Speed of person perception affects immediate and ongoing aesthetic evaluation

Mel W. Khaw\textsuperscript{a,b}\textsuperscript{*}, Phoebe Nichols\textsuperscript{b}, David Freedberg\textsuperscript{c}

\textsuperscript{a} Center for Cognitive Neuroscience & Duke Institute for Brain Sciences, Duke University, USA
\textsuperscript{b} Department of Neuroscience & Visual Arts Department, Bowdoin College, USA
\textsuperscript{c} Department of Art History and Archaeology & Italian Academy for Advanced Studies, Columbia University, USA

\textbf{ARTICLE INFO}

Keywords:
Aesthetics
Empirical aesthetics
Neuroaesthetics
Person perception
Face perception
Attention

\textbf{ABSTRACT}

Recent studies have shed light on how aesthetic judgments are formed following presentations lasting less than a second. Meanwhile, dedicated neural mechanisms are understood to enable the rapid detection of human faces, bodies, and actions. On the basis of cognitive studies of: (i) the speed and acuity of person perception, and (ii) preferential attention given to human imagery (e.g., faces and bodies), we hypothesize that the visual detection of humans in portraits increases the magnitude and stability (i.e., similarity to later responses) of aesthetic ratings. Ease of person perception is also expected to elicit longer durations of preferential viewing time, a surplus measure of viewing behavior that should be positively related to subsequent ratings. To test these ideas, we use a set of cubist portraits previously established to be more or less categorizable in terms of the aggregate time required to perceive the depicted person. Using these images, we track aesthetic judgments made following short and unconstrained presentations; in an intervening task, we measure viewing behavior when subjects are able to selectively reveal regions of these images. We find that highly categorizable artworks (those that require less time to identify the figure as human) elicit higher and more predictive aesthetic ratings following 30 ms presentations while also eliciting longer viewing durations. Changes in ratings throughout the task are positively correlated with cumulative viewing time; critically, an image’s categorizability level further moderates the strength of this relationship. These results demonstrate that a particular kind of visual object recognition – the recognition of human forms – modulates aesthetic preferences at a glance, subsequent viewing patterns, as well as rating changes over time.

1. Introduction

Recent results in empirical aesthetics highlight the brief time window from which meaningful aesthetic judgments are declarable. For example, judgments of beauty following glimpses of stimuli (e.g., following presentations lasting 30–500 ms) are positively correlated with analogous judgments following unrestricted viewings (Verhavert, Wagemans, & Augustin, 2018). This type of investigation has also been performed with more dynamic stimuli; accurate judgments of liking can be made after listening to brief passages of music – durations as brief as 500 ms suffice for select genres of music (Beli, Kasdan, Rowland, Vessel, & Starr, 2018). Studies utilizing brief time windows have also focused on specific visual attributes relevant to aesthetic processing. For instance, content-wise similarity in representational artworks can be discriminated upon following 10 ms views, while similarity in style requires durations upward of 50 ms to disambiguate (Augustin, Leder, Hutzler, & Carbon, 2008). Earlier studies of the timecourse of aesthetic judgments focused on the collative properties of artworks and their relation to arousal levels and hedonic experiences (Rautmann & Vipper, 1983; Berlyne, 1960). In one such study, verbal judgments of ‘complexity’ and ‘orderliness’ are obtainable after 50 ms glances at pattern (synthetic) and painting (genuine) artistic images (Cupchik & Berlyne, 1979). The present study also focuses on aesthetic evaluations at short timescales. In addition, we examine a specific form of visual object recognition and its relevance toward quick aesthetic evaluations: the visual detection of people. This form of visual perception, along with downstream social impressions (of gender, race, personality, etc.), is studied in the wider cognitive sciences as person, people, or social perception (Phillips, Weisbuch, & Ambady, 2014; Rutherford & Kuhlmeier, 2013). Our investigation looks at ratings and viewing behavior elicited by a set of cubist portraits previously studied by Hekkert and van Wieringen (1990). Within this set of portraits, Hekkert and van...
Wieringen (1990) identified varying levels of image categorizability – this metric was based on the amount of time subjects required to identify the depicted human figure within each portrait. Based on physiological and behavioral findings surrounding person perception, we predict that the speed of forming such percepts – as reflected by known categorizability levels – positively affects the magnitude and stability of aesthetic ratings over time. In what follows, we briefly review: (i) relevant timing benchmarks during the extraction of visual information; (ii) social judgments and attention modulation using human imagery; and, (iii) links between person perception and other theoretical constructs in aesthetic processing.

The general timescale of visual information extraction contains several relevant benchmarks for aesthetic processing. In agreement with rapid-content-based judgments of artworks (Augustin et al., 2008), categorical detection of faces or animals in natural scenes can be performed well within 250 ms (Rousselet, Macé, & Fabre-Thorpe, 2003). The visual accumulation of such information resembles a continuous process – using a series of precisely-timed masks, behavioral accuracy associated with brain activity both plateau following 40–60 ms of pre-mask exposure time (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005). Indeed, semantic details regarding people and objects are reportable from glances at complex natural scenes following views as short as 27 ms (Fei-Fei, Iyer, Koch, & Perona, 2007). Despite the quick flow of coarse object information (Riesenhuber & Poggio, 2000), more nuanced behavioral judgments nonetheless vary depending on processing type and the amount of processing that is allowed to take place. For example, color similarity judgments are driven primarily by retinal color information after 200 ms exposure and declines with further exposure time (Schulz & Sanocki, 2003). Select forms of fundamental perceptual grouping are also more or less difficult to perform. The grouping of small pattern elements and individuation of large ones appear to be intrinsic and immediate, while the opposite ability develops between ages 5–10 (Kimchi, Hadad, Behrmann, & Palmer, 2005). Crucially, the integration of configural elements into impressions of people is relatively quick and arguably innate. Sufficient representations of people include 8–12 joints of a moving figure at 150 ms of viewing time (Johansson, 1973), while specific face recognition takes place within 100–200 ms of exposure (Jacques & Rossion, 2006; Sugase, Yamane, Ueno, & Kawano, 1999). In addition, geometric shapes in face-like arrangements are capable of eliciting preferential gaze from newborn infants (Morton & Johnson, 1991).

