Published in: *Acta Psychologica* 192 (2019) 11-22
https://doi.org/10.1016/j.actpsy.2018.10.012

Art expertise in construing meaning of representational and abstract artworks

David L. Bimler[1], Megan Snellock[2], Galina V. Paramei[2]

[1] Massey University, NZ; [2] Liverpool Hope University, UK

* Corresponding author:
Dr. David L. Bimler
School of Psychology
Massey University
Private Bag 11-222
Palmerston North, New Zealand

Email: d.bimler@massey.ac.nz

Word count (excl. figures/tables/References): 7635

Keywords: Art expertise; aesthetic appraisal; semantic differential; factor analysis; Abstract / Representational; cognitive mastery

RUNNING HEAD: Art expertise in construing the meaning of paintings

Abstract

Aesthetic appraisal of artwork can present the observer with visual problems to solve in the process of grasping of its meaning and 'visual rightness' (i.e. "good" structure; Locher, 2003), with an elaboration on perceptual, semantic and affective dimensions (e.g. Marković, 2011). Thus observer's expertise is a factor in aesthetic appraisal. To examine the influence of art training on the aesthetic response, and to clarify the nature of the Representational / Abstract distinction, 30 *Experts* and 33 *Non-experts* (Art and Psychology students, respectively) were asked to rate 24 paintings on six affective and affective-evaluative semantic differential scales. Stimuli were images of paintings from the period 1900-1935, 12 broadly Representational and 12 broadly Abstract. Relative to *Non-experts*, *Experts* rated Abstract artworks as more *Interesting, Beautiful, Informative* and *Sophisticated*, distinguishing them less markedly from Representational artworks. Aggregate *Expert* and *Non-expert* ratings, processed by factor analysis, resulted in a two-factor solution. The first factor, contrasting Abstract and Representational artworks, appeared more salient for *Non-experts*. The second factor, *Cool–Warm*, separating vibrantly-colored paintings from those with a blue-dominated/dull palette, was more salient for *Experts*. While *Non-experts* exaggerated differences between Abstract and Representational paintings, *Experts* appraised these two types of art similarly, attending more to artwork collative properties. We conclude that appreciation of art by *Experts* involves 'cognitive mastery' (Leder, Belke, Oeberst, & Augustin, 2004), i.e. more complex, cues-based visual schemata which equip them with more sophisticated strategies for analysing collative properties and semantics of an artwork while parsing 'visual rightness' to unfold its visual meaning.

Suppose there was some great God of Aesthetics. And then whenever you made a painting, no matter how much you liked it, … you would submit it to the great God of Aesthetics and the God would say, “This is good,” or, “This is bad.” After a while the problem is for you to develop an aesthetic sense that fits with the thing, not just your own personal feelings about it.

Richard Feynman, in: L. Mlodinow, *Feynman’s Rainbow. A Search for Beauty in Physics and Life* (2011, p. 131)

1. Introduction

Two individuals may differ in their personal tastes, perhaps preferring different degrees of complexity or novelty. But independently of that, observers often disagree on the qualities they extract from an artwork to compare against those ideals – that is, they vary in how they appraise the artwork, ascribing it different values on a scale such as ‘simple/complex’, or different degrees of similarity to a second artwork. These variations invite an examination of the processes involved in art appraisal. A particular focus of the present study is the elusive concept of “art expertise”, one conspicuous form of individual variation. A second focus is the concept of “Abstraction”, which is in contrast a quality that varies between artworks. This Introduction integrates various studies on these two concepts, how they interact and where they might fit in a ‘Unified Theory of Aesthetic Response’. We also sketch some of the methods available for quantifying aesthetic responses, to provide a certain context and justification for the questions addressed here.

1.1. Multi-stage models of aesthetic experience as frameworks for art expertise

Numerous studies have examined changes in observers’ performance resulting from art training and “expertise”. These changes manifest in visual search strategies (Nodine, Locher, & Krupinski, 1993; Pihko et al., 2011), with greater attention to abstract features and structural relations between depicted objects (Vogt & Magnussen, 2007). Art experts also differ from novices in their projective test responses (Bilotta & Lindauer, 1980); strategies for sorting artworks by similarity (Augustin & Leder, 2006; Frechtling & Davidson, 1970); and ratings on scales of complexity, novelty, pleasingness (Stojilović & Marković, 2014), etc.

Collectively, these findings point out that compared to laypersons, art experts follow qualitatively different schemata of artwork aesthetics, focusing on non-denotative sensory qualities and higher-order semantic properties of artworks.

