Using an oculomotor signature as an indicator of aesthetic preference

Tim Holmes*

Royal Holloway, University of London, Department of Psychology, Egham Hill, Egham, Surrey, TW20 0EX, UK;
email: t.holmes@rhul.ac.uk

Johannes M. Zanker

Royal Holloway, University of London, Department of Psychology, Egham Hill, Egham, Surrey, TW20 0EX, UK. email: j.zanker@rhul.ac.uk

Abstract. Eye movements are strongly influenced by the task given to an observer. The immediacy of such eye movements, which are difficult to control consciously, offers the potential to explore highly variable subjective evaluations, such as aesthetic preference, with reliable objective measures. We presented a variety of images in sets of 2, 4, or 8 items for different durations and analyzed oculomotor statistics such as cumulative fixation duration, refixations, and the sequence of fixations while participants searched for their preferred image, after which participants indicated their preference using a button press. The total amount of time spent looking at any image correlates with selection preference and does so increasingly well with longer presentation duration. For short presentations, the first and last fixations correlate better with image preference. All response measures become increasingly variable as the number and complexity of presented images are increased. A weighted combination of these measures can significantly improve the correlation with preference, suggesting a "signature" which could be used as a reliable indicator for task-free subjective evaluation of stimuli in visual psychophysics. Its role as an improved fitness function in visually driven evolutionary algorithms is discussed.

Keywords: aesthetics, visual perception, attention, eye movements, eye tracking, top-down, bottom-up.

1   Introduction

The empirical study of the aesthetic experience is not easy. Fechner (1865, as cited in Berlyne, 1971) was well aware of this when he first established the field of experimental aesthetics and observed that

> "aesthetic phenomena are among the most complex phenomena within the domain of psychology, and this is often used as an excuse by psychologists for being able to say so little about them".

This difficulty stems, in part, from a wide range of definitions of what constitutes, or contributes to, the aesthetic experience (Arnheim, 1974; Berlyne, 1971; Hassenzahl, 2008) as well as the range and complexity of the stimuli, including paintings and sculpture, architecture, musical compositions and performances, film and dance, which can be evaluated aesthetically. These challenges frequently result in a reductionist approach, which allows more precise questions about the relationship between some feature of the stimulus and the consciously articulated preferences to be explored in a highly controlled manner. However, as observed by Arnheim (1974), the aesthetic experience is rather fragile and typically relates to the gestalt, or whole, rather than the sum of any isolated parts meaning that a reductionist approach will never be able to explain the effects of the complex interaction of many individual features on the aesthetic evaluation of the stimulus. Previously, we have shown (Holmes & Zanker, 2008) that evolutionary algorithms (Bentley, 1999; Goldberg, 1989; Holland, 1975), which provide a robust method for searching complex feature spaces for highly fit interactions, can be used to answer questions relating to aesthetic preference, using the time spent fixating each image in a simultaneously presented array as a measure of fitness. In this paper, we identify a more powerful oculomotor

*Corresponding author.

signature which correlates with selections based on visual aesthetic preference using a wide range of photographic stimuli, with a view to integrating this signature as a fitness measure in the gaze-driven evolutionary algorithm (Holmes & Zanker, 2008) to facilitate the aesthetic evaluation of complex visual stimuli without the need for conscious articulation of that preference.

The association between the task being performed and the eye movements made while performing the task is well established. In picture viewing (Buswell, 1935; Yarbus, 1967), music and word reading (Land & Furneaux, 1997; Rayner, 1998), visual search (Connor, Egeth, & Yantis, 2004), complex real-world scene perception (Henderson, 2004), and motor tasks (Land, 2007), consistent patterns of eye movements have been detected across participants which point to the crucial role that the task plays in the direction of overt attention. Free viewing tasks, where the participant is not given any task to perform but is merely encouraged to explore the visual scene, are typically used to highlight the influence of low-level, or bottom-up, influences on eye movements stemming from the visual scene itself. However, such effects are not stable over time (Tatler, Baddeley, & Gilchrist, 2005), suggesting the presence of some internally motivated, top-down tasks generated by the participant. It has been suggested that humans seek to create a stimulating and pleasing visual environment for themselves (Berlyne, 1971) and that such behavior might be biologically conditioned as a result of the positive sensations that such stimuli provide (Biederman & Vessel, 2006). If that is indeed the case, then the pattern of eye movements made while performing an aesthetic evaluation task might allow preferences to be predicted under free viewing conditions, allowing aesthetic preferences to be explored experimentally without the many problems associated with conscious reporting of subjective preferences (Nisbett & Wilson, 1977).

Artworks have long been used to elicit eye movements for scientific study, whether for the purposes of exploring task demand effects (Yarbus, 1967) or for providing an insight into the relationship between artistic images and the viewer's aesthetic response to it (Buswell, 1935; Locher, Krupinski, Mello-Thoms, & Nodine, 2007; Stratton, 1906; Tatler, Wade, & Kaulard, 2007; Wooding, Mugglestone, Purdy, & Gale, 2002). Most attempts to relate eye movements to the aesthetic qualities of the stimulus have concentrated on the looking behavior of participants while viewing paintings one at a time (Buswell, 1935; Henderson & Hollingworth, 1998; Rayner & Pollatsek, 1992; Wooding, 2002) and have shown a degree of consistency between viewers with respect to the number and duration of fixations in regions of the artwork which are highly detailed and thus potentially more informative. Attempts to correlate such fixational measures with respect to the outcome of a relative evaluation of several stimuli presented simultaneously are less common, perhaps due to the large individual differences in aesthetic preference (Arnheim, 1974; Berlyne, 1971), but this does not preclude a potential consistency in the oculomotor behavior during the collection of visual information about the stimuli prior to the decision being made, and there is evidence to support a tendency for gaze to linger on objects or artistic images we prefer (Pieters & Warlop, 1999; Plumhoff & Schirillo, 2009; Santella & DeCarlo, 2004).