The second-speed speed of recognizing faces and making social inferences (Johnson, 2005) are putatively enabled by sub-cortical nuclei such as the amygdala (Vuilleumier, Armony, Driver, & Dolan, 2001), in tandem with a cortical network centered at the fusiform gyrus (Kanwisher, McDermott, & Chun, 1997; Kanwisher & Yovel, 2006). Visual recognition of human portrayals might also involve brain regions selectively dedicated to action recognition and observation (Cross, Hamilton, Kraemer, Kelley, & Grafton, 2009; Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). While the subliminal influence of faces on abstract art has been explored (Fleax et al., 2013), the influence of person perception per se, as well its effects on aesthetic evaluations over time, are unknown. Seeing as minimal visual detail and exposure time are required for the detection of faces and bodies, time-constrained assessments of artworks should be particularly impacted by the emergence of such percepts. In particular, the relative ease of detecting portrait subjects should strengthen potential correlations between aesthetic first impressions and evaluations made following extensive viewing.

Given the perceptual benchmarks above, the ability to produce spontaneous high-level judgments toward people might appear less surprising. Locher, Unger, Sociedade, and Wahl (1993) show that stereotypic factors of attractiveness (e.g., masculinity, femininity) exert their influence on judgments following 100 ms-long presentations. Bar, Neta, and Linz (2006) demonstrate that stable personality impressions are declarable within the first 40 ms of viewing faces. Furthermore, inferences about a person’s attractiveness and trustworthiness (Willis & Todorov, 2006), as well as (above-chance) classifications of sexual orientation (Rule & Ambady, 2008; Rule, Ambady, & Hallett, 2009) can be elicited following presentation windows of 100 ms and below. Critically, for the study here, images of faces and people are also known to elicit preferential judgments, decisions, and attention. On this issue, we restrict the focus of our study to person perception independent of varying levels of personal beauty or attractiveness. At this basic level, perceptual realizations of faces embedded in abstract patterns have been shown to coincide with increases in aesthetic appreciation (Muth & Carbon, 2015). Furthermore, experiments in economic decision-making demonstrate that people readily trade work effort and money for opportunities to look at faces (Hayden, Parikh, Deearer, & Platt, 2007). The inclination to perform such tradeoffs has since also been demonstrated in monkeys, where viewing their conspecifics is “paid for” using consumables (Deearer, Khera, & Platt, 2005). In these decision-making experiments, such tradeoffs are taken to suggest that images of faces are associated with valuable information; such visual stimuli might provide information about social hierarchies, membership, and attributes relevant to social interactions (Klein, Shepherd, & Platt, 2009). Indeed, as highlighted previously, the perception of humans is often accompanied by secondary inferences, such as immediate judgments of personality traits from faces (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015) and emotional affect from postures (de Gelder, De Borst, & Watson, 2015). Using minimalistic figures of human motion, it has also been shown that socially-interacting figures are more likely to be consciously perceived compared to isolated or scrambled figures, when these “options” are presented to individual eyes (Su, van Boxtel, & Lu, 2016); the same type of human figures in motion triggers reflexive attentional orienting when used as task-irrelevant distractors (Shi, Weng, He, & Jiang, 2010). Person perception can thus be expected to influence aesthetic evaluation by facilitating the visual transfer of social information – information potentially impacting aesthetic appraisal (this link is discussed in further detail below). In social neuroscience, humanized perception – the recognition of others’ mental lives – has recently been identified as a promoter of pro-social behavior, along with familiarity and similarity to the sense of self (see Harris and Fiske (2009) for a review). From this perspective, person perception in artworks might engender thoughts about those depicted, and their relevance to the observer. Indeed, the assessment of self-relevancy has been postulated to be a central component of aesthetic processing (Vessel, Starr, & Rubin, 2013). Given what might be a common visual preference for humanized content, we furthermore hypothesize that aesthetic ratings (and subsequent viewing time) will favor images with easily-perceived human subjects, even following brief initial glances. In relation to existing theoretical frameworks in empirical aesthetics, the role of person perception/recognition ties into two major topics: (i) processing fluency and familiarity, as well as (ii) the elicitation of empathetic responses. “Fluent” stimuli – those that are more easily processed cognitively – are theorized to obtain greater degrees of aesthetic appreciation; theoretical mechanisms include positive affect arising from said processing (Reber, Schwarz, & Winkielman, 2004), as well as deviations from expectations (regarding fluency) resulting in pleasure and/or interest (Graf & Landwehr, 2015). At present, data-driven measures have been used to identify general determinants (e.g., visual contrast, symmetry, etc.) of fluency (Mayer & Landwehr, 2018). In relation to our study, fluency is also driven by factors that aid in the categorization of stimuli, such as the prototypicality of shapes (Winkielman, Halberstadt, Fazendeiro, & Catty, 2006), prior exposure to stimuli (Bornstein & D’Agostino, 1994), and matching perceptual primes (Reber, Winkielman, & Schwarz, 1998). In this respect, and as the perceptual literature above makes clear, the detectability of people within portraits is predicted to increase an image’s level of conceptual or semantic fluency (Reber et al., 2004; Whittlesea, 1993). In our study, the presence of people within artworks is expected to imbue glances with readily interpretable information; faces and bodies likely facilitate
social inferences more easily compared to other perceptual groupings such as shapes, textures, and inanimate objects. Indeed, using a similar set of cubist paintings, the explicit recognition of depictions resulted in pupil dilation, with preferences given toward easily construed landscapes and portraits (Kuchinke, Trapp, Jacobs, & Leder, 2009). Furthermore, individuals possess an inclination to actively associate geometric shapes with person-like characteristics – a bias in favor of perceiving animacy, originally reported by Heider and Simmel (1944). It should be noted, however, that the present study is by no means intended to reduce the merits of abstract, inanimate, or otherwise challenging art; indeed, the process that leads to the developed liking of such art has recently been documented empirically (Belke, Leder, & Carbon, 2015). Nonetheless, we intuit that the clarity with which people are depicted likely aids the short-term assignment of aesthetic values, during which relative familiarity might encourage valuation (Alter & Oppenheimer, 2008). Familiarity (having prior experience with a particular stimulus) is itself a likely promoter of aesthetic engagement in our task. This general prediction stems from the ‘mere exposure’ effect (Zajonc, 1968) – where prior experience increases aesthetic liking (Martindale, Moore, & West, 1988), possibly through the selective priming of different image attributes (Martindale & Moore, 1988). Finally, human portrayals also tie in with theories that emphasize the role of embodied imagery and empathetic responses in art (e.g. Crozier & Greenhalgh, 1992; Freedberg & Gallese, 2007; Lips, 1935), though once again this does not exclude the possibility of projecting human and embodied forms onto abstract shapes. To illustrate, the observation of actions in art elicits motor responses that have been interpreted as empathetic reactions (e.g., Battaglia, Lisanby, & Freedberg, 2011). Such reflexive motoric reactions also arise from viewing abstract visual information, such as textures and brush strokes in artworks (Freedberg & Gallese, 2007; Sbriscia-Fioretti, Berchio, Freedberg, Gallese, & Umlita, 2013). If the basis of empathetic responses lie in these patterns of matching neural responses (to some inferred activity or action), then human figures should represent a direct source of such action-relevant information. The neural intersection between embodied responses and aesthetic appreciation has also been reviewed extensively by Ticini, Urgesi, and Calvo-Merino (2014) as well as by Kirsch, Urgesi, and Cross (2016). Overall, the relevance of human portrayal in theories are wide and varied, though the general prediction would be that such content plays a significant positive role in aesthetic engagement and the assignment of aesthetic value.

