What Gaze Fixation and Pupil Dilation Can Tell Us About Perceived Differences between Abstract Art by Artists vs. by Children and Animals

(preprint)

Sergio A. Alvarez

Boston College

Ellen Winner

Boston College

Angelina Hawley-Dolan

Boston College

Leslie Snapper

Boston College

Running Head: EYE GAZE AND PUPIL DILATION

Corresponding Author:

Ellen Winner

winner@bc.edu

Dept of Psychology

McGuinn 300

Boston College

Chestnut Hill, MA 02467

Abstract

People with no arts background often misunderstand abstract art as requiring no skill. However, adults with no art background discriminate paintings by abstract expressionists from superficially similar works by children and animals (Hawley-Dolan & Winner, 2011). We tested whether participants show different visual exploration when looking at paintings by artists' vs. children or animals. Participants sat at an eye tracker and viewed paintings by artists paired with "similar" paintings by children or animals, and were asked which they preferred and which was better. Mean duration of eye gaze fixations, total fixation time, and spatial extent of visual exploration were greater to the artist than child/animal images in response to quality but not preference. Pupil dilation was greater to the artist images in response to both questions, and greater in response to the quality than preference question. Explicit selections of images paralleled total fixation times: participants selected at chance for preference, but selected the artist images above chance in response to quality. Results show that lay adults respond differently on both an implicit as well as explicit measure when thinking about preference vs. quality in art and discriminate abstract paintings by artists from superficially similar works by children and animals, despite the popular misconception by the average viewer that "my kid could have done that."

What Eye Gaze and Pupil Dilation Can Tell Us About Perceived Differences between Abstract Art by Artists vs. by Children and Animals

The question of what we like and how we judge quality in art is part of the larger question of how we make valence-based judgments about objects and experiences in the absence of an objective metric (Eisner, 2004). How do we decide what we like, and how do we decide what we think is good? The domain of art provides a lens through which to study this problem -- in particular, the domain of abstract painting, which is little understood by the lay public.

Abstract art is sometimes disparaged as requiring little skill and as indistinguishable from works by preschoolers. We have all heard some variant of the claim that "my kid could have done that." Of course, artists and art historians reject this view as naïve, and recognize the intention, planning, and skill behind canvases that may appear as if they were made without skill (Temkin, 2013; Varnedoe, 2006). Nonetheless, there have been embarrassing incidents in which even connoisseurs have been fooled into thinking that a work by a child was in fact a work by an emerging, yet unknown, abstract expressionist (Bar-Lev, 2007; Chittenden, 2007; Hussain, 1965).

Recent evidence should help to put to rest the claim that for lay adults, abstract expressionist paintings are indistinguishable from those of a preschooler. Observers with no familiarity with abstract art are able to distinguish works by abstract expressionists from superficially similar works by children and animals, as shown by more often preferring and judging as better works by artists when compared to those by children and animals. Not only do adults lean towards works by artists when thinking about what they like and judge as better, but they also perceive more intentionality and structure in the works by the artists (Hawley-Dolan & Winner, 2011; Snapper, Oranc, Hawley-Dolan, Nissel, & Winner, 2014). Even children as young as 4-7 are able to distinguish works by artists from those by children and animals, although their preferences and value judgments are not aligned with those of adults (Nissel, Hawley-Dolan, & Winner, 2014). These demonstrations make clear that there is a perceivable difference that the untrained observer can detect between abstract works by masters and those by children and animals.

In Hawley-Dolan and Winner's (2011) initial study on this question, 30 paintings by famous abstract expressionists were paired with paintings by children or animals (apes, monkeys, elephants). The pairings were constructed holistically to look as similar as possible in at least two of the following characteristics: color, line quality, brushstroke, medium, and composition. An example of one of these pairs can be found in Hawley-Dolan and Winner. Participants viewed these pairs and were asked two questions about each one: "Which image do you like more?" (preference) and "Which image do you think is the better work of art?" (quality). Preference and quality judgments are different ways of reacting to a work of art. A preference is more of an automatic, perceptually based response, while a judgment about quality is a more objective, cognitively based judgment (Hagtvedt, Hadtvedt, & Patrick, 2008; Leder, Belke, Oeberst, & Agustin, 2004; Zajonc, 1980). Pairs were presented in three labelling conditions: the first ten pairs were presented unlabeled; the next 20 had labels: one was labeled "artist" and the other was labeled either "child," "monkey," or "elephant." Ten of the labelled pairs were correctly labelled, and ten were reverse (wrongly) labeled, with both kinds randomly intermixed.

Across all labeling conditions, participants preferred and judged as better the works by the artists significantly more than those by children and animals. Of course it is not surprising that people showed a bias towards the works correctly labelled "artist", and this bias consistent with the finding that people's evaluations of art works are influenced by knowledge of the artist's fame (Chittenden, 2007; Farnsworth & Misumi, 1931). What is noteworthy, however, is that people showed the same bias for the first ten pairs, for which there were no labels. Participants were not told that they were actually choosing between works by artists and works by children and animals, and so the prestige of the label "artist" could not have been a factor. Even more surprising is that people continued to be biased towards artists' works when these works were labeled incorrectly as by a child or animal, thereby overruling prestige in favor of what they actually saw in the painting. A follow up study showed that people can discriminate these two kinds of images even when they are presented singly rather than paired, and that this discrimination is based on perceiving more intentionality and greater visual structure in the works by artists (Snapper, Oranc, Hawley-Dolan, Nissel, & Winner, 2014).

The above findings are all based on explicit measures in which participants respond verbally to a preference and a quality question. Do preference and quality judgments also reveal themselves in implicit ways, and are these implicit signs consistent with explicit responses? This was the question addressed here. Using a subset of the pairs from Hawley-Dolan and Winner (2011), we asked whether lay observers gaze differently at works by artists than those by children and animals, and also whether they show greater pupil dilation when looking at the works by the masters. We examined several implicit measures in response to both the preference and the quality questions.

We assessed eye gaze and pupil dilation using an eye tracker. Eye-tracking is a technique that has been used in an attempt to understand how viewers look at art – where they look and when, as well as how long they look (Busswell, 1935; Locher, 2006; Nodine and Krupinski, 2003; Nodine & McGinnis, 1983; Yarbus, 1967). Typically eye tracking studies examine viewers' scan paths as they look at a painting to determine how people visually analyze a work of art. Here we did not examine the relationship between scan path and visual aspects of the specific art works (e.g., composition), but instead assessed the values of several measurements related to the eye gaze path, including cumulative preferential eye gaze fixation time (total fixation time), a measure that is closely related to looking time, which is widely used in infant research to examine what infants are drawn to look at (e.g., Cohen & Cashon, 2003). Looking time to works of art has been theorized to be an index of cognitive analysis since individuals look longer at what they are thinking about (Kapoula, Daunys, Herbez, & Yang, 2009; Locher, 1996, 2006; Locher, Krupinski, MellopThoms, & Nodine, 2006; Molnar, 1981; Yarbus, 1967), as well as pleasure, since looking time then becomes rewarding (Plumhoff &Schirillo, 2009). Looking time to scenes has also been argued to reflect cognitive analysis (e.g., Henderson & Hollingworth, 1998). We predicted greater total fixation time on works by artists compared to works by children and animals. Other eye gaze variables were also considered, such as the mean duration of individual fixations, the total length of the scan path, and the number of alternations between the two images of a pair. The latter two are cumulative measures related to saccadic eye motions, which, like fixation time, have been connected with cognitive analysis in visual scanning and reading tasks (Liversedge and Findlay, 2000).

Pupil dilation has been shown to be an implicit measure of looking pleasure. Pupils dilate more in response to looking at images perceived as pleasing compared to those perceived as not pleasing, as shown decades ago by Hess (1965) and Hess and Polt (1960). However, pupil size also increases in response to lower luminance, and Hess's studies failed to control for luminance. This confound was corrected in a study examining pupil size in response to abstract paintings by the artist Piet Mondrian (Johnson, Muday, & Schirillo, 2010). Images by Mondrian were presented in various rotations and participants rated each image for pleasingness while their pupil diameter was measured. Pupil diameter was positively correlated with pleasingness ratings. In the two studies reported here we controlled for luminance.