Preferential selection based on eye movements is not new and is central to the preferential looking paradigm (Dobson, 1983; Teller, 1979) and further supported by the gaze cascade model of preference decision making (Glaholt & Reingold, 2009; Shimojo, Simion, Shimojo, & Scheier, 2003). Fundamental to both of these is the relationship between accumulated fixation time and task relevance first identified by Yarbus (1967). Supporting empirical studies are traditionally performed with just two concurrently presented stimuli, with preference being attributed to a stimulus if it attracts the gaze above chance, that is, more than 50% of the presentation time. In other words, the preferential looking paradigm utilizes a two-alternative forced-choice (2AFC) method of selection which reduces the noise in the decision making by controlling the cognitive load on the participant and the differences between the images in the binary choice (Houston-Price & Nakai, 2004). However, this inevitably constrains the variability between stimuli to one or two features (i.e., stimulus dimensions) and means that an exploration of interactions among multiple stimulus features results in a large number of trials, introducing the risk of additional noise arising from participant fatigue.

Most models of overt visual search involving multifeatured stimuli, suggest two phases; the first being associated with scanning the visual field for relevant locations and the second being associated with the more detailed identification, comparison, and evaluation of these image regions (Locher et al., 2007; Van der Lans, Pieters, & Wedel, 2008; Velichkovsky, Joos, Helmert, & Pannasch, 2005). This first phase has been shown to be highly susceptible to the distribution of low-level image features over the visual field (saliency; Itti & Koch, 2000) and is characterized by frequent, short fixations (Antes, 1974; Glaholt & Reingold, 2009). The second phase introduces top-down effects, for exam-

ple, driven by the task being performed, the emotional content of the image, or personal preferences (Glaholt & Reingold, 2009; Land, 2007; Locher et al., 2007) and results in fewer but longer fixations (Antes, 1974; Glaholt & Reingold, 2009). Over a prolonged viewing period, the two operation modes are likely to alternate repeatedly and interact, because participants look for what they like but their preferences can also be influenced by what they see (Glaholt & Reingold, 2009; Shimojo et al., 2003). This would suggest that the time course of the eye movements, over and above the cumulative fixation time used in the preferential looking paradigm, is likely to provide substantial information, which relates to the participant's decision making. A variety of oculomotor measures, other than fixation time, are available from an eye tracker, which have, to varying degrees, been correlated with viewing strategies during visual search, single image viewing, and decision making. These include number, spatial, and temporal distribution of fixations (Carbon et al., 2006; Duncan & Humphreys, 1989; Land, 2007; Pannasch, Helmert, Roth, Herbold, & Walter, 2008; Shimojo et al., 2008; Wooding, 2002) and saccade size (amplitude) which has been show to have an inverse correlation with fixation duration during the later (top-down initiated) phase while viewing artworks (Pannasch et al., 2008). This paper will focus on the spatiotemporal distribution of gaze positions and their associated durations, which should already contain implicit information that also could be derived from analysis of saccade amplitudes.

To test whether the oculomotor measures generalize across aesthetic preference tasks, it is important that a range of images is evaluated, and four different categories were used in the current experiment. Common small objects and buildings are known to have distinct representations in the human brain with differing typical locations in the visual field—small objects are typically located in the fovea, while buildings are more frequently seen in the peripheral visual field (Aguirre, Zarahn, & D'Esposito, 1998; Levy, Hasson, Avidan, Hendler, & Malach, 2001). These two stimulus types are used along with familiar packages of commercial products, which represent an everyday example of preference-based visual search in visually cluttered environments such as supermarket shelves (Van der Lans et al., 2008). We have shown previously that simple images such as monochromatic rectangles with varying height to width ratio and colored shapes can be evolved using an evolutionary algorithm in which the fitness of each image is based on the cumulative fixation duration (Holmes & Zanker, 2008). To link to such work on basic figures, a fourth category of simple colored geometric shapes is included in the present study to further validate the generality of the experimental findings. Within each main category, images representing four different types of stimuli (subcategories) are used to control for any effects unique to a semantically associated type of image or to a particular class of features. In addition to manipulating the image type, the display size and presentation duration are varied in our experiments, with the intention of determining an optimal signal to noise ratio for sampling aesthetic preference with a view to providing sufficient competition between images without fatiguing the participant. This is particularly important if the signature is to be incorporated in an evolutionary algorithm where competition for a limited resource, such as the attention of the participant, is vital to establish relative preference, or fitness scores, for the images being presented in an iterative process.

2 General Methods

2.1 Participants

Seven participants with normal, or corrected-to-normal, vision were recruited from the Psychology department at Royal Holloway, University of London. The project was approved by the institutional ethics committee.

2.2 Stimuli

Samples of two, four, and eight images from one of four categories, each comprising four subcategories (see Figure 1), were displayed using a Cambridge Research Systems ViSaGe (Visual Stimulus Generator). Luminance was not normalized across the images, as we were interested in identifying oculomotor signature that would work despite uncontrolled luminance. Background luminance (white) was 86.3 cd/m². Stimuli were displayed on a 48-cm diameter cathode ray tube (CRT) monitor at a distance of 57 cm from the participant. The physical size of an image on screen ranged from 20 mm $\times$ 30 mm to 50 mm $\times$ 40 mm. Random sampling without replacement from a set of eight images for each subcategory ensured that no image was displayed more than once in each screen presentation, but the same image could be displayed in more than one trial (in case of eight displayed images, the full set was presented in each trial). This was intended to introduce contextual effects because the aesthetic evaluation of each image is relative to the other individuals presented (Shen, Elahipanah, & Reingold,

Figure 1. Complete image set for Experiments 1 and 2. All eight images for each category and subcategory are shown in one of their 8AFC presentations. Reading top-down and left to right: objects (scissors, cameras, watches, mobile phones); buildings (houses, churches, schools, shops); product packages (Bio powder, color powder, bio tablets, color tablets); shapes (circles, triangles, squares, pentagons). Random samples from the set of images formed 2AFC and 4AFC presentations. The spatial position of each image was randomly selected for each trial, so, for example, the red square could appear in any of the 2/4/8 locations for presentations from the squares subcategory.

2007). Individuals were displayed in a radial fashion (see Figure 2) with their center at 8° from a central fixation cross-hair presented immediately prior to the stimuli, that is, the stimuli were presented peripherally with respect to the fixated location at the start of each screen presentation.

Participants were presented with samples of images and instructed to look for the image they most preferred. The following variables were manipulated:

Display size: Two, four, or eight images randomly placed at the cardinal/ordinal locations. Examples of the locations of two, four, or eight image presentations are shown in Figure 2.

Presentation duration: The images were displayed for 1,500, 2,500, or 5,000 ms before being obliterated by a number indicating the button to be pressed to select the image.