In the present study, we examine the assignment of aesthetic value on a set of known cubist portraits featuring differing levels of categorizability. These artworks were originally used by Hekkert and van Wieringen (1990), where the authors defined 3 levels of categorizability (low, intermediate, and high) based on the time it took for their participants to identify the human figure in each painting. We hypothesize that these pre-defined levels affect aesthetic evaluations following short presentation durations as well as conventional, unconstrained presentations. These differences are expected to manifest in terms of rating magnitudes, correlations between measures, as well as rating changes as a function of viewing time. In addition to ratings, we also predict that preferential viewing time – time freely dedicated toward images beyond initial glances – will be allocated favorably toward highly categorizable images. Such a prediction would be consistent with the desirability of discovering faces (Muth & Carbon, 2013), and viewing socially-relevant imagery (Hayden et al., 2007). Notably, this prediction lies in opposition to the cost of time incurred by effortless processing of artworks (e.g., Kuchinke et al., 2009; Reber et al., 2004), whereby longer processing times are found to be negatively related to preference ratings. To test these predictions, we employ a study with 3-phases. First, we employ a short presentation ratings task akin to studies of the micro-genesis of aesthetic preferences (Augustin et al., 2008; Verhavert et al., 2018). Second, we use a mouse-tracking task where subjects selectively reveal portions of images (using their mouse cursor) from different categories. This measure was used in previous studies to gauge the processing weight, as indicated by viewing time, put on specific information (e.g., time and money amounts) when economic choices are made (Khaw, Li, & Woodford, 2018; Payne, 1976; Reek, Wall, & Johnson, 2017). In this study, mouse-tracking allows us to track viewing time preferentially distributed toward specific images, while enabling us to subsequently visualize particular areas of interest within these images. Finally, we utilize a conventional ratings task at the end of the experiment where presentation durations were not limited.

2. Method

2.1. Observers

20 adults (11 female) participated in all phases of the main study consecutively. The average age of this sample was 22.53 (SD = 3.36). Participants were recruited from the Columbia University community using advertisements placed throughout the university’s main campus. Participants completed each task at an individual computer station in about 45 min. During post-task interviews, none of these 20 participants reported having an a priori familiarity with the specific images used, though several noted a familiarity with the cubist style more generally. All participants provided informed consent and were compensated with $15 for their participation.

Following a reviewer’s suggestion, we recruited a further 17 adults (9 female) from the Duke University community for a follow-up experiment to validate the categorizability levels assigned by Hekkert and van Wieringen (1990). The average age of this sample was 23.63 (SD = 3.46). These participants also provided informed consent before participating, and were compensated with $8 for their participation. None of these additional participants reported familiarity with the images that were presented to them during the task.

2.2. Stimuli

The stimuli from this experiment were cubist paintings originally selected by Hekkert and van Wieringen (1990) in their study of prototypicality. We replaced one painting that was ambiguously titled in the High Categorizability category, with the full list of works appearing in Table 1. Based on the average time it took for their subjects to find the human figure within each painting, Hekkert and van Wieringen (1990) divided the paintings into three categories: Low Categorizability (LC), Intermediate Categorizability (IC), and High Categorizability (HC). In that study, these images were also rated on ‘beauty’, ‘prototypicality,’ and ‘complexity’. The authors concluded that aesthetic preferences for abstract works are more influenced by complexity, while seemingly representational works are more influenced by prototypicality.