Pupil dilation has also been shown to be an implicit measure of mental effort (e.g., Beatty, 1982; Johnson, Miller-Singley, Peckam, Johnson, & Bunge, 2014; Just & Carpenter, 1992; Kahneman, 1971). Given that pupil dilation is a measure of pleasure, we predicted greater pupil dilation for the artist images. And given that pupil dilation is a measure of mental effort, we also predicted greater pupil dilation in response to the quality than the preference question, irrespective of which image was looked at longer.

The study reported here was approved by the Boston College Institutional Review Board.

Methods

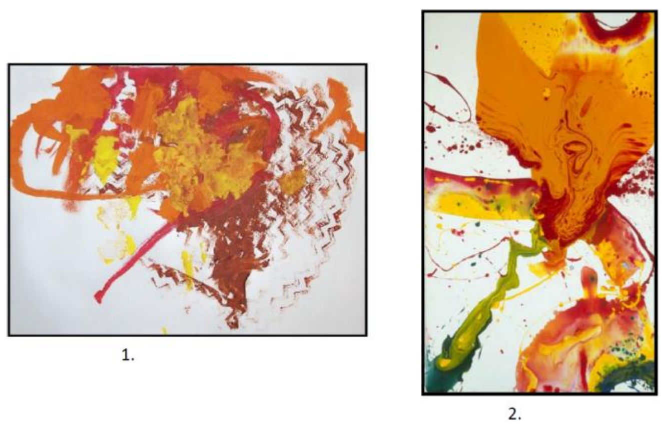
Participants

Forty-eight undergraduate psychology majors ranging in age from 17-20 years of age participated in this study as part of a course requirement. Participants were students majoring in psychology from a private university in the Northeast of the United States. Students received research credit for participation.   Participants wearing glasses were not included. Participants with contact lenses were included because these did not interfere with eye tracker calibration.

Materials and Procedure

Stimuli consisted of a random subset of 12 of the pairs of paintings used by Hawley-Dolan and Winner [2011]). Image pairs used are listed in Appendix 1. Images were presented on a laptop screen. Each screen showed a pair of images, one a painting by a well-known abstract expressionist, the other a superficially similar painting by a child (retrieved from preschool artwork online databases) or nonhuman (monkey, gorilla, ape, chimpanzee, retrieved from online databases of zoo galleries). As mentioned, artist and child/animal paintings were matched according to the following qualities: color, line quality, brushstroke, and medium. Paired images shared at least two of these qualities. Relative left right position of the artist and child/animal painting was randomized, with the same random order presented to all participants. Images were equated as much as possible in size and resolution and appeared with a small black border. The image on the left was numbered "1"; the image on the right was numbered "2". All signatures were removed using Photoshop. A sample image pair is shown in Figure 1.

Figure 1. Sample Image Pair with Child Painting on Left, Artist Painting on Right.



Note. Left: Preschool child painting by Brice Haedge, "Autumn Splendor", Kinderfrogs School @ TCU, Pre K United States. Reprinted by permission of Dana Haedge. Right: Artist painting by Sam Francis, Untitled, 1998. Reprinted by permission of the Artist Rights Society;© 2014 Sam Francis Foundation, California / Artists Rights Society (ARS), NY; © 2014 Artists Rights Society (ARS), New York.

Artist and child/animal paintings were equated for relative luminance, a normalized measure on a scale from 0 (for reference black) to 1 (for reference white). Relative luminance can be more important than absolute (physical) luminance in determining lightness perception (Li and Gilchrist, 1999). Mean relative luminance of an image was calculated as the mean of the red, green, and blue levels of the digital image pixels, normalized to a value between 0 and 1. A weighted variant of this luminance measure that accounts for differential sensitivity of the human visual system to red, green, and blue light was also considered (respective weights: $0.2, 0.7, 0.1$), but the results did not differ qualitatively from those obtained through use of the preceding measure. The mean relative luminance of the artist images (0.51) was not significantly different from that of the child/animal images (0.56), as shown by a *t*-test for paired samples, $p = 0.24, n = 12$. The difference in mean luminance between images presented on the left (0.54) and right (0.53) sides was also not significant, as shown by a *t*-test for paired samples, $p = .78, n = 12$. A Lilliefors test found no significant deviation from normality for the distributions of relative luminance values of the groups of images considered, $p > 0.5$ in all cases. Image contrast was measured using the standard deviation of the local relative luminance over the entire image, also known as Root Mean Square (RMS) contrast (Moulden and Gatley, 1989; Pelli, 1990), as well as by an additional measure of the fraction of an image for which large local variations in luminance occur. On neither of these contrast measures did the mean value of artist and child/animal images differ significantly; and on neither of these measures did the images presented on the left and right sides differ significantly, with *t*-tests yielding *p* values of .20 or greater.

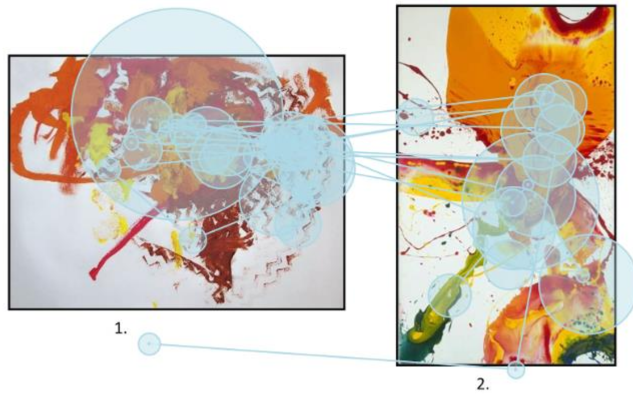
Participants were told that they were going to see 12 pairs of images and that each pair would appear twice. Each image pair was shown for a total of 20 seconds. Images were numbered either 1 or 2. At the first onset of an image pair, participants were asked: “For this image pair, which image do you like more?” Participants responded orally, stating 1 or 2, based on the image numbers, and the response was recorded in writing by the experimenter. The experimenter then asked “Why?” and again the response was given orally and recorded. After 20 seconds, the image pair was replaced by a blank gray screen, which was controlled by the experimenter and left on for as long as it took participants to answer and for the oral response to be recorded. The eye tracker data recorded for the blank gray screen were omitted from our analysis. Once the experimenter had finished recording the response, the experimenter proceeded to the next screen, in which the image pair was displayed for the second time, appearing again for 20 seconds, followed by another blank screen. As soon as the pair appeared for the second time, participants were asked: “Which image do you think is a better work of art?” followed by the question, “Why?” If participants said they liked them equally (Question 1) or they were both equally good (Question 2), the experimenter asked them which one they would choose if they had to pick one. The procedure took approximately 15 minutes. Because of technical difficulties, oral responses to the questions were lost for the first 23 participants tested. Thus, analyses examining oral responses are based on 25 participants. All of the other analyses are based on 48 participants.

Eye gaze was recorded with a SensoMotoric Instruments (SMI; Teltow, Germany) eye tracker. Participants were seated at eye level with the center of a 22 inch diagonal (56 cm) computer monitor at a distance of 24 inches (61 cm). The SMI system recorded participants’ left eye gaze at 120Hz via a SMI iView X RED mobile tracking system.

Each eye tracking session began with a 4-point calibration procedure using the calibration module within the iView X software. Participants were instructed on how to complete the pre-testing calibration, which required them to follow a fixation circle that moved to four quadrants on the display screen, maintaining their fixation on that circle until it moved to the next quadrant. During calibration, participants were instructed to look at the specified target points on the display while the system observed the corresponding pupil position, allowing the mapping of pupil position to on-screen gaze position. Mean recording accuracy for the group of 48 participants as reported subsequently by the calibration validation procedure was 0.55° and 0.58° for x and y positions, respectively. Preceding the onset of the eye gaze recording, participants were instructed to remain as still as possible and keep their gaze on the screen until testing was completed. Eye gaze recording then commenced. During recording, the experimenter watched the participant's head and body position to ensure consistency with pre-testing calibration. Participants were instructed to adjust slightly at times to ensure their eye gaze could be accurately recorded based on the pre-recording calibration.