Image category: The images in a single screen presentation were taken from a single subcategory of the following categories: objects, buildings, commercial products, or shapes.

Image subcategory: Each of the four image categories comprised four semantically related subcategories, for example, the shape category featured differently colored circles, squares, triangles, and pentagons. The full list of subcategories is given in Figure 1.
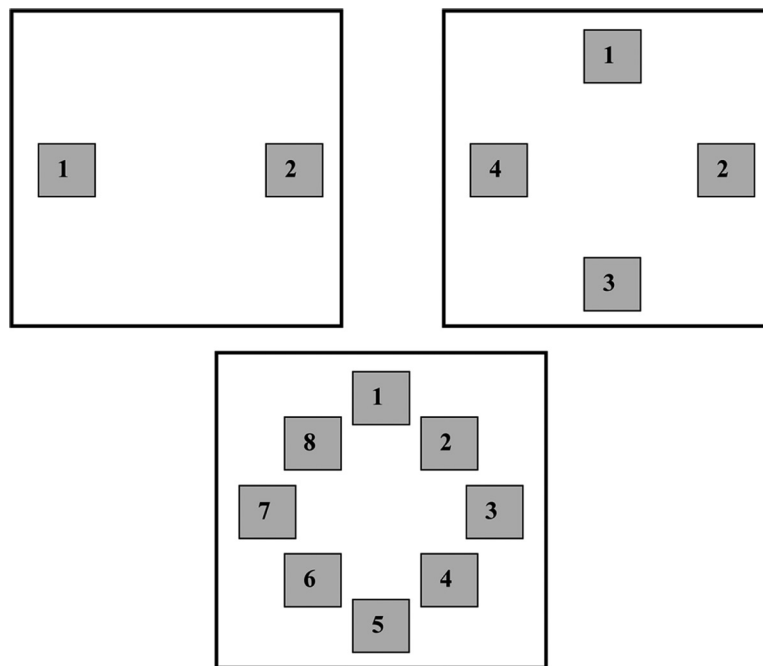
Figure 2. Image positions when two, four, or eight images were presented in a single screen presentation. The numbers shown in each configuration correspond to the button on the numeric keypad used to identify preference for the image at that location.

Each subcategory comprised eight distinct images taken from Google Image searches on the category names that were converted to a 256-color bitmap and resized for presentation. The full image set is shown in the 8AFC configuration in Figure 1.

2.3 Timing

After a central fixation cross-hair was displayed for 1,000 ms, participants were presented with a screen of two, four, or eight images for evaluation for 1,500, 2,500, or 5,000 ms. After this period, numbers corresponding to buttons on a keypad were displayed at the stimulus locations, obliterating a large part of each image; this was to ensure that enough of the image was visible so that participants did not need to spend time encoding the spatial location of their preferred image during the image presentation stage but were unable to continue evaluating the images after the image presentation time. Participants then indicated their preferred image by pressing the number corresponding to the image (see Figure 2) on a physical keypad. No time limit was imposed on the manual response, but participants were encouraged to do this as quickly as possible. On making their selection, a color noise mask was presented for 250 ms to eradicate after images, before the next trial began.

2.4 Eye tracking

A Cambridge Research Systems 50-Hz Video Eye-Tracker (CRS VET) was used to sample raw x and y coordinates for gaze location every 20 ms. Fixations were defined as distinct periods of 100 ms or more with gaze location remaining within a 2.5-mm window on the screen, that is, 0.25° of visual angle at a viewing distance of 57 cm. Areas of interest extending 5 mm beyond the edges of each image were used, which corresponds with the 0.5° spatial resolution of the eye tracker. Oculomotor measures were calculated as follows:

Duration was calculated by simply summing the duration of all fixations for each image.

The sequence of first fixations was recorded on the basis of the first fixation on an image location, for example, in the 4AFC condition, a first fixation sequence of (Images 1, 3, or 4) would indicate that Image 1 was fixated first, then Image 3 followed by Image 4, with possible off-target fixations and refixations of Image 1 prior to the first fixation of Image 3, possible off-target fixations and refixations of Images 1 and 3 prior to the first fixation of Image 4 and Image 2 never being fixated.

Last fixation was recorded as simply the last location fixated prior to images being obliterated by the manual selection numbers.

Returns were defined as distinct refixations, that is, they were separated by at least one fixation of a different image or empty space.

2.5   Experiment 1

Experiment 1 evaluated the effects that array size and presentation length have on the predictive potential of the oculomotor measures. It comprised two stimulus presentations for each of the four categories (random selection of subcategory), for presentation durations of 1,500, 2,500, and 5,000 ms and display sizes of two, four, and eight, that is, 72 trials, for each of the seven participants. All conditions were randomly interleaved, meaning that the participant was unaware of how much time they had to locate their preferred image.

3   Results

Each of the oculomotor measures (duration, first fixation, last fixation, returns) was examined for its correlation with the manually chosen image, that is, the number of times the longest fixation, first fixation, last fixation, and most frequently returned to location matched the manually chosen preferred location was calculated for each participant, for each of the display sizes. If a measure is uncorrelated, then the number of matches should be at chance level (i.e., 50% in the 2AFC condition, 25% in the 4AFC condition, 12.5% in the 8AFC condition), meaning that the ratio of matches to chance (proportion of correct matches/chance level for the presentation size) should be 1, with results higher than 1 indicating a positive correlation and results less than 1 indicating a negative correlation. The results for the ratios of correct matches to chance across the 3 display conditions and 3 presentation durations are shown in Figure 3.

The results of two-tailed repeated measures ANOVAs, corrected using Bonferroni, for each of the oculomotor measures showed a significant main effect from the number of images presented: fixation duration, $F(2, 12) = 11.367$, $p < .005$; first fixation, $F(2, 12) = 9.270$, $p < .01$; last fixation, $F(2, 12) = 13.095$, $p < .005$; returns, $F(2, 12) = 36.619$, $p < .001$. Two-tailed post hoc comparisons between the three AFC levels identified significant differences between presentation of two and four images for fixation duration ($p < .05$), presentation of two and eight images for last fixation ($p < .05$) and returns ($p < .01$), and presentation of four and eight images for last fixation ($p < .05$) and returns ($p < .005$). The results of two-tailed repeated measures ANOVAs, corrected using Bonferroni, for each of the oculomotor measures showed only a marginal main effect from the presentation duration on fixation duration, $F(2, 12) = 4.913$, $p = .056$, with no significant effect on any of the other measures ($p > .05$). Similarly, the interaction between number of images and presentation duration was only significant for fixation duration, $F(4, 24) = 4.284$, $p < .05$. Post hoc analysis of this interaction for fixation duration revealed a significant effect from presentation in the 8AFC condition, $F(2, 12) = 5.825$, $p < .05$, and from the presentation size in the 1,500 ms, $F(2, 12) = 5.427$, $p < .05$, and 5,000 ms, $F(2, 12) = 19.906$, $p < .001$, conditions.