Multi-stage models of aesthetic response are especially relevant for present purposes (reviewed by Pelowski, Markey, Lauring, & Leder, 2016). Among these, Leder, Belke, Oeberst, and Augustin's model (2004; also Leder & Nadal, 2014) is couched in terms of information processing. In its latest elaboration, the Vienna Integrated Model in Art Perception (VIMAP; Pelowski, Markey, Forster, Gerger, & Leder, 2017), aesthetic experience is conceived as a cognitive process that starts with analysis of low-level features (color, luminance, contrast etc.) and progresses to higher-level processing by involving interpretations of the artwork.

The stages of sophistication of aesthetic appraisal are explicitly developmental (see also Parsons, 1987). Crucially, multi-stage models allow for individuals to differ in which stage of processing is dominant, or which level of visual complexity they prefer, on a scale from low-information simplicity to random unpredictability. Aesthetic experience progression unfolds from naïve, single-perspective hedonic ('liking') responses to highly conceptualized evaluations of the image as a cultural artifact ('cognitive mastery'). At the 'liking' stage, the art stimulus is observed in a conceptual vacuum, while in 'cognitive mastery' it is embedded in a context as part of a cultural dialog embracing historical meaning and artistic factors of the artwork, then judged on these additional terms. Augustin and Leder (2006, p. 136) go on to argue that "the 'outcome' of any aesthetic episode in the sense of pleasure, displeasure or any kind of judgment of an artwork cannot be fully understood without knowledge of the stimulus aspects and the cognitive concepts that were relevant during the processing and interpretation."

Along the developmental pathway to cognitive mastery, individuals differ in their preferred level of artwork complexity. As formulated by Silvia (2013), complexity and novelty in an artwork distinguish experts from novices by evoking a positive value of 'coping potential' from the former and a negative value from the latter; 'complexity' combines with these to produce the 'knowledge emotions' of Interest or Confusion, respectively.

A complex or ambiguous image (e.g. a pattern of irregular blotches on a contrasting background) requires the observer to marshal cognitive resources. If the ambiguity is then resolved in a delayed closure, filling in missing details to reveal a simple reinterpretation, then those resources are released again and aesthetic satisfaction follows. In the words of Belke. Leder, and Augustin (2006, p. 116), “the experience to dissolve perceptual or conceptual ambiguity is assumed to influence the continuous affective evaluation of an artwork positively”. In their multi-stage information-processing model, this occurs at the intermediate stage of ‘explicit classification’ where an artwork is categorized – encoding unique artist’s style and the content’s historical/contextual meaning. In a dialectic between bottom-up intrinsic perceptual variables and top-down acquired knowledge, visual complexity of an artwork is reduced to schematic, propositional form (Lindell & Mueller, 2011; Pelowski et al., 2017).

1.2. Abstraction

As foreshadowed above, there is an interaction between observer expertise and each artwork’s level of abstraction. It is widely reported that naïve and expert observers tend to agree in their assessments of Representational art; assessments of Abstract artworks are where they diverge (e.g. Leder, Gerger, Brieber, & Schwarz, 2014; Leder, Gerger, Dressler, & Schabmann, 2012; Mullennix & Robinet, 2018; Pihko et al., 2011). Artists (or at least art students) attend to ‘style schemes’ in Abstract art to which non-artists (e.g. psychology students) are oblivious (Cela-Conde, Marty, Munar, Nadal, & Burges, 2002).

Hekkert and van Wieringen (1996) presented subjects with an array of post-impressionist works and made the images more abstract by removing color and/or fine figurative detail; these manipulations had negative impact on aesthetic appraisal, which decreased with expertise (see also Neperud & Marschalek, 1988, for a similar interventional approach). This might even serve as an operational form of the Abstract/Representational distinction: the more differently an image is processed by experts and non-experts, the higher its level of abstraction.

In another intervention study, Locher (2003) displaced compositional elements in a range of art stimuli and asked subjects to recognize the original from the manipulated version. A breakdown of their stated rationales revealed that Abstract paintings tended to evoke a more cognitive level of consideration relative to the concrete, hedonic approach applied to Representational ones. This further illustrates the integral connection between the stages of aesthetic processing, and differences among art styles and types – in particular, the distinction between 'Abstract' and Representational types.

As the name implies, the Abstract mode aspires to generality, progressively stripped of the litany of particulars (e.g. recognizable "thingness": scenes, motifs, or persons). The absence forces the observer's attention to focus on aspects such as line, palette, technique, composition, although all of these are also important aspects of Representational art. The rise of Abstract styles and movements is a relatively recent development in art history, and accompanies an opening separation between experts and naïve observers. We note that Marković (2011) and Leder et al. (2012) write about a Representational/Abstract *dimension*, or level of abstractness, while Pihko et al. (2011) recognize three intermediate categories between the two extremes. Although it is convenient to treat the distinction as a sharp dichotomy, it is more of a continuous gradient: even the most realistic, academic painting has elements of stylization and convention (Chatterjee, Widick, Sternschein, Smith II, & Bromberger, 2010).

