 (fractals, mean ± standard error), 0.39 ± 0.06 (grayscale nature), and 0.50 ± 0.09 (satellite images), the changes in the following years were: fractals, +0.065; grayscale nature, −0.030; and satellite images, +0.107. Interestingly, despite the lack of explicit memory for satellite images, inter-observer consistency increases, pointing to factors common to all observers, i.e., image-immanent, sensory-driven variables, which determine conspicuity ratings. Although observers do not explicitly recall their previous conspicuity ratings, their overall judgments remain highly consistent after a period of one year which hints for an internal metric that observers keep when making judgments, with the same caveats (changes with time between 1st–2nd/3rd observations), and the same variance with respect to other people.

Table 2
Order of categories used in experiment 2

	Session 1		Session 2	
	Block 1	Block 2	Block 3	Block 4
Subjects 2 and 7	Fractals	Grayscale nature	Satellites	Magazines
Subjects 4 and 9	Fractals	Grayscale nature	Satellites	Color nature

3.5. Experiment 3—Effect of presentation duration

In experiments 1 and 2, each image was presented for 600 ms. Does this choice of presentation duration affect inter- and/or intra-observer consistency? We presented a set of 100 colored natural scenes to 6 additional observers, none of whom had participated in experiment 1 or 2. Images were shown at 3 different presentation durations in four blocks, for 20 ms, 600 ms, 3 s, and again 20 ms. Each block consisted of the same 100 images in random order, (i.e., there were four blocks with 4 presentation durations in total). All pair-wise correlation coefficients were positive for all presentation durations. The mean correlation coefficient across all 15 pairs of observers was significantly different from 0 for all presentation durations ($p < 10^{-4}$, *t*-test for any duration). It is lowest for 20 ms presentations (first block: 0.17 ± 0.10—mean ± SD, fourth block: 0.23 ± 0.16), intermediate for 600 ms presentation (0.26 ± 0.14) and highest for 3000 ms (0.33 ± 0.15). An ANOVA for factor presentation time (for the first three blocks) shows a significant effect at $p = 0.0068$. That is, within-observer consistency increases with prolonged viewing time.

4. Discussion

The present study demonstrates that judgments of an image's overall conspicuity are consistent across observers and across multiple presentations within the same observer. Despite some dependence on image class, a significant correlation between conspicuity ratings is observed for a wide variety of classes, ranging in semantic content from fractals to magazine covers. Consistency within an image class does not depend on memory for images in this class.

In principle, the observed consistency can arise from a variety of sources: one possibility that explains both within- and between-observer consistency is that conspicuity judgments are primarily sensory-driven, image-immanent. This view is supported by the relatively high consistency that already exists for very short presentation durations of 20 ms. An alternative explanation for within-observer consistency would be that observers rely on memory rather than on a stereotyped bottom-up evaluation (Standing, 1973). While observers have good recognition memory for most of our images, such an account would not explain between-observer consistency. Furthermore, in the long-term test after a year, we found that observers did not explicitly remember the images and required time to once again set their metric (revealed by increase of intra-observer correlation over each session). Nevertheless, their ratings were consistent with the ratings made in the previous year. Finally, images with recognizable content and objects are more easily remembered than images with less readily identifiable objects (Underwood, Foulsham, van Loon, & Underwood, 2005). In our case, recognition memory for overhead satellite images was worst of all categories, while memory for fractals was better. Nevertheless, satellite images were rated more consistently than fractals, further arguing against a purely memory-based account.