The results for the returns measure show performance below chance in both the 2AFC and 4AFC conditions. This is largely because returns were very scarce in those trials, suggesting that participants either did not have enough time to complete first fixations on all the stimuli, as would certainly be the case in the 1,500-ms presentation for eight images, or were able to maintain representations of up to four locations in working memory from initial scanning of the images prior to any prolonged fixation based on preference. Overall, the results clearly show an increased performance above chance on all measures when the participant is presented with more choice, that is, the oculomotor statistics are more informative when there are more candidates to be evaluated, with little or no effect from extended presentation time, especially for the single fixation measures.

3.1   Experiment 2

The second experiment evaluated the effects that image category and presentation length have on the predictive potential of the oculomotor measures. It comprised four stimulus presentations with a fixed display size of 8 for each of the 16 subcategories, for presentation durations of 1,500, 2,500, and 5,000 ms each, that is, 192 trials for each of the three participants, selected from the initial seven. This meant that the participants saw images from Experiment 1 supplemented by images they had never seen be-
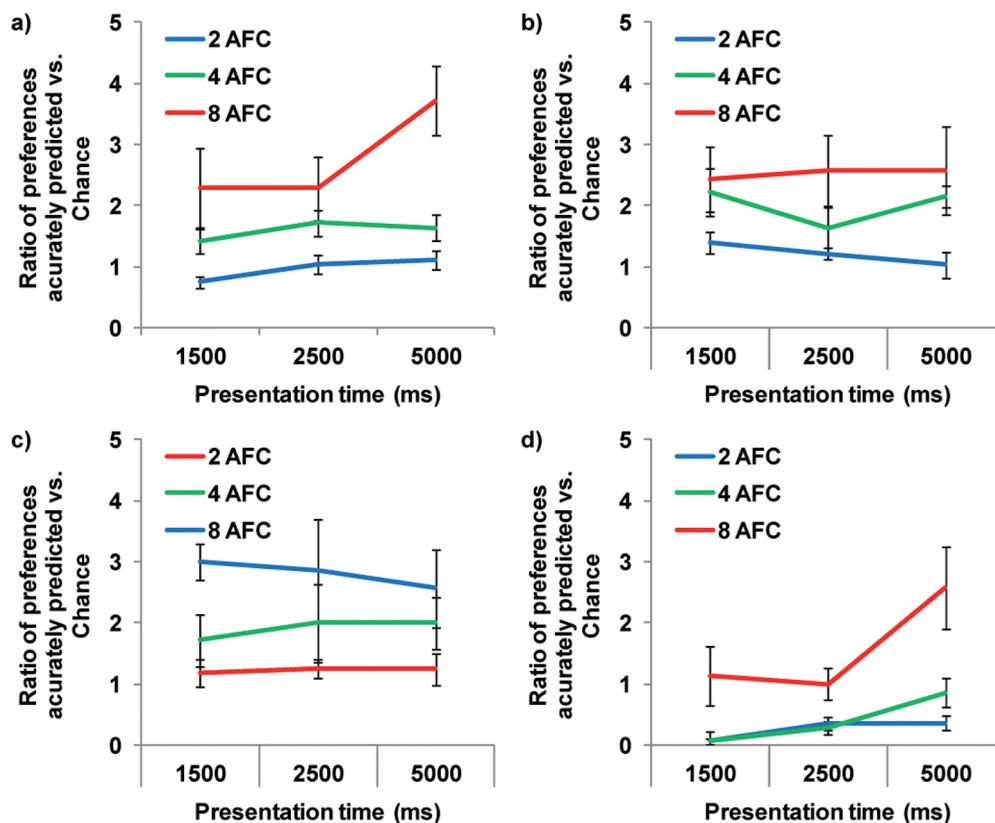
Figure 3. Ratio of number of times the four oculomotor measure corresponds with the subsequent manual AFC selection, relative to chance levels: (a) fixation duration; (b) first fixation; (c) last fixation; (d) returns. Consistent improvements are visible across measures with increased number of images, with no consistent effect from presentation duration. Results are averaged over 24 trials for each condition, $N = 7$. Bars show standard error.

fore. Randomization of the location of the images meant that participants were highly unlikely to see the same arrangement viewed in the 8AFC condition in the first experiment. As before, all conditions were randomly interleaved, meaning that the participant was unaware of how much time they had to locate their preferred image.

4   Results

Figure 4 shows little difference in the overall performance of the four oculomotor measures for the four categories, with all performing better than chance. The results of two-tailed repeated measures ANOVAs, corrected using Bonferroni, for each of the oculomotor measures showed only a significant main effect from the category for first fixation ($F(3, 6) = 11.271$, $p < .05$). There was a significant main effect from the presentation duration for the returns measure ($F(2, 4) = 17.777$, $p < .05$), and a significant main effect from the interaction of category and duration for first fixation ($F(6, 12) = 4.723$, $p < .05$). No significant effect was observed between subcategories of image type for any of the four major categories.