In order to preserve subjects’ lack of familiarity with these images (in accordance with other studies of initial, first impressions-based, aesthetic ratings), we did not conduct a confirmatory pre-test of image categorizability. Furthermore, a post-task assessment of categorizability was also not possible with these subjects as they would have learned the relevant spatial locations within various images. Thus, we employ the categorizability ranges defined by Hekkert and van Wieringen (1990), which we later validate with an independent sample of subjects (see below). In their report, LC images required mean response times above 7000 ms for the identification of the human figure. Similarly, IC images necessitated average response times between 3000 and 7000 ms; lastly, HC images were those that required less than an average of 3000 ms for figure identification.

In order to control for potential differences in low-level visual properties between categories, all images underwent a set of pre-processing procedures. We equated images’ histograms in terms of their luminance, contrast, and average amplitude spectra using the SHINE Toolbox (Willenbockel et al., 2010). The average response times using this set of normalized images, relative to the original set, as well as the
relative rankings of image discriminability are plotted in Fig. 2 and detailed in the Results section (Categorizability Speeds).

2.3. Procedure

2.3.1. Categorizability task

Following a reviewer’s suggestion, we sought to replicate the robustness of categories defined by Hekkert and van Wieringen (1990) with the digitally normalized images that were utilized in the main study. Toward this end, we employed an analogous recognition task using an independent sample of participants. Subjects on this task were instructed to press the spacebar key (with their finger constantly positioned on the keyboard) as soon as they perceived the depicted portrait subject. The onset of images was cued by a 3-second countdown which was always triggered by the subject (to prevent presentations before the subject was ready). To further encourage subjects to produce meaningful responses, subjects were instructed to indicate with the mouse cursor where they perceived the portrait subject immediately following their response.

2.3.2. Rating task

In the first rating phase, subjects were presented with members of each stimulus category in randomized order. Subjects were instructed to rate on a continuous slider bar how pleasing they perceived the on-screen stimuli, ranging from 0 (not pleasing) to 100 (most pleasing). In order to discourage rote memorization of prior ratings, ratings were displayed with a precision of two decimal places; in addition, subjects were not limited in the time available to declare their ratings. In the first version of the rating phase, presentation durations were constrained to 30 ms. This specific duration was chosen as it meets experimental benchmarks of aesthetic, social, and visual judgments. Following the findings of Verhavert et al. (2018), reliable judgments of beauty (in terms of significant correlations with unconstrained ratings) are obtainable from 30 ms onward, with the average correlation utilizing 40 ms appearing to be only marginally greater. Second, as described in more detail within the Introduction section, high-level social judgments of face stimuli such as of emotion (Morris, Öhman, & Dolan, 1998), sexual orientation (Rule et al., 2009; Rule & Ambady, 2008), and personality (Carney, Colvin, & Hall, 2007) are obtainable within the 30–50 ms range. The 30 ms constraint thus represents a duration that allows us to replicate prior observations regarding aesthetic judgments, while being in a range that allows for the visual extraction of person-related semantic associations. Subjects subsequently completed a viewing procedure (Spotlight task, described below). In the following final rating phase, images were left on the screen for as long as subjects deliberated on their judgments. In both rating phases, each image was presented 3 times for a total of 120 trials, comprising 40 unique images (Table 1).

2.3.3. Spotlight task

During the spotlight task, participants were presented with stimuli only viewable through a small circular aperture. The position of the aperture is controlled by mouse movement in order to uncover specific portions of the stimuli. The aperture was a Gaussian transparency filter featuring a standard deviation of 33 pixels, centered on a square mask with a height and width of 200 pixels (Scarf, 2015). Subjects were able to reveal any on-screen region of their choice through this “spotlight” for 5 s during each trial. Before subjects were allowed to maneuver the mouse, stimuli were presented briefly for 30 ms in their respective locations (one image at a time, in random order). All subjects completed 48 trials, comprising 3 repetitions of each complete set of images. Images in the IC and LC categories received 4 additional repetitions each, as their categories each included 4 fewer images than the HC set. These additionally repeated images were randomly selected from each category so that 16 triplets were presented in each repetition. The spatial location of each category’s images were randomized during each trial (one image from each category occupied either the middle, right, or left on-screen position; Fig. 1B). During the entirety of the 5 s provided for each trial, the momentary locations of participant’s mouse’s movements (in the form of (x,y) screen coordinates) were recorded. A circular timer above the screen indicated the portion of 5 s remaining on each trial.

3. Results

3.1. Categorizability speeds

Based on a reviewer’s suggestion, we ran a follow-up study to validate the categorical boundaries proposed by Hekkert and van Wieringen (1990). We first confirm that the average times for subjects to declare recognition (of a person in each portrait) are not significantly different between the previous and current study (Fig. 2A). Here we perform paired two-tailed t-tests between the previously reported