1.3. Experimental approaches for artwork appraisal

In an influential paradigm of quantifying aesthetic experience the focus is on *subjective similarities* among images (Augustin & Leder, 2006). Observers' ratings of pairwise similarity are summarized by converting them into a geometrical model or 'map', via multidimensional scaling (MDS; e.g. Avital & Cupchik, 1998; Berlyne, 1975; Berlyne & Ogilvie, 1974; Cupchik, 1974; Goude, 1972). In the context of art expertise, MDS provides individual-difference models that handle systematic observer variation (O'Hare & Gordon, 1977). The axes locating points within the spatial solution can tentatively be identified as attributes used to compare and contrast the images, each contributing to perceived dissimilarity, e.g. "Realistic–Unrealistic" and "Clear–Indefinite" (O'Hare, 1976). The task of eliciting similarities for all pairs for the MDS approach limits the number of paintings that can conveniently be sampled. This in turn limits the ranges of inter-stimulus variation that can be captured.

In contrast, in the semantic-differential method stimuli are presented one at a time, asking subjects to rate each one on a battery of affective or evaluative scales, with scales defined by pairs of antonymous extremes. Thus it allows a richer stimulus set to be probed.

The field has not reached consensus on the optimal number of scales, and which ones, to capture the observers' appraisals. In his pioneering studies Berlyne (1976; Berlyne & Ogilvie, 1974) opted for 12 bipolar scales, such as Complexity, Familiarity, Unexpectedness, Regularity, describing collative (informational) properties of artworks, sampling their Valence (Hedonic tone), Arousal and Uncertainty.

In the following years the number of employed scales varied substantially depending on the researchers' theoretical framework. Hekkert and van Wieringen (1996) provided five scales for assessing monochrome images and two more for the colored versions. Van Paasschen, Bacci, and Melcher (2015) employed four scales that, according to their a priori paradigm, suffice to capture the high-level qualities of artworks: Beauty, Valence, Liking and Arousal.

At the other extreme, Augustin, Carbon, and Wagemans (2011) started from first principles and tasked their subjects with a fine-grained pool of 77 unipolar scales, which they reduced to 23 in a second experiment. Marković and Radonjić (2008) elicited 25 and 43 antonym pairs for quantifying 'explicit' and 'implicit' qualities, respectively. Marković (2011) developed these into 32 scales arranged in a typology of three domains: 12 perceptual-descriptive, 12 affective-aesthetic, and eight (unipolar) 'semantic' scales. These domains correspond loosely to the levels of information processing in the multi-stage models addressed above, from reductionist objective descriptors to abstract subjective evaluations, i.e. comparable to the bottom-up/top-down extremes in Lindell & Mueller (2011) and Pelowski et al. (2017).

This level of detail may be unwarranted, though, and the exact choice of scales may not be crucial in studies of this kind. The presence of subtle distinctions among the pool of scales – e.g. between *Beauty* vs. *Pleasingness* (Valence) vs. *Preference* – is no guarantee that the subjects will use these niceties, or understand the analytical, philosophical refinements that the researchers intended. Indeed, Marković (2011) noted that a *Representational/Abstract* distinction could be discerned in each of his three parallel domains, implying an element of redundancy.

1.4. Broader context: multifaceted nature of artwork appraisal

The multifaceted nature of art appraisal is a recurring theme in the literature on experimental aesthetics. Paintings can provide visual pleasure by their color harmony, symmetry, simplicity, balanced composition (or perceptual goodness and closure, to borrow the terms of Gestalt Psychology), familiar motifs or 'fluency of processing' (Lindell & Mueller, 2011); in other cases, artwork complexity, novelty, intricacy, ambiguity or stimulation may afford more cerebral satisfaction (Muth, Hesslinger, & Carbon, 2015).

Berlyne (1976; Berlyne & Ogilvie, 1974) grouped the collative properties of artworks under the rubrics of Valence (Hedonic tone), Arousal and Uncertainty: three factors of subjective response inspired by Osgood's Evaluation, Potency and Activity (EPA) scheme (cf. Osgood, May, & Miron, 1975). The Valence/Hedonic tone factor also emerged from a number of analyses (e.g. Neperud, 1970), loading on scales such as "Ugly−Beautiful" or "Unpleasant−Pleasant", that capture the accessible 'liking' aspects of appraisal.