These results are similar to those for the 8AFC condition in the Experiment 1 and suggest that the image type had little effect on the measures. The strong performance of the first fixation measure in the 1,500-ms shape is probably a result of the simplicity of the eight images being used in each of the trials allowing an almost reflexive saccade to be made to the preferred color in later trials. The consistent improvement in the performance of the returns measure for longer presentation times across all image types confirms the importance of this behavior when making comparative preference choices and suggests that in the absence of prior knowledge of the image characteristics, this might be the best single indicator of preference.
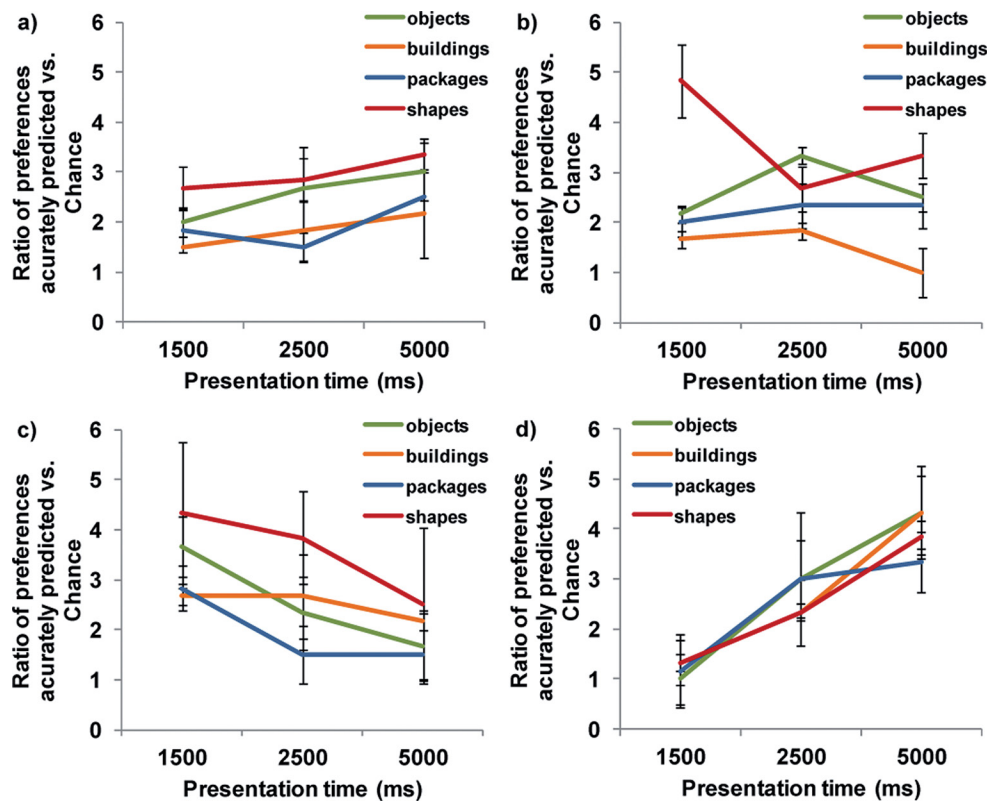
Figure 4. Ratio of number of times the four oculomotor measure corresponds with the subsequent manual AFC selection, relative to chance levels: (a) fixation duration; (b) first fixation; (c) last fixation; (d) returns. Results are averaged by image category (object, building, package, or shape) over 16 trials for each category and each presentation duration (192 trials in total), $N = 3$. Bars show standard error.

4.1   Oculomotor signature

The results from Experiments 1 and 2 suggest that predictive potential of each of the oculomotor measures depends on the size of the stimulus sample and duration of presentation while only being mildly affected by the image category and, in this case, associated image complexity. Figure 5 shows an example of the cumulative fixation plot for a single participant (P4) in a single trial (5,000 ms, 8AFC, buildings category). From this plot, it can be seen that neither the first image fixated nor the image being fixated at the end of the trial was identified as the preferred image by the viewer, highlighting the problem with single fixation measures, as the first fixation is made without information about the other choices available, and the final fixation might be made after a decision has been made.

The number of times first fixation matches the selected preference, expressed as a percentage of all trials in which at least one image was accurately fixated (396 of the $6 \times 72 = 432$ trials reported in Experiment 1), is shown in Figure 6 for all items presented. Here, it can clearly be seen that the first item fixated is less likely to reflect the final selection as both presentation duration and presentation size increases and although always above chance for the display size, as choice and presentation time increase, late fixated items are increasingly likely to correlate with the manual preference selection. However, this presents a problem for a cumulative duration-based fitness evaluation, as the later an item is first fixated, the less chance it has to accumulate a long fixation duration. This suggests the need to weight the fixation duration according to the lateness of the first fixation for each item.

On the basis of these observations and the group results for Experiments 1 and 2, a new oculomotor measure, or "signature," was calculated using a multiplicative weighting which effectively scales the cumulative fixation duration, according to the likelihood that this image will be the preferred image, using the sequence of first fixations and the number of returns (distinct refixations after first fixation) to the image, as follows:
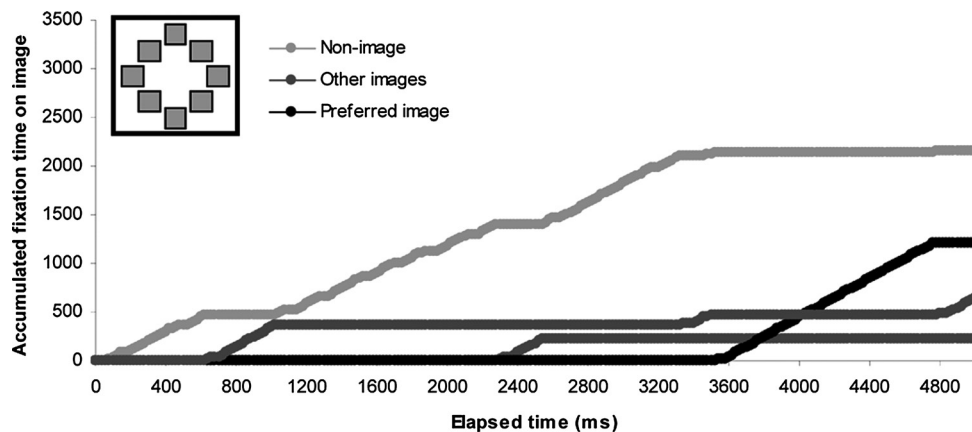
Figure 5. Example of a cumulative fixation plot from a single trial for participant P4 in Experiment 1 (Buildings category). Fixations on the preferred image are shown, as well as those on each of the other images which attracted at least one fixation, and for fixations outside of the image areas of interest ("nonimages"). $N = 1$ observer and presentation.

Let f_i be the amount of time spent fixating on item i, where $i = 1$ to $i = display\ size$. In Experiments 1 and 2, display size was 2, 4, or 8.

Let F be the total amount of time spent fixating on images for $i = 1$ to display size, that is, f_i summed for all i.

Then,

$$C_i = \frac{f_i}{F} \tag{1}$$

is the relative cumulative fixation duration for each image. Note that C_i is always in the range of 0 to 1.