SMI BeGaze software, version 3.0, was used for detection and initial processing of eye movement and pupil size data. Eye gaze data were extracted in the form of sequences of time stamped records, each describing the duration of an uninterrupted gaze fixation to either the artist image or the child/animal image, together with the mean pupil dilation (diameter) during this fixation. Visual fixations, saccades, and eye blinks were identified in accordance with standard techniques (Salvucci and Goldberg, 2000). Fixations were detected when eye gaze location presented a horizontal plus vertical dispersion of at most 100 pixels for at least 80 ms. Fixation duration values were recorded with a precision of 1 ms. Horizontal and vertical location values were averaged over the duration of the corresponding fixation event, and recorded with a precision of 0.1 pixels. A visualization of sample eye tracker data appears in Figure 2. Fixations centered outside the boundaries of the image stimuli (as seen in the Figure) were discarded before data analysis. Pupil size was measured by the number of black pixels detected in the eye image by the eye tracking system. Eye blinks were identified as fixation-like events shorter than 70 ms in which pupil size is less than 1 pixel or gaze location is 0 horizontally and vertically (as described in the SMI BeGaze 3.0 Manual), and were discarded before data analysis.

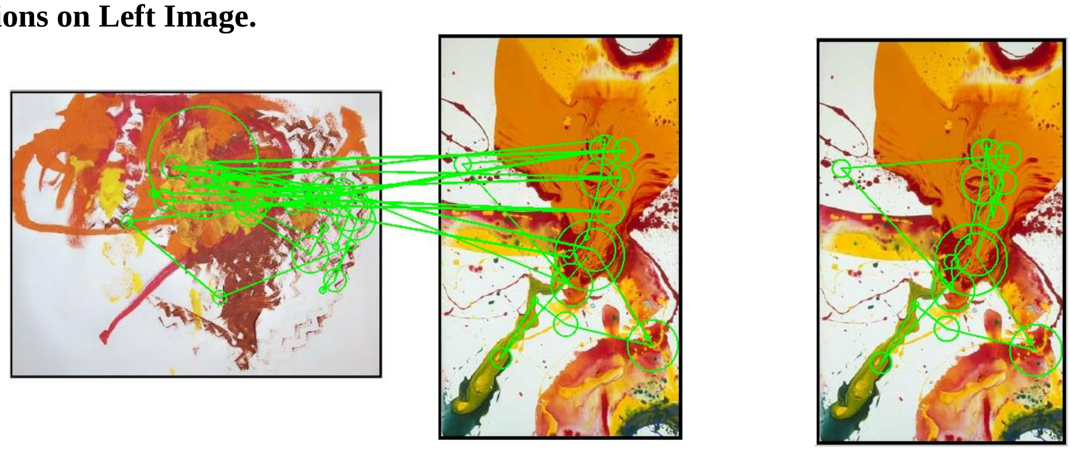
Figure 2. Raw Gaze Fixation to Image Pair from Figure 1 with Size of Circles Proportional to Fixation Duration



Fixation data processing and statistical analysis was carried out in MATLAB and the Statistics toolbox (The Mathworks, Natick, MA, USA). Preprocessing of the raw eye gaze data files was performed by scripts written in the Python programming language (www.python.org).

Variables of interest. Several variables derived from eye gaze fixation events were considered, including fixation duration, total number of fixations, summed fixation duration over all points of a given image (total fixation time), and mean pupil dilation per fixation, which is the result of dividing the sum of the individual pupil dilation samples during the fixation by the total number of those samples. Saccades, the eye movements that mediate between consecutive fixation events, often involve fixations on both images of a pair. A saccade-related measurement specific to each image was obtained by considering the angular distance between consecutive fixation locations within the given image only (see Figure 3). This angular distance is not always associated with an individual visual saccade between the given fixation locations, because intermediate saccades that cross from that image to the other and back may occur on the visual path between the start and end fixation locations, and any such cross-image saccades are ignored. Hence, the term ``fixation distance'' is used when referring to the angular distance between consecutive same-image fixations, instead of the more commonly used term ``saccade amplitude''. Two other variables were derived from the within-image fixation sequence: total path length (sum of distances between consecutive within-image fixation locations), and path dispersion (a measure of the spatial spread of the collection of fixation locations across an image, computed by summing the eigenvalues of the covariance matrix of the fixation locations). Like fixation distance, the latter two measures are rooted in saccade-related events. However, unlike fixation distance, total path length and path dispersion are mainly measures of the extent of spatial exploration of a given image, and are therefore measured more naturally in pixel units in the image itself. A bilateral (across images) variable was also considered: the number of crossings, that is, the total number of saccades that cross from one of the two images of a pair to the other.

Figure 3. Fixation Sequence After Preference Question with Fixations Outside Image Area Removed. a. Original Fixations on Image pair; b. Within-Image Fixation Sequence for Right Image, with Each Fixation Joined to the Next Fixation within the Image, Ignoring Fixations on Left Image.



a b

Note. Radius of each circle is proportional to the duration of the associated fixation. Consecutive fixations are joined by line segments.

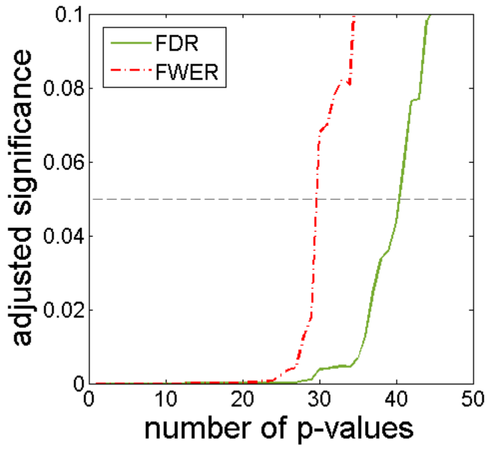
With the exception of the number of crossings between images, which is a bilateral measure, the values of selected variables were computed separately for the images in each pair, and separately for each question. Statistical hypothesis testing was carried out to detect distributional differences between the values of eye movement variables corresponding to fixations on the artist-authored image, and values of the same variables corresponding to fixations on the untrained-authored side. Data were pooled across the set of participants and image pairs. Because the groups being compared were matched precisely, with the same image pairs and human participants in all cases, statistical significance of differences in means (or medians) of times or dilations was measured by a paired test. In most cases, a Lilliefors / Kolmogorov-Smirnov test [Conover, 1999] rejected normality of the data at the level $p < 0.05$, and therefore a paired nonparametric Wilcoxon signed rank test was used instead of a *t* test unless otherwise stated.

For each within-image variable selected, X, an additional probability variable P(X) was introduced to assess directly, on a per-participant basis, whether greater values of the selected variable on one of the images in a pair were predictive of master authorship of that image (the number of gaze crossings between images is excepted, as it provides a single value for the pair of images rather than one value per image). First, for each human participant, h, the probability P(h,X) was computed that X takes greater values for participant h on the artist side than the untrained side, as the fraction of all image pairs I, for which participant h's eye movements when viewing I yield higher values of X on the master side than on the untrained side. The mean probability P(X) of higher values of X on the master side was then computed as the mean of the values P(h,X) over all participants, h. The probability P(X) is a measure of the reliability of X as a predictor of master authorship, on average, for a randomly selected human participant. Values of P(X) significantly greater than 0.5 for a particular variable, X, would indicate that the given variable is a better than random predictor of authorship. Statistical significance of such comparisons was assessed by a paired Wilcoxon signed rank test as described above. Because comparisons involving the probability variables P(X) are performed on a per-participant basis, sample sizes are smaller than for the comparisons of values of X directly. Hence, comparisons between the probability variables P(X) generally yield larger p-values than for the raw variables X.