---

Table 1

<table>
<thead>
<tr>
<th>Category</th>
<th>Artist</th>
<th>Painting</th>
</tr>
</thead>
<tbody>
<tr>
<td>High categorizability</td>
<td>Braque</td>
<td>Standing Female Nude (1907)</td>
</tr>
<tr>
<td></td>
<td>Cezanne</td>
<td>Cubist figure (1913)</td>
</tr>
<tr>
<td></td>
<td>Gleizes</td>
<td>Portrait of Jacques Naylor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1910 - 11)</td>
</tr>
<tr>
<td></td>
<td>Gleizes</td>
<td>Man on a balcony (1912)</td>
</tr>
<tr>
<td></td>
<td>Gris</td>
<td>Portrait of Picasso (1912)</td>
</tr>
<tr>
<td></td>
<td>Gris</td>
<td>Portrait of Josette (1916)</td>
</tr>
<tr>
<td></td>
<td>Gris</td>
<td>Harlequin with guitar (1917)</td>
</tr>
<tr>
<td></td>
<td>Léger</td>
<td>Harlequin seated beside a table (1919)</td>
</tr>
<tr>
<td></td>
<td>Larionov</td>
<td>Woman walking on the boulevard (1912)</td>
</tr>
<tr>
<td></td>
<td>Léger</td>
<td>Woman sewing (1909)</td>
</tr>
<tr>
<td></td>
<td>Metzinger</td>
<td>Lady at her dressing table (1916)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Clovis Sagot (1909)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Nude woman in an armchair (1909)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Wilhem Uhde (1910)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Seated female nude (1910)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Harlequin (1918)</td>
</tr>
<tr>
<td>Intermediate categorizability</td>
<td>Braque</td>
<td>Woman playing a mandolin (1910)</td>
</tr>
<tr>
<td></td>
<td>Braque</td>
<td>Female figure (1910 - 11)</td>
</tr>
<tr>
<td></td>
<td>Braque</td>
<td>La musicienne (1913)</td>
</tr>
<tr>
<td></td>
<td>Braque</td>
<td>The musician (1917 - 18)</td>
</tr>
<tr>
<td></td>
<td>Gleizes</td>
<td>Dancer (1917)</td>
</tr>
<tr>
<td></td>
<td>Léger</td>
<td>Seated woman (1913)</td>
</tr>
<tr>
<td></td>
<td>Léger</td>
<td>Soldier smoking (1916)</td>
</tr>
<tr>
<td></td>
<td>McDonald</td>
<td>Synchrony in purple (1917)</td>
</tr>
<tr>
<td></td>
<td>Wright</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Nude with draperies (1907)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Portrait of D.H. Kahnweiler</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Nude (1910)</td>
</tr>
<tr>
<td></td>
<td>Vilhon</td>
<td>Portrait of madame Y.D. (1913)</td>
</tr>
<tr>
<td>Low categorizability</td>
<td>Braque</td>
<td>Seated man with a guitar (1911)</td>
</tr>
<tr>
<td></td>
<td>Duchamp</td>
<td>Nude descending a staircase (1911 - 12)</td>
</tr>
<tr>
<td></td>
<td>Léger</td>
<td>The typographer (1919)</td>
</tr>
<tr>
<td></td>
<td>Mondrian</td>
<td>Female figure (1912)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Female nude (1910 - 11)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Clarinet player (1911)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Man with a pipe (1915)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Man with a mandolin (1911)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Ma jolie (1911 - 12)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Man with a violin (1912)</td>
</tr>
<tr>
<td></td>
<td>Picasso</td>
<td>Man leaning on a table (1916)</td>
</tr>
<tr>
<td></td>
<td>Udaltsova</td>
<td>At the piano (1914)</td>
</tr>
</tbody>
</table>
response times and those measured in our replication sample. No significant differences were found for HC ($t(15) = -0.82, p = 0.42$), IC ($t(11) = -0.32, p = 0.75$), nor LC ($t(11) = 0.11, p = 0.91$) images. In order to confirm that these category assignments affected the speed of recognition on a within-subjects basis, we computed the subject-specific rankings that were assigned to each image based on their individual response times (i.e., ranks 1 through 40, comprising fastest to slowest response for each subject). With these distributions of relative rankings, we can confirm – using two-sample $t$-tests – that HC images occupy higher rank values than the neighboring category of IC images ($t(474) = -16.09, p < 0.001$) while IC images occupy higher ranks than LC images ($t(406) = -5.99, p < 0.001$). The distribution of these ranks and their interquartile ranges can be seen in Fig. 2B. As a final comparison to the response times reported by Hekkert and van Wieringen (1990), we compare the distribution of our replication samples’ response times to their reported means for each image (Fig. S2A, B, C). We find that 34/40 of their reported averages fell within the 95% C.I. bounds of our measured distributions of response times. The averages that fell outside of these bounds were associated with 3 images from the HC category, 2 from the IC category, and 1 LC image. In the absence of further data from the original experiment by Hekkert and van Wieringen (1990), we are unable to speculate further on these individual image discrepancies. We are nonetheless able to conclude that the original category assignments are still effective at eliciting broad differences in the speed of person perception.

### 3.2. Ratings task(s)

We first sought to test the hypothesis that HC (least abstract) images elicited the greatest magnitude of ratings between subjects. First, a one-way between subjects ANOVA confirmed that ratings were differentially distributed across the three image categories.

This was true in both the 30 ms presentation condition ($F(2, 2397) = 9.55, p < 0.001$) and the subsequent unlimited duration condition ($F(2, 2397) = 13.47, p < 0.001$). Additional post-hoc two-sample $t$-tests confirm the predicted directionality of this ordering (Fig. 3A). Across both viewing conditions, HC images received higher ratings than IC images ($t(1678) = 2.15, p < 0.05$ for the 30 ms viewings; $t(1678) = 2.65, p < 0.01$ for unconstrained viewings). Meanwhile, IC images received greater ratings than LC images ($t(1438) = 2.06, p < 0.05$ for the 30 ms viewings; $t(1438) = 2.37, p < 0.05$ for unconstrained viewings). Interestingly, we also observe that ratings decrease between each mode of presentation across all three categories – an observation that is covered in more detail in the analysis below, on viewing times and subsequent changes in ratings.