Although a large body of psychophysical, electrophysiological and computational studies focus on the rapid recognition of the main content, or "gist" of a scene (Biederman, 1981; Evans & Treisman, 2005; Li et al., 2002; Oliva & Torralba, 2006; Potter & Levy, 1969; Renninger & Malik, 2004; Rousselet et al., 2002; Schyns & Oliva, 1997; Thorpe et al., 1996), the rapid evaluation of the conspicuity of scenes has received little investigation. We instructed observers to rate the "saliency" of a scene. The rationale behind the choice of such a vague term to characterize conspicuity was to evoke very different associations between individuals as compared to well-defined terminologies. In light of this vagueness, the high consistencies we found are more convincing than if we had used a much more detailed and explicit instruction or pre-labeled example images. In the modeling literature, "saliency" is typically used in the context of objects within an image rather than across images, i.e., as a term of spatial attention rather than global preference (Itti & Koch, 2001; Koch & Ullman, 1985). Wright (Wright, 2005) demonstrated that such within-image saliency is closely linked to the probability of change detection and a subjective saliency rating. Since there is evidence that visual processing of individual objects recruits different early mechanisms from gist recognition in entire scenes (Oliva & Torralba, 2006), our findings on image-wide consistency are consistent, but distinct from Wright's results. Nevertheless, inter-observer consistency is a prerequisite to construct sensory-driven (bottom-up) models that predict non-idiosyncratic aspects of human behavior. As neurally inspired bottom-up models have been successfully applied to spatial attention (Itti & Koch, 2001; Peters et al., 2005; Tsotsos et al., 1995), object recognition (Riesenhuber & Poggio, 2000) and gist recognition (Oliva & Torralba, 2006; Renninger & Malik, 2004), our results spur the possibility that such sensory-driven models could also partially predict human preference for certain scenes. Such a putative model would have a variety of applications, ranging from art to advertisements to human-factors usability design.

Acknowledgments

We wish to thank Laurent Itti and the anonymous reviewers for their valuable comments on previous versions of the manuscript. This research was supported by the National Geospatial-Intelligence Agency (NGA), the National Science Foundation, and the Swiss National Science Foundation (WE, PA00A-111447).

References

Biederman, I. (1981). On the semantics of a glance at a scene. In M. K. J. R. Pomerantz (Ed.), *Perceptual organization* (pp. 213–263). Hillsdale, New Jersey: Lawrence Erlbaum.

Deco, G., & Schurmann, B. (2000). A hierarchical neural system with attentional top-down enhancement of the spatial resolution for object recognition. *Vision Research, 40*(20), 2845–2859.

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology-Human Perception and Performance, 31*(6), 1476–1492.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience, 2*(3), 194–203.

James, W. (1890). *The principles of psychology*. New York: Henry Holt.

Johnson, S. P. (2001). Fleeting memories: Cognition of brief visual stimuli. *Contemporary Psychology-Apa Review of Books, 46*(6), 561–564.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46*(11), 1762–1776.

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

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 of the United States of America, 99*(14), 9596–9601.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*(2), 205–231.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155*, 23–36.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107–123.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research, 45*(18), 2397–2416.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology-Human Learning and Memory, 2*(5), 509–522.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*(1), 10-&.

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology-Human Perception and Performance, 28*(5), 1163–1175.

Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research, 44*(19), 2301–2311.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3*(Suppl.), 1199–1204.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5*(7), 629–630.

Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception, 26*(8), 1027–1038.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207–222.

Sun, Y. R., & Fisher, R. (2003). Object-based visual attention for computer vision. *Artificial Intelligence, 146*(1), 77–123.

Thorpe, S., Delorme, A., & Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks, 14*(6–7), 715–725.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y. H., Davis, N., & Nuflo, F. (1995). Modeling visual-attention via selective tuning. *Artificial Intelligence, 78*(1-2), 507–545.

Underwood, G., Foulsham, T., van Loon, E., & Underwood, J. (2005). Visual attention, visual saliency, and eye movements during the inspection of natural scenes. In *Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach*, *Pt. 2*, *Proceedings* (Vol. 3562, pp. 459–468).

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London Series B-Biological Sciences, 265*(1394), 359–366.

Wright, M. J. (2005). Saliency predicts change detection in pictures of natural scenes. *Spatial Vision, 18*(4), 413–430.