The other two factors identified by Berlyne (counterparts of Osgood's Potency and Activity) are recurring too: Arousal (unexpectedness, unbalance, stimulation) and Uncertainty (complexity, ambiguity) – "the extent to which a stimulus pattern (an art object, for instance) calls for information processing on the part of the viewer" (Alluisi, 1975, p. 520). Stojilović and Marković (2014) recently arrived at a second-order factor of General Aesthetic Experience, but this subsumed three first-order factors, Arousal, Aesthetic Experience, and Relaxation Tone.

As discussed above (1.1), experts appear to differ from naïve observers in having access to additional frames of reference for interpreting and grasping the underlying simplicity within a novel work. Van Paasschen et al. (2015) recently reported that experts scored stimuli more highly than novices on an Aesthetic scale (Beauty) without differing on an Affective scale (Valence), as if their training allowed the experts to see beauty invisible to novices and separate from Hedonic tone. Mullennix and Robinet (2018), indeed, found that novices and experts did not differ in their 'liking' but expertise was positively correlated with 'understanding' judgments. Thus, in line with the 'cognitive mastery' stage of the multi-stage model (Pelowski et al., 2017), 'expertise' might be rephrased as 'possession of multiple schemata'. Specifically, art-trained viewers possess "more elaborate and complex categorizations of art" (Neperud, 1988, p. 293), which enable them to actively process the information available in artworks to make sense of them, or exercising the 'coping potential' (Silvia, 2013), resulting in deeper aesthetic experiences. It follows that the optimal level of complexity – enough to call upon cognitive resources, but not enough to obscure the ambiguity completely – is training-dependent.

1.5. Hypotheses

In the present study, we compared appraisal of artworks by university Art students ('Experts') and Psychology students ('Non-experts'). We anticipated that an important aspect of art expertise is the possession of multiple frameworks or schemata for 'explicit classification', leading to an increase in comprehension of artworks and, hence, their aesthetic appreciation (Pelowski et al., 2017), by resolving their ambiguities and recognizing 'visual rightness' (Locher, 2003). We hypothesized that art *Experts* will (1) reveal more elaborate judgments of evaluative-affective meaning of artworks compared to *Non-experts*, who will focus on simpler criteria to dominate their judgments [Expertise]; (2) vary more among themselves than Non-experts do, especially in their appraisal of *Abstract* artworks; and (3) construe the meaning of *Abstract* works more positively than Non-experts, who will judge *Representational* artworks higher than *Abstract* ones [Artwork Type].

For stimuli we selected 24 paintings, from a period in the early 20th century, to avoid diachronic forms of variation. Artwork appraisal was assessed using six semantic differential scales. In effect, for each of *N* observers the stimuli were located as a point within a six-dimensional space; collectively, observers' responses located each stimulus in 6 x *N* dimensions. The goal of analysis was to reduce the data to a parsimonious account, taking advantage of any interdependence and redundancy among the scales to fit them and the points into a lower-dimensional space, while exploring differences in the spaces of art *Experts* vs. *Non-experts*.

2. Method

2.1. Participants

Sixty-three participants were recruited at a North-West England university; all were 1st, 2nd, 3rd-year students or undertook internship shortly after graduation. Fine Arts students (N=30; 5 males), *Experts* (*E*s), were recruited as volunteers through opportunity sampling; they were aged 19-52, with mean age 22.9 ± 6.0 years. Psychology students (N=34), *Non-experts* (*NE*s), were recruited through opportunity sampling, apart from 1st-year Psychology students who received course credits for their participation. Data of one Psychology student, a 55-year old male, had to be removed from the analysis since the instruction was apparently misunderstood – see below. In the final sample (N=33; 9 males), age ranged between 18-40 years, with mean age of 22.3 ± 5.7 years. Ethical approval was obtained from the Ethics Committee of the Department of Psychology. All participants gave their informed consent prior to participation in the study.

2.2. Stimuli

Twenty-four artworks were selected from the period 1900-1935, 12 *Representational* (*R*) and 12 *Abstract* (*A*) (see Table 1 for the list of artists and paintings). This time window was intended to capture the period in International Modernism when the concept of 'abstraction' emerged as an autonomous category (Greenberg, 1960). The *Representational* artworks were chosen in line with the definition of depicting objects, scenes or events known by the viewer; conversely, the *Abstract* artworks were defined as those that do not depict literal objects encountered in the physical world.

Table 1 about here

To reduce confounding factors and focus on the role of *Artwork Type*, where possible, pairs of stimuli were loosely matched between the two types by subject matter: e.g. Paul Klee's *Senicio* (1922), an *A* painting of a face, was matched with the *Portrait of Anna Akhmatova* (1922), *R* painting by Kuzma Petrov-Vodkin. When the subject matter of an *A* artwork was not recognizable, a *R* counterpart was chosen on the basis of similar color palette or composition.