Measurement of effect sizes. The present studies report results obtained under two closely matched experimental conditions. Paired statistical hypothesis tests are therefore used to assess significance as described above. As discussed by Morris and DeShon (2002, pp. 106-107), paired and unpaired tests correspond to different conceptualizations of the underlying populations, and this fact can inform the measurement of effect sizes. The *t* scores resulting from paired tests are based on the change scores (matched pairwise differences between groups); the *t* scores resulting from unpaired tests are based on the raw scores within groups. We thus used the standard deviation of the change scores (matched pairwise differences between groups) in the formula for effect sizes: $d = \frac{(mean_1 - mean_2)}{std}$. The resulting effect size measure can be calculated in terms of the *t* (or *z*) statistic resulting from a paired *t*-test on *n* pairs (or paired Wilcoxon signed rank test), as follows: $d = t\frac{\sqrt{2}}{\sqrt{n}}$, where the factor $\sqrt{2}$ ensures that the resulting values precisely match those of the pooled standard deviation version when the two groups are statistically independent.

Correction for multiple statistical comparisons. Statistical significance results were corrected using the Holm-Bonferroni procedure (Holm, 1979) to guarantee a familywise error rate (FWER) less than 0.05. Control of the false discovery rate at the level FDR < 0.05 was also considered, using the Benjamini-Hochberg (1995) procedure; also see (Nakagawa, 2004). Both the Holm-Bonferroni and the Benjamini-Hochberg procedure may be viewed as adjusting individual p-values upward before comparing them against the significance threshold. The adjusted p-values in the present paper are displayed in Figure 4.

Figure 4. Adjusted *p*-Values for Control of Family-Wise Error Rate (FWER) and False Discovery Rate (FDR) at $\alpha < 0.05$.



Results

Mean values and standard deviations for the variables of interest are shown in Table 1. Mean values of the variables of interest for the artist and child/animal images in response to the preference and quality questions separately appear in Table 2. Asterisks (respectively, underlines) denote statistical significance as assessed by a Benjamini-Hochberg correction to control the false discovery rate, FDR < 0.05 (respectively, by a Holm-Bonferroni correction to control the familywise error rate, FWER < 0.05). The mean values in Tables 1 and 2 were calculated by first averaging the values of a given variable of interest for each participant, over the pool of all 12 artist images, and separately over the pool of all 12 child/animal images. There were 19 instances (among a total of 1152 trials, corresponding to 2 questions and 12 image pairs for each of 48 participants), in which a participant fixated on only one of the two images when responding to a question. In such instances, no data was recorded on the unviewed side, and the corresponding values are missing from the computation behind Tables 1 and 2. A summary of significance findings appears in Table 3. Details for specific variables are discussed below.

Table 1. Means (Standard Deviations) of Variables of Interest.

Total fixation time (s)	7.91	(1.57)
Fixation count	27.4	(5.83)
Fixation duration mean (ms)	293	(71.3)
Fixation duration var /10^4	3.67	(5.22)
Total path length /10^3 (pixels)	2.26	(0.650)
Path dispersion /10^3 (pixels^2)	8.36	(3.85)
Fixation distance mean (degrees)	3.02	(0.597)
Fixation distance var (degrees^2)	4.19	(1.95)
Number of crossings	10.3	(4.70)
Pupil dilation mean (pixels)	14.9	(1.91)
Pupil dilation variance	0.729	(0.775)

Total fixation time. The mean total fixation time (averaged separately over the pool of all artist images and the pool of all child/animal images, for each participant) was significantly longer for the artist image (9.17 s) than the child/animal image (6.66 s) when responding to the quality question, (Wilcoxon signed rank test for paired samples, $p < 0.000001, z = 8.34, d = 0.49, n = 576$). This finding retains significance under a Holm-Bonferroni correction (FWER < 0.05) and therefore also under a Benjamini-Hochberg correction (FDR < 0.05). In response to the preference question, the difference between the mean total fixation time on the artist image (8.12 s) and the mean total fixation time on the child/animal image (7.68 s) was not significant (Wilcoxon $p = 0.074, z = 1.79, d = 0.11, n = 576$). See Table 2. The difference in total fixation times between the two questions for each image pair (averaged over the artist and child/animal images) was not significant, $p = 0.90, z = 0.119, d = 0.005, n = 1152$.

Table 2. Mean Values of Selected Variables in Response to Preference and Quality Questions

	Pref			Qual			Pref vs Qual
	Artist	Child/animal		Artist	Child/animal		
Total fixation time	8.12	7.68		9.17	6.66	*̲	
Fixation count	28.7	27.6		30.3	22.8	*̲	*̲
Fixation duration mean	288	282	*	309	296	*̲	*̲
Fixation duration var / 10^4	3.04	2.81	*	4.87	4.00	*̲	*̲
Total path length / 10^3	2.37	2.29		2.49	1.88	*̲	*̲
Path dispersion / 10^3	7.81	7.48		9.34	8.83	*	*̲
Fixation distance mean	2.96	3.03	*	2.99	3.07	*	
Fixation distance var	4.10	3.91		4.35	4.35		*̲
Number of crossings		11.8	-		8.75	-	*̲
Pupil dilation mean	14.8	14.7	*̲	15.2	15.0	*̲	*̲
Pupil dilation variance	0.799	0.803		0.662	0.655		*̲

Note. * Denotes Benjamini-Hochberg corrected statistical significance of differences, FDR < 0.05; underline denotes Holm-Bonferroni corrected significance, FWER < 0.05.

Because total fixation times were averaged for each participant over the pool of all artist images, and separately over the pool of all child/animal images, the mean total fixation times do not capture participants' implicit selections of one of the two images within each pair. In order to address this point, the probability of greater total fixation time on the artist side of a given image pair was computed for each participant as described earlier. In response to the preference question, the mean probability of fixating longer on the artist images was 0.51, at chance as indicated by a Wilcoxon signed rank test, $p = 0.56$, $z = 0.58, d = 0.12, n = 48$. For the quality question, the mean probability of fixating longer on the artist images was 0.65, which is significantly above chance, $p < .000001$, $z = 5.04$, $d = 1.0, n = 48$. The probability of fixating longer on the artist images was significantly greater for the quality than the preference question, Wilcoxon $p = .000049$, $z = 4.06, d = 0.83, n = 48$. These findings retain significance after a Holm-Bonferroni correction (Table 3).

Number of fixations. Results were qualitatively similar to those for total fixation time. For the preference question, no significant difference was observed between the mean numbers of individual fixations on the two images of each pair (28.7 vs. 27.6), Wilcoxon $p = 0.16, z = 1.40, d = 0.082, n = 576$, and the probability of a greater number of fixations on the artist side did not differ significantly from chance, $p = 0.92, z = 0.099, d = 0.020, n = 48$. For the quality question, the mean number of fixations was significantly greater on the artist image (30.3) than on the child/animal image (22.8), Wilcoxon $p < 0.000001, z = 7.82, d = 0.46, n = 576$, and the probability of a greater number of fixations on the artist image than the child/animal image was significantly greater than 1/2, Wilcoxon $p = 0.0000081, z = 4.46, d = 0.91, n = 48$. The mean number of fixations for each image in response to the preference question (28.2) was significantly greater than for the quality question (26.6), Wilcoxon $p = 0.000083, z = 3.94, d = 0.16, n = 1152$. Significance of all of these comparisons is retained after a Holm-Bonferroni FWER correction (Table 3).

Fixation duration. Significantly greater mean fixation duration was observed on the artist image than on the child/animal image in response to both questions, Wilcoxon $p = 0.0020, z = 3.09, d = 0.18, n = 576$ (means 288 ms and 282 ms) in response to the preference question, and $p = 0.0000046, z = 4.58, d = 0.27, n = 576$ (means 309 ms and 296 ms) in response to the quality question. Significance persists after a Benjamini-Hochberg correction for both questions. However, a Holm-Bonferroni correction eliminates significance in the case of the preference question. The per-participant, per-image pair probability of greater mean fixation on the artist image behaved similarly to total fixation time and number of fixations. The probability was significantly greater than 1/2 in response to the quality question, $p = 0.020, z = 2.33, d = 0.48, n = 48$ (mean probability 0.57), but not in response to the preference question, $p = 0.10, z = 1.64, d = 0.33, n = 48$ (mean probability 0.54). Significance is retained after a Benjamini-Hochberg correction but not a Holm-Bonferroni correction.