Now, let s_i be the order of the first fixation of image i, divided by the display size. For example, if image i was the first image to be fixated in the 8AFC condition, $s_i = 1/8$; if image i was the second image to be fixated in the 2AFC condition, $s_i = 2/2\ (=1)$. Thus, s_i is always in the range of 0 to 1.

Let r_i be the number of returns (see earlier definition) of items i, that is, the total number of distinct fixations excluding the initial fixation. For example, if image i is fixated for three distinct periods during a tracking session, $r_i = 2$, if it is only fixated once, then $r_i = 0$. Note that r_i is effectively in the
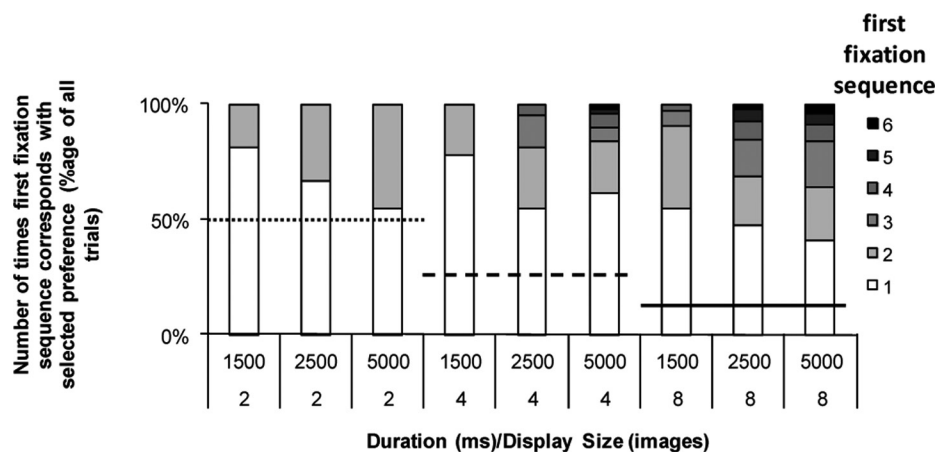

Figure 6. Proportion of any particular item in the stimulus set matching selected preference, with gray level indicating the sequence of fixation. Sequence of fixations is simply the order in which images are fixated for the first time. Bars show the proportion of items matching preference by sequence number totaled over all trials from Experiment 1a in which at least one image was successfully fixated (396 trials in total/all image categories). Horizontal lines show the chance levels for the 2AFC (50%, dotted line), 4AFC (25%, dashed line), and 8AFC (12.5%, solid line). The increased likelihood that an image other than the first fixated image will match the selected preference as display duration and display size increase is clearly visible.
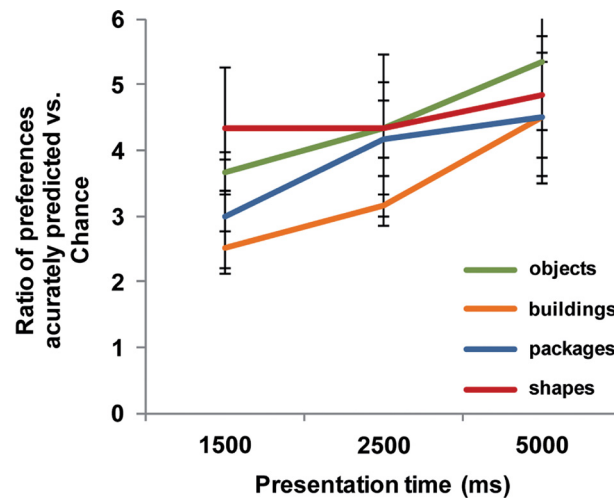
Figure 7. Effect of using weighted combinations of oculomotor measures with the new subjective fitness function described in *Equation 3*. The new measure performs consistently across all four image types and performance improves monotonically with presentation time. All presentations were in the 8AFC configuration, 5,000 ms. Error bars show standard error over all trials, $N = 3$.

range of 0 to ∞, but its maximum value for any single tracking session is the total number of distinct fixations minus 1. For this reason, r_i is normalized to be in the range 0 to 1 by dividing by the total number of refixations for all items $i = 1$ to display size.

Then,

$$w_i = \frac{s_i + r_i}{2} \qquad (2)$$

is a weighting which represents the sustained interest in the image following its initial fixation. Note that w_i is in the range of 0 to 1 for all images.

Finally,

$$a_i = w_i C_i \qquad (3)$$

is the weighted relative cumulative fixation duration scaled according to sustained interest and forms the proposed oculomotor signature of aesthetic preference.

Figure 7 shows the results for the new signature applied to each image type using the gaze data from Experiment 2. Overall, the new ratio is now positively correlated for all presentation times and all image types, matching manual preference selection significantly better than chance at all presentation times: 1,500 ms, mean ratio = 3.378, $t(2) = 5.529$, $p < .05$; one-tailed; 2,500 ms, mean ratio = 4.000, $t(2) = 5.322$, $p < 0.05$; one-tailed; 5,000 ms, mean ratio = 4.7917, $t(2) = 4.149$, $p < 0.05$; one-tailed. Results of one-tailed *t*-test comparisons between the new signature and the individual oculomotor measures showed significant improvements at 1,500 ms for fixation duration, $t(2) = 4.200$, $p < .05$, and returns, $t(2) = 3.560$, $p < .05$; at 2,500 ms for fixation duration, $t(2) = 3.308$, $p < .05$, first fixation, $t(2) = 4.015$, $p < 0.05$, and returns, $t(2) = 16.000$, $p < .005$; and at 5,000 ms for first fixation, $t(2) = 3.004$, $p < .05$, and last fixation, $t(2) = 6.513$, $p < .05$. These results are consistent with a signature that now accounts for some of the time-dependent deficiencies of some of the individual measures.

5 Discussion

The results from Experiments 1 and 2 confirm that accumulated fixation duration correlates with preference better than chance, and combining it with measures such as the order of first fixation and the number of returns to an image, which relate to sustained interest in the image, results in an oculomotor signature which correlates significantly better with the conscious preference decision (see Figure 7).