All images were of high resolution. They were presented on a computer monitor (13.3 in) as singletons (approximately 413px x 536px), centered (in an area of 75% x 75% of total screen size), on a gray background (RGB values 191,191,191).

2.3. Instruments

To assess responses to the paintings, we chose six semantic differential scales (SDs), guided by previous studies in empirical aesthetics that employed SD as the tool (Berlyne & Ogilvie, 1974; Biaggio & Supplee, 1983; Marković, 2011; Neperud & Jenkins, 1982; Silvia, 2006). These were 7-point scales, anchored at each extreme with antonymous adjectives (cf. Osgood et al., 1975). In the terminology of these previous studies, the scales were as follows:

1. SD1: *Boring–Interesting*
2. SD2: *Ugly–Beautiful*
3. SD3: *Annoying–Pleasing*
4. SD4: *Uninformative–Informative*
5. SD5: *Cool–Warm*
6. SD6: *Naïve–Sophisticated*

2.4. Procedure

Participants viewed single images presented in a pseudorandom sequence. As each image appeared, it was accompanied by a first SD scale, presented beneath the image, where it remained until a response was made by pressing a key 1 to 7 on the computer keyboard. The response was saved and the next scale was presented (in pseudorandom order within each trial). After the sixth scale, the image was replaced by a blank grey screen for 2000 ms before the next stimulus appeared, to minimize the possibility of imposing an afterimage upon the latter. This procedure was repeated until all 24 images had been rated, a process that took less than an hour in total. E-Prime 2.0 software was used for presenting the stimulus sequence and recording the responses. The experiment proper was preceded by a practice trial, including one *R* (Southall, *Belgium Supported by Hope*, 1918) and one *A* artwork (Bomberg, *The mud bath*, 1914), not part of the main stimulus set, to familiarize the participant with the task.

The experiment took place in a quiet room on the university campus, one student at a time, with only the researcher present. The room provided ambient daylight through a window; the participant faced the window, from which the monitor was shielded, and was seated ca. 30 cm from the monitor.

2.5. Data analysis

A natural approach with rating data is to apply Factor Analysis (FA) to reduce them to an underlying structure. However, the three-mode nature of the present data (subjects by stimuli by scales) pose an analytic challenge. Algorithms do exist for factorizing three-mode data sets, and have been applied in the art-aesthetic context (Kroonenberg, 1983; Neperud, 1970), but the results are not straightforward to interpret, and are fraught with rotational indeterminacy.

Studies (e.g. Berlyne & Ogilvie, 1974) sometimes postulate that variance within a defined subject sub-group can be ignored and removed (by averaging the ratings for each stimulus/scale combination) prior to conventional FA. This is far from rigorous, and imposes the very structure that we are testing for. By way of counter-argument, it is conceivable that a novice and an expert might differ in how they understand and apply (for instance) the 'Sophistication' scale, depending on what they bring to the art encounter. In other words, the scale in question might probe two different concepts (quite apart from different appraisals of the images themselves). But it is also possible for experts to differ among themselves. Finally, it is not obvious that novices or experts are themselves a homogeneous subgroup, with all subjects at a given stage of aesthetic development interpreting and grouping the scales identically. (Few of the Art students would have reached high mastery level, while some of the Psychology students may have acquired some art expertise independently.)

Instead, to explore *Expert* vs. *Non-expert* difference, we first applied FA to data arranged in the “stringing-out” technique, used by Osgood et al. (1975) to simplify the variance among the *scales* (cf. also Marković & Radonjić, 2008); here, though, it was used to reduce the variance among the *subjects* into their factor loadings. A second level of FA, targeting the resulting factors themselves, clarified the factors’ structures of SD scales. Then we repeated the process of stringing-out FA, focusing on the variance among the *paintings* to allow differing appraisal of the two *Artwork Types* to emerge.

As mentioned above, data of one participant were excluded from the analyses since he appeared to have rated each painting identically on all six scales, e.g. scoring the Braque at 1, Lempicka at 7, and so on.

3. Results

3.1. Factor analysis: expert and non-expert artwork spaces

Normally we would begin with a basic description of the data, reporting mean values and trends, before progressing to analysis of the internal structure. Here we reluctantly reverse that sequence. Our goal is to avoid averaging responses across any variable without first examining the *variation* on that variable to test whether there is sufficient consensus (across subjects, or scales). We also want to be able to justify grouping subjects by Expertise, or paintings by Artwork Type, and not to simply *assume* that they are distinct groups.