Like the mean of fixation duration, the variance of fixation duration was significantly greater on the artist image in response to both questions, with Wilcoxon $p < 0.000001, z = 5.04, d = 0.30, n = 576$ (mean $4.87\ 10^4$) for the quality question, and $p = 0.0066, z = 2.72, d = 0.16, n = 576$ (mean $3.04\ 10^4$) for the preference question. Again, only the result for the quality question retains significance after a Holm-Bonferroni correction (Table 3). The mean probability of greater fixation duration variance on the artist side (0.58) was significantly greater than 0.5 in response to the quality question, though significance does not survive a Holm-Bonferroni correction, Wilcoxon $p = 0.0033, z = 2.94, d = 0.60, n = 48$. In response to the preference question, the mean probability of greater fixation duration variance on the artist side (0.53) did not differ significantly from chance, Wilcoxon $p = 0.15, z = 1.44, d = 0.29, n = 48$.

Table 3. Significance of Differences between Artist and Child/Animal Images in Response to Preference and Quality Questions.

	Pref	Qual
Total fixation time		<u>**</u>
Fixation count		<u>**</u>
Fixation duration mean	*	<u>*</u>*
Fixation duration variance	*	<u>*</u>*
Total path length		<u>**</u>
Path dispersion		**
Fixation distance mean	**	*
Fixation distance variance		
Pupil dilation mean	<u>**</u>	<u>**</u>
Pupil dilation variance		

Note. * indicates Benjamini-Hochberg significance, FDR < 0.05 for data pooled across participants; ** indicates Benjamini-Hochberg significance, FDR < 0.05 for pooled data and for individual participants separately; underline indicates stronger Holm-Bonferroni significance, FWER < 0.05.

Fixation path length and dispersion. Fixation path length is significantly greater on the artist image ($2.49\ 10^3$) than the child/animal image ($1.88\ 10^3$) in response to the quality question, $p < 0.000001, z = 6.97, d = 0.41, n = 576$, but not the preference question ($2.37\ 10^3$ vs. $2.29\ 10^3$), $p = 0.34, z = 0.954, d = 0.056, n = 576$. Likewise, the mean probability of greater fixation path length on the artist side for a given image pair and participant in response to the quality question (0.61) is significantly greater than chance, Wilcoxon $p = 0.0000030, z = 4.67, d = 0.95, n = 48$, while the probability of greater fixation path length on the artist side in response to the preference question (0.51) does not differ significantly from chance, $p = 0.70, z = 0.390, d = 0.080, n = 48$.

Path dispersion was significantly greater on the artist image than the child/animal image in response to the quality question, Wilcoxon $\mathrm{p} = 0.0017, \mathrm{z} = 3.15, \mathrm{d} = 0.19, \mathrm{n} = 576$, but not in response to the preference question, Wilcoxon $\mathrm{p} = 0.28, \mathrm{z} = 1.07, \mathrm{d} = 0.063, \mathrm{n} = 576$. The mean probability of greater spatial dispersion on the artist image was also significantly greater than chance in response to the quality question, $\mathrm{p} = 0.018, \mathrm{z} = 2.36, \mathrm{d} = 0.48, \mathrm{n} = 48$, while the probability was not significantly different than chance in response to the preference question, $\mathrm{p} = 0.80, \mathrm{z} = 0.247, \mathrm{d} = 0.050, \mathrm{n} = 48$. Significance for the quality question is retained after a Benjamini-Hochberg correction, but not a Holm-Bonferroni correction. The mean spatial dispersion in response to the quality question was significantly greater than in response to the preference question, $p < 0.000001, z = 11.4, d = 0.47, n = 1152$. Significance of this difference is retained even after a Holm-Bonferroni correction.

Fixation distance. The mean fixation distance, that is, the mean distance between consecutive within-image fixations, is significantly *smaller* on the artist image than the child/animal image, for both the preference question (2.96° vs. 3.03°), $\mathrm{p} = 0.019, \mathrm{z} = 2.35, \mathrm{d} = 0.14, \mathrm{n} = 576$, and the quality question (2.99° vs. 3.07°), $\mathrm{p} = 0.010, \mathrm{z} = 2.56, \mathrm{d} = 0.15, \mathrm{n} = 576$. Neither of these findings survives a Holm-Bonferroni FWER correction, but both results persist after a Benjamini-Hochberg FDR correction. A significantly above chance probability of smaller fixation distance on the artist side (0.54) is found in response to the preference question, $\mathrm{p} = 0.016, \mathrm{z} = 2.41, \mathrm{d} = 0.49, \mathrm{n} = 48$, though not after a Holm-Bonferroni correction. Significance of the difference between the probability of smaller fixation distance on the artist side in response to the quality question (0.54) and chance does not survive a Benjamini-Hochberg correction, $\mathrm{p} = 0.030, \mathrm{z} = 2.16, \mathrm{d} = 0.44, \mathrm{n} = 48$. No significant difference is observed between the mean fixation distances for the preference and quality questions, $p = 0.17, z = 1.38, d = 0.057, n = 1152$.

No significant difference in the variance of fixation distance is observed between the artist image and the child/animal image, in response to either the preference question ($4.10\ vs.\ 3.91, p = 0.47, z = 0.719, d = 0.042, n = 576$) or the quality question ($4.35\ vs.\ 4.35, p = 0.39, z = 0.867, d = 0.051, n = 576$), and the probability of greater fixation distance variance on the artist image does not differ significantly from chance in response to either the preference question ($p = 0.20, z = 1.29, d = 0.26, n = 48$) or the quality question ($p = 0.45, z = 0.752, d = 0.153, n = 48$). A significantly greater variance of fixation distance is observed for the quality question (4.35) than the preference question (4.01) over the pool of all participants, even after a Holm-Bonferroni correction ($p = 0.000046, z = 4.08, d = 0.17, n = 1152$), but the probability of greater variance of fixation distance does not differ significantly between the two questions ($p = 0.17, z = 1.39, d = 0.28, n = 48$).

Number of crossings. The mean number of saccades that cross from one of the two images to the other was significantly greater in response to the preference question (11.8) than the quality question (8.75), a significant difference as determined by a Wilcoxon signed rank test for paired samples, $p < 0.000001, z = 13.5, d = 0.80, n = 576$. Significance persists after a Holm-Bonferroni FWER correction.

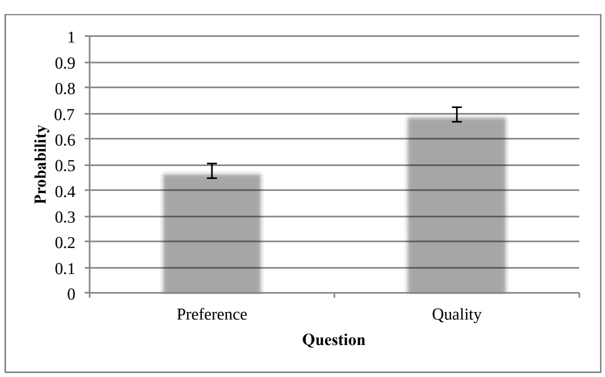
Pupil dilation. Mean pupil dilation was greater on the artist image than on the child/animal image in response to the preference question (14.8 pixels vs. 14.7 pixels, Wilcoxon $p < 0.000001, z = 5.89, d = 0.35, n = 576$), as well as in response to the quality question (15.2 pixels vs. 15.0 pixels, $p < 0.000001, z = 8.21, d = 0.48, n = 576$). Mean pupil dilation averaged over the two images of a pair was 14.8 pixels for preference vs 15.1 pixels for quality, a highly significant difference ($p < 0.000001, z = 14.5, d = 0.60, n = 1152$). This finding is consistent with prediction, based on the assumption that pupil dilation is an index of cognitive load. The mean probability of greater pupil dilation when looking at the artist images in response to the preference question was 0.60, which is greater than chance, Wilcoxon $p = .000011$, $z = 4.39, d = 0.90, n = 48$. For the quality question, the mean probability of greater pupil dilation when looking at the artist image was 0.64, again greater than chance, $p = .0000057$, $\mathrm{z} = 4.54, d = 0.93, n = 48$. The mean probability of greater pupil dilation when looking at the artist image did not differ significantly in response to the preference vs. quality questions (Wilcoxon , $p = .07$, $\mathrm{z} = 1.84, d = 0.38, n = 48$). These findings are consistent with prediction, based on the assumption that pupil dilation is an index of pleasure.