The differences between categories were observed despite the fact...
that the preference orderings for each image were highly specific to each individual. To confirm this, we computed the average Spearman rank-order correlation between each subject and the remaining subjects in the sample. The Spearman rank-order correlation was used here as a measure of agreement in the ordering of each category's images. Preference ranks were used instead of ratings in order to account for subject-specific baseline ratings and rating differences between ranks. Each image's rank was determined based on the ordering of the mean rating provided to each image (recall that each image was presented 3 times each during each phase). The distribution of these rank-order correlations (comprising 19 observations per subject) was not significantly different from zero for either the HC ($t(379) = -0.88$, $p = 0.38$ for 30 ms views; $t(379) = -0.41$, $p = 0.68$ for unlimited duration views) or IC images ($t(379) = 1.59$, $p = 0.11$ for 30 ms views; $t(379) = 1.58$, $p = 0.11$ for unlimited duration views), supporting the notion of individual-specific preference orderings. For the LC images, there was a statistically-significant, albeit modest level of agreement between subjects in both presentation conditions. The average Spearman's correlation ranged between 0.06 and 0.12; Fig. 3B ($t(379) = 3.88$, $p < 0.001$ for the 30 ms condition while $t(379) = 8.39$, $p < 0.001$ for the unlimited duration condition).

Next, we sought to replicate the previously reported correlations in studies of briefly presented artworks (Verhavert et al., 2018). We do this by computing Pearson correlations between each subject's 30-ms and unlimited viewing judgments (for each image). We replicate the range of correlations previously reported – with the greatest aggregate correlations being observed in the HC category of images (Fig. 3C). Each distribution of correlation coefficients were significantly greater than zero ($t(19) = 10.82$, $p < 0.001$; $t(19) = 5.29$, $p < 0.01$; and $t(19) = 2.99$, $p < 0.01$ for HC, IC, and LC images respectively). A one-way between subjects ANOVA furthermore confirms that the means of these correlations between groups were not equal ($F(2, 57) = 4.41$, $p < 0.05$). Further post-hoc two-sample t-tests indicate that the correlations in the HC category were greater than those in the LC category ($t(38) = 3.17$, $p < 0.001$), though there was not a significant difference between HC correlations and its neighboring IC category ($t(38) = 1.72$, $p = 0.09$, in a two-tailed test). Overall, all three image sub-types replicate significant rating correlations following brief viewing durations; in addition, the average correlations obtained follows the predicted ordering, with HC images eliciting the most highly-predictive ratings following brief presentations.

3.3. Spotlight task

We first confirmed that the selective revelation of individual images was concentrated at known regions of interest, as demonstrated in similar eye-tracking exercises in the past (e.g., Yarbus, 1967). As has been previously shown, the areas of images that elicited greater viewing densities were generally the face and bodies of the persons being portrayed (Figs. 3 and S1, S2, S3). In order to visualize the density of mouse-coordinates that were associated with each image, we summed the overall density of mouse-coordinates recorded from each subject in individual pixel-sized bins (belonging to each image's presentation frame size of $333 \times 572$ pixels). The resulting density counts were smoothed using a Gaussian kernel (a convolution between the density counts and a square Gaussian filter matrix of 20-pixels, featuring a standard deviation of 5 pixels). The resulting heatmaps are plotted for visualization purposes in Fig. 4, with the full range of images available in Supplemental Figs. 1–3. The heatmaps are overlaid above an edge-detected version of each image, in order to best visualize the regions that correspond to areas of greater viewing density.

The first quantitative hypothesis we sought to test using the spotlight paradigm was whether HC images elicited greater amounts of relative attention between categories; we measure this in a non-parametric manner using the overall density of recorded mouse coordinates, collected within each frame. Here we are interested in testing the specific hypothesis that HC images are preferentially viewed over both IC and LC images. Again, we began with a one-way between subjects ANOVA performed on the average time dedicated to each image category across trials (Fig. 5A). There is a significant difference between mean viewing times across categories ($F(2, 2877) = 13.46$, $p < 0.001$). Subsequent post-hoc two-sample t-tests confirm that HC images receive greater viewing time compared to IC images ($t(1918) = 4.02$, $p < 0.001$); while IC images receive more viewing time than LC images ($t(1918) = 4.80$, $p < 0.001$).

Given that there is consistently one depicted person per image, we also examined whether the above viewing trend is robust over time. In other words, do subjects keep choosing to view HC images even after repeated viewings? To answer this question, we compare the fraction viewing time between the first and last blocks of trials (each comprising
presentations of the complete image sets). We find no statistically-significant differences between blocks for either HC ($t(651) = 1.54, p = 0.12$) or LC images ($t(651) = −1.43, p = 0.15$), categories where such effects are most likely to occur. We note however, the apparent trend in which viewing time lowers for HC images and increases for LC images. In order to visualize and report this potentially emerging effect, we computed the fraction viewing time over each trial and averaged this metric across individual trials (Fig. 5B).

3.4. Viewing time and subsequent changes in ratings

Finally, we sought to explain the variance in ratings observed between the beginning of the study (with the 30 ms ratings task) and the end of the study (with the unlimited viewing task), as a function of intervening viewing time. We first observe above that ratings generally decline by the end of the study, regardless of the image’s category (Fig. 3A). Given that the main intervening experiences of the subject (between the two rating blocks) lie in the spotlight task, we ask: what is the relation between cumulative viewing time, and subsequent rating changes? To inspect this, we computed the average rating change for images according to their cumulative total viewing time (Fig. 6). Since the total viewing time is unique to each trial, subject, and image, we binned this variable into bins of equal 0.25 s widths; this interval represents increases of 5% of the total possible viewing time. We then computed the average change in ratings associated with these cumulative viewing time bins. Consistent with the directionality of a positive familiarity bias (wherein greater amounts of viewing time should correspond to greater degrees of appreciation), we find a significant positive correlation between total viewing time spent on an image, and the magnitude of the subsequent rating change by the end of the task ($r = 0.60, n = 60, p < 0.001$). We also hypothesize that HC images will be more adept at fostering familiarity-based increases in ratings. We find a supportive trend for this, with HC viewing time exhibiting the strongest positive association with subsequent increases in ratings, followed by IC and LC viewing time respectively (Fig. 6). When testing for this trend at the individual level, we find that the correlation coefficients for HC and IC images (between unique cumulative viewing times, and their ensuing changes in ratings) were not significantly different over blocks of trials, but does appear to trend in the following manner: LC images increase in viewing time, followed by a concomitant decrease in viewing time for HC images. Continuous lines in grey depict the averages of binned trials in 10 trial bins.