3.1.1. Simplifying subject variance

With three-way data, dimension reduction to a simpler form must go through several stages. We arranged the data as a table with the 63 subjects as columns, and each of the 6 (scales) x

24 (paintings) = 144 painting-scale combinations as a row, before applying FA in its simplest form (Principal Components Analysis, PCA). Each of the resulting Principal Components (PCs) is a 'prototypal subject', i.e. a prototype pattern of responses to all paintings on all scales, and the responses from each *actual* subject are approximated by a combination of the PCs, weighted by his or her loadings on them.

The first three PCs accounted for 28.33%, 5.71% and 4.29% of total variance, with eigenvalues of 17.85, 3.60 and 2.70, respectively. Thus the first component (PC1) dominates the variance, and we begin by focusing on it, treating it as an area of agreement or consensus pattern of responses shared across *E*s and *NE*s; in contrast, the second and subsequent components capture minor deviations from or modulations of this consensus.

Individual subjects' loadings on PC1 ("Consensus") – how closely they approach the prototypal observer – are almost all positive, with a mean of 0.49; crucially, these loadings differ significantly with expertise (plotted in Fig. 1). Mean loadings for PC1 are 0.41 (± SE 0.04) for *E*s and 0.57 (± SE 0.03) for *NE*s (p = 0.002). Loadings on PC2 also differ ("Distinction"; see Fig. 1): the *E* and *NE* groups have positive and negative mean PC2 values, respectively, with means 0.10 (0.04) for *E*s and –0.05 (0.04) for *NE*s (p = 0.010). On PC3, means were 0.08 (0.03) and –0.05 (0.04), respectively (p = 0.017). The between-group differences are a major mode of variation among the subjects, justifying the separate treatment of *E*s and *NE*s. In addition, as shown by dispersion of the points in Fig. 1, *E*s seem to vary more among themselves, and deviate further from PC1 (the overall dominant pattern of responses) than *NE*s.

Figure 1 about here

PCA provides the option of recovering the underlying PCs themselves as part of its output, as columns of 144 scores. Recall that each PC is a prototypal pattern of responses to individual paintings. For the analysis of *Artwork Type*, we restructured PC1 into a 24-by-6 table, with paintings as rows and scales as columns. It was then possible to apply PCA a second time, focusing on the paintings and simplifying the variance among the scales. (If we had obtained a table directly by averaging the score for each painting and scale across subjects, this would be the *first* application of PCA.) In this second-order solution, scores on the first component $PC1^{(2)}$ arranged the paintings along an *A*-to-*R* axis, although it is not a perfect split: within each *Artwork Type*, the paintings vary in the degree of appraised abstraction (Supplementary Fig. S1). Scores on the second-order $PC2^{(2)}$ indicate a quality of color, contrasting warm palettes with the cool blue palettes of paintings; we term $PC2^{(2)}$ 'Color/Warmth/Vibrancy'. Looking at the scales themselves, SD5 *Cool–Warm* loads highly on $PC2^{(2)}$, while the other five scales, used with little distinction between them, load highly on $PC1^{(2)}$.

Having established a group difference for Expertise, it is possible to use PCA to isolate and exaggerate the differences between the *E* and *NE* groups. This entails rotating the solution to simple structure, where the rotated factors correspond to prototypal *E* and *NE* response patterns, and restructuring each factor for second-order PCA as described in the previous paragraph. Fig. S2 shows the outcome. Notably, a simple *A/R* split dominates the exaggerated *NE* prototype. In the *E* prototype, the corresponding component is more nuanced, while the 'Color/Warmth/Vibrancy' feature becomes evident (Hypothesis 1).

However, it is more intuitive to repeat the whole procedure for the two groups separately, which delivers reassuringly similar results. In the solution for the 30 *E*s in isolation, the first three unrotated PCs had eigenvalues (VAF) of 7.20 (24.01%), 2.11 (7.04%) and 1.90 (6.32%). All but one PC1 loading were positive, with a mean loading of 0.44. In the solution for 33 *NE*s in isolation, the first three PCs have eigenvalues (VAF) of 11.87 (33.96%), 2.08 (6.31%), 1.86 (5.62%). All PC1 loadings were positive, with a mean loading of 0.58. For each group, the consensus is a weighted mean of responses from individual subjects (weighted by reliability, i.e. how closely they approximated the group's collective opinion). These numbers indicate an *E*-specific and a *NE*-specific consensus. As expected (Hypothesis 2), untrained subjects reveal greater within-group agreement, while *Experts* deviate more among themselves from their group consensus.

As before, we rearranged the *E*-only PC1 and *NE*-only PC1 into 24-by-6 painting-by-scale tables. For convenience of comparison and display, we stacked these into a 48-by-6 table in which each painting appears twice: once as a row describing how *E*s characterized it on the 6 scales, and then as a second row for the *NE* description.