Significantly greater pupil dilation variance is observed in response to the preference question (0.80) than the quality question (0.66), Wilcoxon $p < 0.000001, z = 8.17, d = 0.34, n = 1152$, a result that persists after a Holm-Bonferroni correction. No significant difference occurs between artist images and child/animal images in response to the preference question ($p = 0.24, z = 1.16, d = 0.069, n = 576$), and observed borderline significance of the difference in response to the quality question ($p = 0.046, z = 2.0, d = 0.12, n = 576$) does not persist after a Benjamini-Hochberg correction.

Image choices. Although all participants indicated their choices in response to the preference and quality questions, the data for 23 of the 48 participants were lost due to hardware difficulties. Results involving the choices of the remaining 25 participants are reported below.

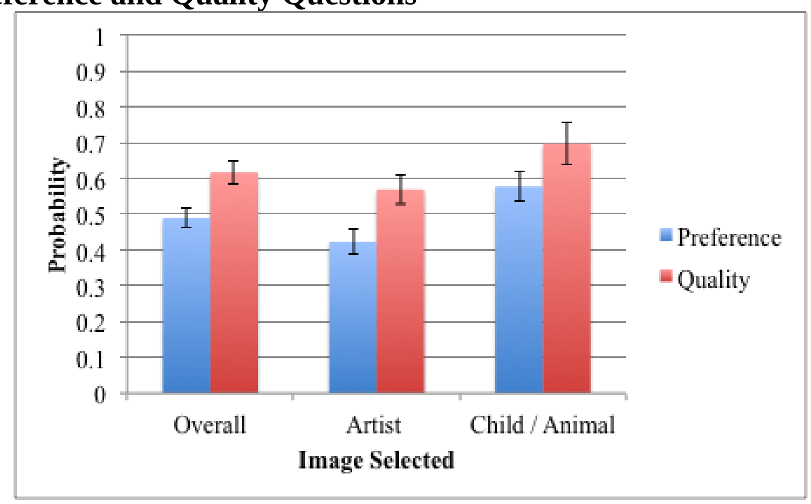
Explicit selections paralleled implicit selections based on total fixation time and other eye movement-based measurements. As shown in Figure 5, the mean per-participant probability of choosing the artist image in response to the preference question in Study 2 was 0.47, at chance as indicated by a Wilcoxon signed rank test, z = 1.03, $p = 0.30$, $d = 0.29, n = 25$. The mean per-participant, per-image pair probability of choosing the artist image in response to the quality question was 0.69, significantly above chance, Wilcoxon z = 4.2, $p <. 001$, $d = 1.2, n = 25$. The per-participant, per-image pair probability of selecting the artist images was significantly greater for the quality than the preference question, as shown by a Wilcoxon signed rank test, z = 4.16, $p < .001, d = 1.2, n = 25$. Justifications for choices were almost always descriptions of visual properties of the image chosen (e.g., colors, shapes, etc.). These justifications did not add anything informative to the findings and therefore were not analyzed.

Figure 5. Mean Probability of Selecting Artist Image in Response to the Preference and Quality Questions



Relation of choice to total fixation time. Surprisingly, for both preference and quality, the probability of fixating longer on the artist image was significantly higher when the child/animal image was selected than when the artist image was selected (Figure 6), with respective Wilcoxon $p = 0.0049, z = 2.812, d = 0.80, n = 25$ for the preference question, and $p = 0.026, z = 2.23, d = 0.63, n = 25$ for the quality question. Thus, fixating longer on an image is perhaps best referred to as an implicit indication of interest rather than as an implicit selection. However, significance is lost after a Benjamini-Hochberg correction in the case of the quality question, and after a Holm-Bonferroni correction in the case of the preference question.

Figure 6. Mean Probability of Fixating Longer on Artist Image Given Image Explicitly Selected for Preference and Quality Questions



Relation of choice to total number of fixations. The results for the number of fixations on each image are similar to those for total fixation time. Explicit selection of the artist image is associated with a higher probability of a greater number of fixations on the child/animal image than the artist image, for both questions. For the preference question, mean probability of a greater number of fixations on the artist side was 0.57 when the child/image was explicitly selected, and only 0.40 when the artist image was explicitly selected, a significant difference, Wilcoxon $p = 0.00068, z = 3.40, d = 0.96, n = 25$. For the quality question, the probabilities of a greater number of fixations on the artist side were 0.66 when the child/animal image was explicitly selected, and 0.54 when the artist image was explicitly selected, again a significant difference, Wilcoxon $p = 0.025, z = 2.24, d = 0.64, n = 25$. Significance is retained only for the preference question after a Benjamini-Hochberg but not a Holm-Bonferroni correction.

Relation of choice to mean duration of individual fixations. For the preference question, the probability of greater mean fixation duration on the artist image was significantly higher than chance when the child/animal image was selected (0.61, Wilcoxon $p = 0.018, z = 2.36, d = 0.67, n = 25$) but not when the artist image was selected (0.50, Wilcoxon $p = 0.94, z = 0.071, d = 0.020, n = 25$). Significance of the difference between the two was borderline, $p = 0.048, z = 1.97, d = 0.56, n = 25$, and does not survive a Benjamini-Hochberg FDR correction. For the quality question, the probability of greater mean fixation duration on the artist image does not differ significantly from chance, regardless of participants' explicit selection, and the difference in probability between the two selections is not significant, $p = 0.75, z = 0.325, d = 0.092, n = 25$. No significant findings occurred for the variance of fixation duration.

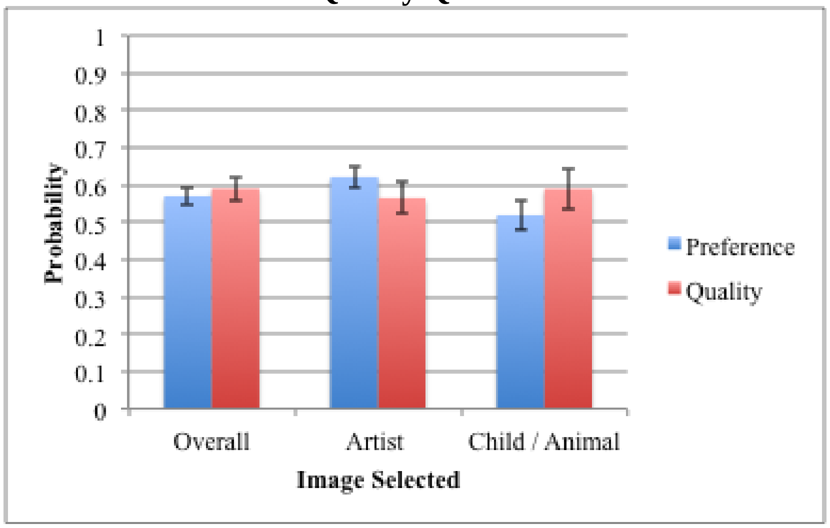
Relation of choice to total fixation path length. The results for path length (sum of distances between consecutive within-image fixations) are similar to those for total fixation time and number of fixations. A significantly higher probability of greater path length on the child/animal image than the artist image is observed when the child/animal image is selected (0.59) than when the artist image is selected (0.42) in response to the preference question, $p = 0.00050, z = 3.48, d = 0.98, n = 25$. In response to the quality question, the difference between the probability of greater path length on the artist image when the child/animal image is selected (0.66) and the probability of greater path length on the artist image when the artist image is selected (0.56) is not significant, $p = 0.10, z = 1.64, d = 0.46, n = 25$. Benjamini-Hochberg and Holm-Bonferroni corrections retain significance for the preference question.