These results confirm a strong relationship between the subjective decisions we consciously make and the eye movements which precede expression of those conscious decisions using conventional methods such as button presses and questionnaire responses. More fundamentally, they demon-

strate the usability of preferential looking paradigms with normal, healthy adult participants and not just with observers who would have difficulty understanding a task or communicating a response (Dobson, 1983; Teller, 1979). In addition, this work provides further support for the gaze cascade model (Glaholt & Reingold, 2009; Shimojo et al., 2003) which suggests an increase in sustained interest that is visible from fixations late in the presentation period. This correlation between accumulated fixation duration and consciously expressed preference also lies at the heart of the gaze-driven evolutionary algorithm (Holmes & Zanker, 2008) and the results from Experiments 1 and 2 further support the use of eye movements as a fitness measure for aesthetic preference. This is because the fixation duration represents a "better than chance" indication of preference and so can be used to provide the selection pressure needed to drive an evolutionary optimization process. Early versions of the algorithm relied on the specification of a task, such as "look for your preferred shape," to motivate the participant's eye movements. While the active selection of a pleasing and stimulating visual environment has been suggested to be a background task for humans (Berlyne, 1971; Biederman & Vessel, 2006), an oculomotor signature, such as that developed here, which correlates with such a task could now be incorporated in a free viewing version of the paradigm, which simply requires the participant to look at the images evolving on screen without any explicit task specification. The removal of task specification, as well as consciously articulated response, would provide a language-independent means for exploring complex preferences relating to aesthetics. Early results using stimuli inspired by Op Artist Bridget Riley suggest that the signature in *Equation 3* can be used to generate aesthetically pleasing complex patterns in the absence of a specific task (Holmes, 2010; Zanker, Voigt, & Holmes, 2010).

Clearly, the oculomotor signature developed here does not represent a comprehensively optimized model for preferential looking, which would need to incorporate parameters for the presentation duration, the image complexity, and the complexity of the decision-making task, perhaps using pupilometry with respect to the amount and variety of the choice (Holmes, 2010). However, the improved correlation between looking and preference could facilitate better fitness estimation than fixation duration alone in gaze-based computer interactions where aesthetic preference is used to drive an optimization algorithm (Holmes & Zanker, 2008). The improved correlation between the new oculomotor signature and aesthetic preference would increase the selection pressure in that evolutionary algorithm and accelerate the location of highly fit solutions (Goldberg, 1989) allowing more complex image spaces to be searched without increasing the demands on the participant (Takagi, 1998, 2001, 2009). This has important consequences for the nonreductionist investigation of visual aesthetic preference, since such an approach is dependent on the availability of techniques which can manipulate and search complex images.

The present experiments highlight the importance of the presentation time for the preferential looking paradigm with normally functioning adults, especially with a measure such as the returns measure used here. If presentations are too short for the number and complexity of images being compared, then there is insufficient time to complete the bottom-up–driven location phase of visual search before initiating the top-down–driven identification phase (Locher et al., 2007; Van der Lans et al., 2008), which is important for a meaningful response to an aesthetic preference task (Glaholt & Reingold, 2009). With short presentation times, it is possible that gaze-based measures such as the one proposed here might better reflect image saliency than image preference (Itti & Koch, 2000), although more recent evidence suggests that the looking behavior changes less during the course of viewing than previously thought (Tatler, Hayhoe, Land, & Ballard, 2007), meaning that the benefit from prolonged exposure might simply be the result of longer fixations later in the time course (Glaholt & Reingold, 2009). However, this interaction between a feed-forward and a feedback loop raises a question of causality: Does fixation precede preference or preference precede fixation? Clearly, correlation results such as those presented here cannot resolve this, and further investigation of this is needed. However, the inclusion of this signature in a task-independent gaze-driven evolutionary algorithm does present an opportunity for learning general trends in a participant's preference, through the frequencies of specific image features in final population, and these can then be used to generate previously unseen stimuli for a forced-choice task where the presentation duration is manipulated to verify the effects of timing on actual preference (Zanker et al., 2010).

The oculomotor signature developed here is a quantitative behavioral measure which can be used to evaluate aesthetic preference with highly complex visual stimuli without any prior knowledge as to the contribution to that preference from individual features. As such, it already addresses several concerns with experimental research into aesthetics including the ambiguity of qualitative

results (Arnheim, 1974; Berlyne, 1971; Hassenzahl, 2008), the limitations imposed by the reductionist methodologies (Arnheim, 1974) and the unreliability of conscious reflection on subjective decisions (Nisbett & Wilson, 1977). Advances in technology are resulting in eye trackers that are more accurate, easier to use, and more affordable. Moreover, off-the-shelf portable eye trackers now present the possibility to study such signature oculomotor behaviors away from the laboratory facilitating quantitative evaluation of artworks in their original forms as they were meant to be seen. It seems that in addition to beauty itself and understanding of beauty might lie in the eyes of the beholder.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron, 21*, 373–383. doi:10.1016/S0896-6273(00)80546-2

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103*(1), 62–70.. doi:10.1037/h0036799

Arnheim, R. (1974). *Art and visual perception: The new version*. Berkley, CA: University of California Press

Bentley, P. J. (1999). *Evolutionary design by computers*. San Francisco, CA: Morgan Kaufmann Inc.. doi:10.1017/S0890060499133018

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York, NY: Appleton-Century-Crofts

Biederman, I, & Vessel, E. A. (2006). Perceptual pleasure and the brain. *American Scientist, 94*(May-June), 249–255. doi:10.1511/2006.59.995

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago, IL: University of Chicago Press

Carbon, C-C., Hutzler, F., & Minge, M. (2006). Innovativeness in design investigated by eye-movements and pupilometry. Psychology Science, 48(2), 173-186.

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention bottom-up versus top-down. *Current Biology, 14*(19), 850–852. doi:10.1016/j.cub.2004.09.041

Dobson, V. (1983). Clinical applications of preferential looking measures of visual acuity. *Behavioural Brain Research, 10*, 25–38. doi:10.1016/0166-4328(83)90147-X

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458. doi:10.1037/0033-295X.96.3.433

Glaholt, M. G., & Reingold, E. M. (2009). Stimulus exposure and gaze bias: A further test of the gaze cascade model. *Attention Perception Psychophysics, 71*(3), 445–450. doi:10.3758/APP.71.3.445

Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. Boston, MA: Addison-Wesley Longman Inc

Hassenzahl, M. (2008). Aesthetics in interactive products: correlates and consequences of beauty. In H. N. J. Schifferstein & P. Hekkert (Eds.), *Product experience* (pp. 287–302). Amsterdam, the Netherlands: Elsevier. doi:10.1016/B978-008045089-6.50013-7

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*(11) 498–504. doi:10.1016/j.tics.2003.09.006