PCA resulted in a two-component solution with eigenvalues (VAF) of 4.10 (68.40%), 0.99 (16.49%). The second component (PC2) is still identifiable as Color or Warmth or Vibrancy, and is dominated by SD5, while the first component (PC1), *Abstract-Representational*, combines the other five SD scales. Scores for the paintings are plotted as Fig. 2 with lines connecting the dual perceptions of each one. (Fig. S1 is effectively the same solution, but does not show the expertise-related differences.)

Figure 2 about here

There is a notable trend in Fig. 2 for the *NE*s' appraisal, compared to their *E* counterparts, to be more polarized, manifested by displacing points for *Abstract* paintings further to the left and for *Representational* paintings further to the right. In other words, the *Non-experts* used the scales to distinguish *A* from *R* paintings more clearly. This can be seen in the mean PC1 values for the 12 *A* and 12 *R* works, which are –0.67 and 0.59, respectively in the *NE* responses, whereas the corresponding means in the *E* responses are –0.27 and 0.36.

We note in addition that the dispersal along PC1 (i.e. the variance of scores) is greater for the *NE* appraisal of the paintings than for the *E* perceptions: 1.30 and 0.74, respectively. Conversely, the dispersal along PC2, Color/Warmth/Vibrancy, is greater for the *E* perceptions than for the *NE*s – 1.26 and 0.71, respectively. That is, *NE*s place relatively more weight on PC1, Artwork Type, when rating the paintings, and less on PC2, their palette.

3.1.2. Simplifying stimulus variance

It is also instructive to apply PCA to a reorganized table where *paintings* are the columns and the 6 (scales) x 63 (subjects) = 378 combinations are rows. This simplifies the painting variance first, reducing the responses to descriptions of a convenient number of 'prototypal paintings'.

The first four PCs had eigenvalues (variance) of 5.28 (21.98%), 2.43 (10.13%), 1.72 (7.18%) and 1.44 (6.01%), declining slowly in magnitude after the third. Guided by the scree test, we retained the first three PCs. After Varimax rotation to simple structure, the artworks were described by the factor loadings plotted in Fig. 3. The solution distinguishes between *R* works, with dominant PC1 loadings, and *A* works, closer to PC2. One can speak of PC1 and PC2 being *Representational* and *Abstract* prototypes, while PC3 could be described as a prototype of Color or Vibrancy, or stylized *Representational*, best approximated by Dix, Gauguin, Hopper and Roerich works (the paintings with highest PC3 loading). That is to say, observers' responses did implicitly distinguish the types of painting.

Figure 3 about here

3.2. Semantic differential scales: subject means, averaged by artwork type

Having established that the *E* and *NE* groups do differ in their responses, and also that *R* and *A* paintings do collectively differ, we finally return to our Hypothesis 1, with a more detailed examination of the broad, descriptive properties of the data.

For each SD scale, two sets of contrasting mean scores were calculated: for each subject *i*, means *mR*(*i*,*s*) and *mA*(*i*,*s*) for the two *Artwork Types* (i.e. across 12 *R* and 12 *A* paintings, separately); and for each painting *p*, means *mE*(*p*,*s*) and *mNE*(*p*,*s*) for the two *Expertise* subject groups (i.e. across 30 *E*s and 33 *NE*s). These contrasts are summarized in Figs. 4 and 5 as scatterplots. Fig. 4 emphasizes variation specific to the *Artwork Type*, where each panel contains a point for *each subject*, located by *mA*(*i,s*) and *mR*(*i,s*) as *x*- (*Abstract*) and *y*- (*Representational*) axis coordinates, respectively. In comparison, Fig. 5 emphasizes variation of responses related to art *Expertise*, and each panel contains a point for *each painting*, located by its *mE*(*p,s*) and *mNE*(*p,s*) values as the *x*- (*Expert*) and *y*- (*Non-expert*) axes. (Fig. S3 of Supplementary Materials displays *mE*(*p,s*) and *mNE*(*p,s*) in the form of bar graphs.) The two distributions of points in each panel are summarized by 70% confidence-region ellipses (Friendly, 2006), fitted in Fig. 4 to *E* and *NE* subjects (red and blue, respectively), and in Fig. 5 to *R* and *A* paintings (green and red, respectively).

Figure 4 about here

Finally, Fig. 6 shows the overall mean response on each scale, within *Artwork Type* and subject *Expertise*. That is, for each *s*, the means of *mR*(*i,s*) and *mA*(*i,s*), averaged separately for *E*s and *NE*s; or conversely, the means of *mE*(*p,s*) and *mNE*(*p,s*), averaged separately for *A* and *R* stimuli.