No significant findings occurred for the remaining fixation-related spatial variables (path dispersion, mean and variance of fixation distance).

Relation of choice to pupil dilation. In response to the preference question, the mean probability of greater pupil dilation when looking at an artist image (compared to when looking at its paired child/animal image) was significantly greater than chance when the artist image was selected (Wilcoxon $p = .0012, z = 3.23, d = 0.91, n = 25$). Significance is retained after a Benjamini-Hochberg correction, but not a Holm-Bonferroni correction. The mean probability of greater pupil dilation when looking at the artist image did not differ significantly from chance when the child/animal image was selected (Wilcoxon $p = 0.52, z = 0.646, d = 0.18, n = 25$).

The mean probabilities differed significantly between the two explicit selections (p = 0.04, z = 2.06, $d = 0.58, n = 25$), though significance is lost after a Benjamini-Hochberg correction. In response to the quality question, the mean probability of greater pupil dilation when looking at the artist image did not differ significantly from chance, regardless of participants' explicit selections (Wilcoxon $p > 0.1, z < 1.62, d < 0.46, n = 25$), and the mean probabilities did not differ significantly between the two explicit selections (p = 0.51, z = 0.67, $d = 0.19, n = 25$), as shown in Figure 7. Two-way ANOVA and the nonparametric Friedman version of two-way ANOVA concur in identifying a significant group difference among the three pupil dilation means corresponding to the two explicit choices and the overall mean in the case of the preference question (p < .02), and no significant group differences among the means in the case of the value judgment question (p > 0.29). Significance of the group differences for the preference question is not retained after a Benjamini-Hochberg correction.

Figure 7. Mean Probability of Greater Pupil Dilation to Artist Image Given Image Explicitly Selected for Preference and Quality Questions



Discussion

Results show that, despite the naïve view that abstract art requires no skill and is something anyone could do, even observers with no special training in the visual arts can discriminate between works by famous abstract expressionist artists and superficially similar, equally bold, and often equally apparently messy works by the untrained – specifically children and animals who have been given painting materials by experimenters (monkeys, chimpanzees, gorillas, orangutangs, and elephants). The ability to discriminate these two classes of works has been reported when people are asked to guess whether a work is by an artist or a child or animal (Snapper, Oranc, Hawley-Dolan, Nissel, & Winner, 2014) and when people are shown a work by an artist paired with a work by a child or animal and asked which they prefer and which they deem to be a better work of art (Hawley-Dolan & Winner, 2011).

These findings become even stronger with the present study, in which we demonstrate using implicit measures obtained through eye tracking that untutored adults respond differentially to these two kinds of works – in terms of total fixation time, duration of individual fixations, number of distinct fixations, spatial extent of visual exploration, as well as in terms of pupil dilation. The results reported here demonstrate that people look significantly longer, and more carefully, at works by artists than at works by children and animals when making quality but not preference judgments. We suggest two reasons for this. First, quality judgments require a more exhaustive look before a reasoned decision can be reached; preference judgments can be spontaneous. Second, works by artists have more visual structure than works by children and animals, and thus invite more exploration. That lay adults do perceive more visual structure in the artist images than the child/animal images used here was shown by Snapper, Oranc, Hawley-Dolan, Nissel, & Winner, 2014. When deciding on preference, participants looked equally long at the two paired images, and their eye gaze explored paths of similar lengths in both images.

We recognize that because the preference question was always asked first, it is possible that longer total fixation times on the quality question were due to the fact that this was a second viewing of the image. However, we see no a priori reason why the opportunity to look a second time would lead to fixating longer. If anything, one might expect people to fixate less long on the second viewing since they have already seen the image.

The difference in fixation time in response to quality vs preference judgments supports theoretical claims that quality judgments and preferences require different ways of responding: making a quality judgment is a cognitive response based on an objective analysis of the work, while deciding upon a preference is a more automatic affective response response (Hagtvedt, Hadtvedt, & Patrick, 2008; Hawley-Dolan & Winner, 2011; Hawley-Dolan & Young, 2013), Leder, Belke, Oeberst, & Agustin, 2004; Zajonc, 1980). We also found that number of crossings between the two images was greater for the preference than the quality question. Perhaps this was because preference was harder to determine than quality. One can like an image even if one thinks it is not great art.

Pupil dilation was significantly greater when fixating on the artist images in response to both questions. This finding is consistent with Hawley-Dolan and Winner's (2011) finding that people report a stronger preference for artist over child/animal images.

Explicit selections of images in response to the quality question were consistent with Hawley-Dolan and Winner's (2011) finding that participants believe the artist images to be the better works of art. However, explicit selections of images in response to the preference question were inconsistent with results reported by Hawley-Dolan and Winner. The present study showed no reliable difference in preference for the artist over the child/animal image, while Hawley-Dolan and Winner showed a reliable preference for works by artists. The present study used only 12 of the 30 pairs that were used in the Hawley-Dolan and Winner study, and it is possible that this could account for the difference in findings.

When we compare explicit vs. implicit responses, however, we find consistency between these two kinds of measures. Explicit selections of images paralleled total fixation time findings. Participants looked equally long at both images in a pair when thinking about which they preferred, and they also showed no reliable difference in the frequency of selecting the artist vs. child/animal images as preferred. However, participants looked longer at the artist images when thinking about which they believed to be the better work of art, and they also selected the artist images over the child/animal images in response to the quality question.

Pupil dilation responses were also consistent with previous explicit response findings. Assuming that pupil dilation is a measure of pleasure, results showed that participants experienced more pleasure when fixating on the artist than the child/animal images. Participants also showed greater pupil dilation in response to the quality than the preference question. Assuming that pupil dilation is also a measure of cognitive load, results show that responding to the quality question requires greater mental effort than responding to the preference question. Our pupil dilation predictions could be accounted for either by pleasure or cognitive load, or both. Future research should attempt to disentangle pleasure from mental effort.

Surprisingly, fixating longer on an artist image did not predict choice of that image. In response to both the preference and quality question, the probability of fixating longer on the artist image was significantly higher when the child/animal image was selected (and the probability of looking longer at the child/animal image was significantly higher when the artist image was selected). We speculate that people may need to look longer at the image they are going to reject to achieve certainty than at the one they are going to choose. This phenomenon is reminiscent of that described in Griffin and Oppenheim (2006): people look longer at an object when naming it inaccurately than when naming it accurately.

Participants were significantly more likely to show greater pupil dilation while fixating on an artist image when that image was chosen as preferred, but not when that image was chosen as the better work of art. This finding suggests that when thinking about what is preferred, participants considered the pleasingness of the image, but while thinking about which one was better, pleasingness was less important as a factor. This finding provides further support for the claim that preferences and quality judgments are different kinds of aesthetic responses (Hawley-Dolan & Winner, 2011; Hawley-Dolan & Young, 2013).

It should be underscored however that these findings do not allow us to determine whether greater pupil dilation is indicative of greater pleasure, greater cognitive load, or both. One way of determining whether greater pupil dilation when looking at artist images is due to pleasure, and greater dilation when responding to the quality question is due to cognitive load would be to have participants rate each image for pleasure, and to rate each question for cognitive load. This remains for future research.

We conclude that people respond differently on both an explicit and implicit level when thinking about preference vs. quality in visual art. Results of this study support two previous studies (Hawley-Dolan & Winner, 2011; Snapper, Oranc, Hawley-Dolan, Nissel, & Winner, 2014) demonstrating that lay adults discriminate between abstract works by artists vs. those by children and animals. These two kinds of work are not indistinguishable, despite the fact that sometimes, given an individual work, even connoisseurs have been fooled.