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading in scene perception* (pp. 269–293). Amsterdam, the Netherlands: Elsevier. doi:10.1016/B978-008043361-5/50002-X

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Cambridge MA: MIT Press

Holmes, T. (2010). *Interactive evolutionary computation driven by gaze: A new paradigm for experimental aesthetics and beyond*. (PhD Thesis, Royal Holloway, University of London, Surrey, UK) http://royalholloway.academia.edu/TimHolmes/Papers/195130/Gaze_Driven_Interactive_Evolutionary_Computation_-_A_New_Paradigm_for_Experimental_Aesthetics_and_Beyond

Holmes, T., & Zanker, J. (2008). *Eye on the prize: Using overt visual attention to drive fitness for interactive evolutionary computation*. Paper presented at the Eighth Annual Conference on Genetic and Evolutionary Computation, Seattle, WA, 1531–1538 doi: 10.1145/1389095.1389390

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development, 13*(4), 341–348. doi:10.1002/icd.381

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506. doi:10.1016/S0042-6989(99)00163-7

Land, M. F. (2007). Fixation strategies during active behavior: A brief history. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain*. (pp. 75–95). Oxford, England: Elsevier. doi:10.1016/B978-008044980-7/50006-9

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions: Biological Sciences, 352*(1358), 1231–1239. doi:10.1098/rstb.1997.0105

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience, 4*(5), 533–539. doi:10.1016/S1053-8119(01)92231-1

Locher, P. J., Krupinski, E. A., Mello-Thoms, C., & Nodine, C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision, 21*(1–2), 55–77. doi:10.1163/156856807782753868

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. doi:10.1037/0033-295X.84.3.231

Pannasch, S., Helmert, J. R., Roth, K., Herbold, A. K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research, 2*(2), 1–19

Pieters, R., & Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research in Marketing, 16*(1), 1–16. doi:10.1016/S0167-8116(98)00022-6

Plumhoff, J. E., & Schirillo, J. A. (2009). Mondrian, eye movements, and the oblique effect. *Perception, 38*(5), 719–731. doi:10.1068/p6160

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology, 46*, 342–376. doi:10.1037/h0084328

Santella, A., & DeCarlo, D. (2004). Robust clustering of eye movement recordings for quantification of visual interest. In A. T. Duchowski, & R. Vertegaal (Chair), *Proceedings of the Eye Tracking Research and Applications (ETRA) Symposium 2004* (pp. 27–34). New York, NY: ACM.. doi:10.1145/968363.968368

Shen, J., Elahipanah, A., & Reingold, E. M. (2007). Effects of context and instruction on the guidance of eye movements during a conjunctive visual search task. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain*. (pp. 75–95). Oxford, England: Elsevier. doi:10.1016/B978-008044980-7/50030-6

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience, 6*(12), 1317–1322. doi:10.1038/nn1150

Stratton, G. M. (1906). Symmetry, linear illusions and the movements of the eye. *Psychological Review, 13*, 82–96. doi:10.1037/h0072441

Takagi, H. (1998). *Interactive evolutionary computation: System optimization based on human subjective evaluation*. Paper presented at the IEEE International Conference on Intelligent Engineering Systems (INES '98), New York, NY. (pp. 2933–2966). doi:10.1145/1274000.1274100

Takagi, H. (2001). Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE, 89*(9), 1275–1296. doi:10.1109/5.949485

Takagi, H. (2009). New IEC research and frameworks. In J. Kacprzyk. (Ed) *Studies in "computational intelligence"*. (pp. 65–76). Berlin, Germany: Springer. doi:10.1007/978-3-642-03633-0_4

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research, 45*, 643–659. doi:10.1016/j.visres.2004.09.017

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5), 1–23. doi:10.1167/11.5.5

Tatler, B. W., Wade, N. J., & Kaulard, K. (2007). Examining art: dissociating pattern and perceptual influences on oculomotor behaviour. *Spatial Vision, 21*(1–2), 165–184. doi:10.1163/156856807782753903

Teller, D. Y. (1979). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior & Development, 2*, 135–153. doi:10.1016/S0163-6383(79)80016-8

Van der Lans, R, Pieters, R, & Wedel, M. (2008). Competitive brand salience. *Marketing Science Articles in Advance*, 1–10. doi:10.1287/mksc.1070.0327

Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the XXVII Conference of the Cognitive Science Society*. (pp. 2283–2288). Stresa, Italy

Wooding, D. S. (2002). Eye movements of large populations: II Deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods Instruments and Computers, 34*(4), 518–528. doi:10.3758/BF03195481

Wooding, D. S., Mugglestone, M. D., Purdy, K. J., & Gale, A. G. (2002). Eye movements of large populations: I. Implementation and performance of an autonomous public eye tracker. *Behavior Research Methods Instruments and Computers, 34*(4), 509–517. doi:10.3758/BF03195480

Yarbus, A. L. (1967). *Eye movements and vision*. New York NY: Plenum Press

Zanker, J., Voigt, K., & Holmes, T. (2010). Evolving illusory motion using eye-movements. *Journal of Vision, 10*(7), 162. doi:10.1167/10.7.162

Tim Holmes received his first degree in Mathematics from the University of Sheffield before completing a postgraduate diploma in Psychology at Royal Holloway, University of London, where he remained for his PhD, under supervision from Professor Johannes Zanker, to explore the relationship between eye-movements and subjective preferences using an innovative methodology based around evolutionary algorithms. Tim's interest in aesthetics and real world applications continues in his ongoing research into visual attention and package design, gaze based assistive technologies and other biometric indicators of preference. For more information visit http://pure.rhul.ac.uk/portal/en/persons/tim-holmes%28e94ddf63-1a12-457a-8df5-61325c5e9ad0%29.html

Johannes M Zanker received his PhD in Zoology from University Tübingen, Germany, and held appointments at various institutions, including the Max-Planck-Institut in Tübingen, Department of Psychology at UCL in England, and Centre for Visual Sciences, at the Australian National University in Canberra, before joining the Psychology Department at RHUL, where he is professor of Neuroscience in the Computational Vision Lab. His main research intereste are related to visual perception and computational models of neural information processing, with a wide range of projects such motion psychohysics, the role of eye movements in illusions, computational aesthetics, or visual ecology. For more information visit http://www.pc.rhul.ac.uk/staff/J.Zanker/.