Fig. 4. Visualizations of filtered mouse-tracking density data gathered from example (edge detection-filtered) images belonging to HC, IC, or LC categories (pictured respectively from left to right). Attention appears distributed focally toward known regions of interest, e.g., facial features of portrayed subjects.

Fig. 5. (A) Viewing time in the spotlight task favors images belonging to the high categorizability category, followed by images classified as medium and low categorizability. (B) The proportion of viewing time dedicated to each category is not statistically-different over blocks of trials, but does appear to trend in the following manner: LC images increase in viewing time, followed by a concomitant decrease in viewing time for HC images. Continuous lines in grey depict the averages of binned trials in 10 trial bins.
different ($t(38) = -1.12, p = 0.27$); however, IC correlations were significantly greater than those measured over LC images ($t(38) = 2.66, p = 0.00$).

4. Discussion

We firstly observe that despite heterogeneous tastes for specific images, highly categorizable portraits were generally rated as more visually pleasing. This is true following first-time views at short presentation durations, as well as by the end of the task when unlimited viewing time was allowed. We also find that viewing time is preferentially devoted toward highly categorizable images. In terms of viewing time was allowed. We also sent presentation durations, as well as by the end of the task when unlimited visually pleasing. This is true following between viewing time and rating changes. Taken together, we find that the relative ease of person perception changes how aesthetic evaluation begins and proceeds. These differences manifest within first impressions, viewing time allocation, and subsequent rating changes. In relation to existing frameworks in empirical aesthetics, these three types of effects can be seen to correspond to different stages of aesthetic processing (Chatterjee, 2003; Leder, Belke, Oeberst, & Augustin, 2004). Firstly, brief duration or single-glance preferences are likely driven by coarse categorical information obtainable during sub-second exposures, thus favoring highly categorizable images. The brief (30 ms) time window from which this effect occurs suggests that visual object recognition – not necessarily arising from artworks – highly influences aesthetic judgments made at these timescales. Further work should examine whether other types of semantic information (e.g., of animate and inanimate objects) can influence early aesthetic judgments. Secondly, the spotlight task demonstrates that viewing time is continuously and preferentially given to highly categorizable portraits, supporting the view that personified content is desirable beyond immediate object recognition (Crozier & Greenhalgh, 1992). Finally, the effects of viewing time seen here underscores a difference between processing viewing time (negatively related to preferences) and preferential or surplus viewing time (positively related to preferences), discussed further below.

Our results also broaden the range of previously described effects regarding realism and aesthetic valuation (Schepman, Rodway, Pullen, & Kirkham, 2015; Vessel & Rubin, 2010). This study presents a comparison of a different kind than those explored in previous reports, e.g., of real-world vs. fractal images in Vessel and Rubin (2010), and of various abstract vs. representational art in Schepman et al. (2015). Departing from distinct categories of image content, our study focuses on the level of abstraction specifically within depictions of people. As in the case of the results from Schepman et al. (2015), we observe that perceived representativeness is preferred over abstraction (Fig. 3A). In terms of agreement between subjects, Vessel and Rubin (2010) and Schepman et al. (2015) also found that their respective real world and representational images elicited greater degrees of inter-observer agreement. Despite this, in the present study, the greatest (albeit modest) level of rankwise agreement was found between ostensibly abstract (LC) images (Fig. 3B). One interpretation of this pattern is that the perceived realism of human figures facilitates individual-specific appraisals of facial or personal attractiveness (Germine et al., 2015; Hönekopp, 2006; Rhodes, 2006). Notably, Leder, Goller, Rigotti, and Forster (2016) using the method of Hönekopp (2006) find that 40% of face attractiveness variability can be ascribed to private (individual-specific) taste, relative to 75% for abstract art. On this topic, our study raises a new observation regarding abstract faces/bodies, for which there might be an increased level of shared beauty standards; though, further study is warranted to overcome the small number of images (12–16) used within each category here. Vessel and Rubin (2010) notably suggested that collectively shared semantic associations lead to greater levels of aesthetic agreement. The relatively low degree of agreement for HC images suggest that easily-perceived people does not necessarily elicit a high level of inter-rater agreement. This might be due to the rich set of social inferences – e.g., of emotions, identity, race, etc. – that can form downstream of person perception. Future extensions of the approach we set out here could involve testing abstract vs. realistic depictions of varying subject matter, e.g., of inanimate objects. In addition, the boost in aesthetic appreciation when humans are easily detectable ought to be investigated in other types of visual stimuli. Previously studied categories of images that might be susceptible to the effects we observe here include natural scene photographs with and without people (e.g., Fei-Fei et al., 2007), as well as abstract patterns embedded with human-like details (e.g., Muth & Carbon, 2013).