It would be unsurprising to show that experts can distinguish abstract paintings by artists from those by children and animals. In this study we examined response to abstract art by people with no background or training in art or art history. These are the people who are likely to misunderstand abstract art as meaningless marks requiring no skill. Our results reveal that even though people unschooled in art consciously believe that abstract art is no different from the paint splotches of a child or monkey, they do in fact, at some non-explicit level, perceive a difference. Despite what one might think, it does not take any specialized training to distinguish works by artists from those by children and animals.

Acknowledgements

We thank David Bonaiuto, Jessica Kenworthy, Chris Wright, and Jacque DeAnda for their help in testing participants; we thank John Ksander and Eric Allard for indispensable technical assistance with the eye tracker; we thank Jenny Nissel for expert assistance with graphics. We thank the anonymous referees for their valuable comments, which have helped us to improve the paper. This research was supported by undergraduate research fellowships from Boston College.

References

Bar-Lev, A. (Director). (2007). *My kid could paint that* [Motion picture]. United States: A&E IndieFilms.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin, 91*(2), 276.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1), 289-300.

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

Chittenden, M. (2007, December 2). Tiny-fingers Freddie dupes art world. *The Sunday Times*. Retrieved November 10, 2009, from http://www.timesonline.co.uk/tol/news/uk/article2983819.ece

Cohen, L. B., and Cashon, C. H. (2003). Infant perception and cognition. In I. B. Weiner (Series Ed.) & R. M. Lerner, M. A. Easterbrooks, & J. Mistry, (Vol. Ed.),, *Handbook of Psychology: Vol. 6. Developmental Psychology,* pp. 65–89. New York: Wiley.

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: John Wiley & Sons.

Griffin, Z. M., & Oppenheimer, D. M. (2006). Speakers gaze at objects while preparing intentionally inaccurate labels for them. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 943.

Hagtvedt, H., Hagtvedt, R., & Patrick, V.M. (2008). The perception and evaluation of visual art. *Empirical Studies of the Arts, 26*, 197–218.

Hawley-Dolan, A., & Winner, E. (2011). Seeing the mind behind the art: We can distinguish abstract expressionist paintings from highly similar paintings by children, chimps, monkeys, and elephants. *Psychological Science, 22,* 4, 435-441.

Hawley-Dolan, A., & Young, L. (2013). Whose mind matters more: The moral agent or the artist? The role of intent in ethics and aesthetics. *PLOS ONE, 8*(9), e70759.

Henderson, J. M., & Hollingworth, A. (1998). *Eye movements during scene viewing: An overview*. In G. Underwood (Ed.), Eye Guidance in Reading and Scene Perception. Oxford: Elsevier.

Hess, E. H. (1965). Attitude and pupil size. *Scientific American*, *212*, 46-54.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*, 349-350.

Holm, S. (1979). "A simple sequentially rejective multiple test procedure". *Scandinavian Journal of Statistics* 6 (2), 65–70.

Hussain, F. (1965). Quelques problèmes d'esthétique experimentale. *Sciencias de l'Art*, *2*, 103–114.

Johnson, E. L., Miller Singley, A. T., Peckham, A. D., Johnson, S. L., Bunge, S. A. (2014). Task-evoked pupillometry provides a window into the development of short-term memory capacity. *Frontiers in Psychology, 13,* 5, 218.

Johnson, M. G., Muday, J. A., & Schirillo , J. A. (2010). When viewing variations in paintings by Mondrian, aesthetic preferences correlate with pupil size. *Psychology of Aesthetics, Creativity, and the Arts, 4*, 3, 161-167.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122-149.

Kapoula, Z., Daunys, G., Herbez, O., Yang, Q. (2009). Effect of title on eye-movement exploration of cubist paintings by Fernand Leger. *Perception, 38(4),* 479-491.

Kahneman, D. (1971). Changes in pupil size and visual discrimination during mental effort. In J.R. Pierce & J.R. Levene (Eds.), *Visual science: Proceedings of the 1968 International Symposium* . Bloomington , IN : Indiana University Press.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology, 95*, 489–508.

Li, X., & Gilchrist, A. L. (1999). Relative area and relative luminance combine to anchor surface lightness values. *Perception & Psychophysics*, *61*(5), 771-785.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*(1), 6-14.

Locher, P. J. (2006). The usefulness of eye movement recordings to subject an aesthetic episode with visual art to empirical scrutiny. *Psychology Science*, *48*, 2, 106-114.

Locher P., Krupinski E. A., Mello-Thoms C., Nodine C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, *21*, 55–77.

Molnar, E. (1981). About the role of visual exploration in aesthetics. In H. I. Day (Ed.), *Advances in intrinsic motivation and aesthetics* (pp. 385-413). New York: Plenum.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs: *Psych. Methods*,7, 1, 105-125.

Moulden, B., & Gatley, L. F. (1989). The standard deviation of luminance as a metric for contrast in random-dot images. *Perception*, *19*(1), 79-101.

Nissel, J., Hawley-Dolan, A., & Winner, E. (2014). "It's really good for an elephant": How children distinguish abstract expressionist paintings from superficially similar works by preschoolers and animals. Submitted for publication.

Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology*, 15(6), 1044-1045.

Nodine C., Krupinski E. (2003). How do viewers look at artworks? *Bull. Psych. Arts 4*, 65–68.

Nodine, C., & McGinnis, J. (1983). Artistic style, compositional design, and visual scanning. *Visual Arts Research, 12*, 1-9.

Pelli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America. A, Optics and image science, 7*(10), 2032-2040.

Plumhoff, J.E., & Schirillo, J. A. (2009). Mondrian, eye movements, and the oblique effect. *Perception, 38*, 719-731.

Salvucci, D. D. & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (pp. 71-78). Association for Computing Machinery, New York, NY, USA. ISBN:1-58113-280-8*Snapper, L., Oranc, C., Hawley-Dolan, H., Nissel, J., & Winner, E. (2014). Your kid could not have done that: Even untutored observers can discern intentionality and structure in abstract expressionist art. Under review.

Taylor, R. P., Spehar, B., Van Donkelaar, P., & Hagerhall, C. M. (2011). Perceptual and physiological responses to Jackson Pollock's fractals. *Frontiers Human Neurosci 5*: 60.

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

Zajonc, R. B. (1980). Feeling and thinking: preferences need no inferences. *American Psychologist*, 35(2), Feb 1980, 151-175.

Appendix 1. List of Image Pairs

Pair 1

Artist: Kenzo Okada "Points No. 19"

Animal (elephant): Yod Yeam Medium

Pair 2

Artist: Karel Appel, "Untitled 196"

Animal (chimpanzee): Congo

http://www.theage.com.au/ffximage/2005/06/21/chimp_congo_wideweb__430x328.jpg

Pair 3

Artist: Helen Frankenthaler "The Caves"

Animal (orangutan): Nonja

Pair 4

Artist: Sam Francis "Untitled" 1998-89

Child: Brice "Autumn Splendor" Kinderfrogs School @ TCU, Pre K United States

Pair 5

Artist: Clyfford Still "1945-R"

Animal (elephant): Kam Lai Tong

Pair 6

Artist: Mark Tobey "New World Stage" early 1960s

Animal (elephant): Punpetch, Elephantart.com

Pair 7

Artist: Elaine de Kooning "On the way to San Remo" 1967

Animal (elephant): Lucky

Pair 8

Artist: Mark Rothko “Multiform” series, 1946-1949

Child: Khayman P. “Green Fingers”

Pair 9

Artist: Ralph Rosenborg, “Autumn Landscape”

Child: Ronan “Fall Colors” La Verne Parent Participation Preschool”

Pair 10

Artist: Boon' oil on canvas painting by James Brooks, 1957, Tate Gallery

Animal (gorilla): Okie

http://gregcookland.com/journal/2007/08/gorilla-art-at-franklin-park-zoo.html

Pair 11

Artist: Joan Mitchell, http://abstractexpressionism.wordpress.com/2007/10/06/joan-mitchell/

Animal (elephant): Naram #66 Medium

Pair 12

Artist: Hans Hoffman Fiat Lux 1963

Child: Phillip C. “Colors” Age 4, South Carolina