Our results validate short presentation durations and apparent levels of abstraction as comparable constraints on aesthetic processing. We observe that the range of intra-subject correlations (between 30 ms and unconstrained viewings) is similar to those reported in Verhavert et al. (2018). Interestingly, our HC images exhibit an average correlation coefficient comparable to their 100 ms and unconstrained viewings. By the same token, correlation coefficients measured using IC and LC images might be seen as comparable to their 50 ms and 10 ms treatments respectively. Our results thus suggest that categorizability...
modifies the stability of aesthetic first-impressions on a similar scale to having different time constraints. Although Augustin et al. (2008) found evidence for content-based aesthetic processing following presentations as short as 10 ms, the results here demonstrate that perceptually ambiguous (LC) images limit aesthetic processing in the same manner as constrained viewing time does. Related to the issue of processing constraints, the finding that time-limited ratings are greater than their unlimited counterparts (Fig. 3A) might appear to be incongruent with the effect of preferential viewing time (in the Spotlight task). Without further duration conditions, we are unable to determine whether fixed initial presentations benefit from shorter durations. Such an effect would be consistent with the known effects of increased processing time on decreased liking (e.g., Kuchinke et al., 2009), discussed and disambiguated from preferential viewing time below. In addition, although we replicate the scale of within-subject correlations documented by Verhaver et al. (2018), their aggregate ratings appear to not significantly change across their range of tested durations. A simple albeit study-specific explanation of this observation is that brief aesthetic judgments incorporate additional assumptions about images relative to unrestricted viewings. In the former, holistic “gist” impressions (Oliva, 2005; Oliva & Torralba, 2006) of the portraits we selected might influence early judgments, taking into account assumed features (e.g., assumed symmetry or simplicity) or sensory-level information (e.g., shading and shape) that are updated upon extended viewing.

When it comes to the resulting changes in ratings, we are able to attribute a significant portion of this variability to a positive familiarity bias. That is, images that accrue more viewing time tend to be rated more highly by the end of the experiment. A causal link between familiarity and processing fluency was initially established under fluency-attribute accounts (Jacoby & Dallas, 1981), whereby memory traces of previously experienced stimuli facilitate performance on a task (e.g., word recognition). The positive effect of viewing time here contrasts results that pair processing time against liking. Measures such as classification time (Reber et al., 2004) and recognition time (Kuchinke et al., 2009) are negatively related to preferences, as longer times in such cases represent a reduced ease of processing. Critically, this experiment involves several passive exposures prior to exploratory viewing during the spotlight procedure. In this way, time distributed freely toward stimuli – beyond requisite processing time – should be predictive of higher liking. Given this distinction between processing and surplus viewing time, future studies ought to delineate when this distinction occurs and how this process aligns with the visual recognition of focal subject matter. Furthermore, although it is difficult to tell from our data alone whether increasing aesthetic value causally promotes increased viewing time (or the other way around), this result is consistent with other non-aesthetics studies demonstrating that: (i) gaze positively modulates resulting preferences for faces (Shimojo, Simion, Oliva, & Scheier, 2003), (ii) long-term familiarity determines liking for face views (Mita, Dermer, & Knight, 1977), and (iii) multiple subliminal exposures lead to increased positive affect (Monahan, Murphy, & Zajonc, 2000). Interpreted collectively, these results suggest that either a fluency-based (Winkielman & Cacioppo, 2001) or general familiarity effect is operating here, with HC portraits cultivating the most positive increases in ratings following a fixed duration of views. Given that these effects are likely driven by the perceptual discovery of singular portrait subjects, future studies might investigate whether the same results hold for non-naïve individuals already familiar with such artworks. For instance, subtle features of portraits such as profile angles (Humphrey & McManus, 1973; Schirillo, 2000), and pose similarity with the viewer’s posture (Bertamini, Byrne, & Bennett, 2013) are known to affect viewer appreciation – such effects ought to arise after the initial process of person perception. Thus, the effects of viewing time here can be understood as describing short-term aesthetic value assignment (under 30 s), following person recognition in untrained viewers.

There are also indicators regarding limits on these preferential viewing effects. Our data suggest that the attribution of viewing time might further change over a larger number of trials than those tested here (Fig. 5B). Indeed, the maximum possible viewing time per non-HC image in our task (25 s) is less than the mean viewing time (< 30 s) measured in a live museum setting (Smith & Smith, 2001). Although there are no significant differences in viewing time distribution between repeated viewings, the trend in Fig. 5B (as well as the ordering of means in Fig. 5A) suggests that LC views might eventually exceed HC views after sufficient exposure. This is a natural trend to expect, if one assumes that subjects eventually seek out additional perceptual discoveries (i.e., “a-ha” moments as by Muth & Carbon, 2013) within seemingly abstract images (once the content of HC and IC images are processed), or if the ostensibly more challenging art was meta-cognitively chosen to be viewed/appreciated later (Kahn, Ratner, & Kahneman, 1997). Further questions on this issue will center on extended presentation and viewing durations (beyond the range our subjects performed), to gauge the larger range of possible associations between viewing time, presentation time, and rating changes.

Overall, our results highlight how one kind of visual object recognition – person perception – influences the assignment and development of aesthetic ratings. These effects occur at several different timescales, ranging from ratings made following 30 ms of initial views, to those made without any duration constraints. We also document how this form of perceptual categorizability modifies viewing patterns in tandem with rating changes per unit time of viewing. The detectability of human figures appears to boost aggregate evaluations as well as the rate of evaluative changes over time. Further research is warranted to determine whether these effects are specific to depictions of people or generalizable to other focal content in artworks. In the latter case, aesthetic evaluation might be generally facilitated by levels of ease in forming conceptual or semantic associations, consistent with broader accounts of processing fluency. If so, further experiments of this kind will be able to identify other visual attributes that independently affect early and late phases of aesthetic processing, a distinction emphasized in other recent studies (e.g., Belke et al., 2015; Graf & Landwehr, 2015; Thönnes & Hübner, 2014).

Appendix A. Supplementary figures

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actapsy.2019.05.006.

References


Affective priming while using facial expressions modulates liking for abstract art. PLoS One, 8(11), e80154.