Combined, Figs. 4 and 6 (and Fig. S3) indicate that *E* and *NE* participants gave more positive scores to *R* than to *A* artworks, so the totality of points tend to lie above the diagonals in Fig. 4. In 2 x 2 mixed ANOVA with *Artwork Type* as the within-subjects factor contrasting *mR*(*i,s*) and *mA*(*i,s*) and *Expertise* as the between-subjects factor, this emerged as a significant main effect of the *Artwork Type* for all scales (marginal for SD1) (see Table 2 and Fig. 6).

Table 2 about here

Notably, much of these overall *R/A* differences were specific to the *NE* group. That is, the distribution of *NE*s was *further* above the diagonal in each panel of Fig. 4, while the *E* distribution overlapped it. Measuring the difference between the ellipses, the *Artwork*

Type–Expertise interaction effect reached significance for all scales (marginal for SD2), i.e. the tendency of *R* to elicit more positive responses than *A* paintings was higher for *NE*s. The interaction effect accords with previous findings that artistic training results in higher *Pleasingness* ratings for *Abstract* artworks (Leder et al., 2012; Stojilović & Marković, 2014).

Although individual *E*s varied in their use of a given scale, they tended to be *consistently* biased: if an *E*'s mean rating across *A* paintings lay at one or other extreme of a scale, he or she tended to rate *R* paintings equally high or low, resulting in ellipses that are elongated and aligned with the diagonal (e.g. SD2, *Ugly–Beautiful*). SD4, *Uninformative–Informative*, evoked most variation from *E*s and *NE*s alike, and the largest distributions. At the other extreme, SD5, *Cool–Warm*, showed least variation and the smallest ellipses (despite the elongation of the *E*s' ellipse).

*NE*s were not so even-handed. Their ellipses are typically tilted at a < 45° angle, almost horizontal for SD2 and SD6. This is due to their $mA(i,s)$ means varying across a wider range than their $mR(i,s)$ means: the latter tend to be confined to the upper end of the scale, with the exception of SD5. However, for *R* and *A* paintings combined, neither *E*s nor *NE*s gave more positive scores than the other (which would have separated their ellipses *along* the diagonal). That is, there were no significant main effects of *Expertise* for any scale.

Figure 5 about here

3.3. Semantic differential scales: artwork means, averaged by expertise

Turning to the aggregate scores for each painting within the *E* and *NE* groups, there are consistent displacements between the two ellipses in each panel of Fig. 5. *R* paintings received more positive scores from *NE*s than from *E*s, displacing the corresponding ellipse above the diagonal (see SD1, SD5); conversely, *A* paintings received more positive scores from *E*s than from *NE*s, shifting their ellipse below the diagonal (most obvious for SD4, but also SD2, SD6), or both (SD3). That is to say, the *NE* ratings (*y*-axes) tend to separate *R* from *A* artworks more clearly, with *NE*s rating *R* artworks as more *Interesting*, *Pleasing* and *Sophisticated* than *E*s (*x*-axes). In 2 x 2 mixed ANOVA with *Expertise* as a within-painting measure contrasting *mNE*(*p,s*) and *mE*(*p,s*) and *Artwork Type* as the between-painting factor, the *Artwork Type–Expertise* interaction terms reached significance throughout (see Table 3 and Fig. 6).

The main effect of *Expertise* was only significant for SD2 and SD4, viz. *E*s tended to describe paintings as more *Beautiful* and more *Informative* than *NE*s, displacing the *combined* distribution below the diagonals of those panels, but this was not a general effect.

There were also main effects of *Artwork Type* for SD2, SD3, SD4 and SD6. The combined subjects rated *R* paintings as more *Beautiful*, *Pleasing*, *Informative*, and *Sophisticated* than *A* paintings, separating the ellipses along the diagonal. (See also Fig. S3.)

Finally, the distributions of *A* paintings are more diagonally elongated than their *R* counterparts. That is, *E*s and *NE*s both saw greater diversity among *A* than among *R* paintings in their contents of *Interest* and *Sophistication* in particular, but also of *Beauty*, *Pleasingness*, and *Informativeness*. (With only *NE* subjects, Schepman, Rodway, Pullen, & Kirkham, 2015 also observed a narrower range of responses to *R* stimuli.) The exception here is SD5 (the affective *Cool–Warm* scale), where both ellipses are elongated diagonally, with similar response ranges for both *A* and *R* artworks.

The difference in diversity across the six SDs is combined with a tendency for ellipses to be tilted at an angle above 45°, rotated towards the vertical. This is most notable in the *A* distributions on SD1, SD2, SD6, where mean ratings *mNE*(*p,s*) from *NE*s varied more – ranging down into the *Boring*, *Ugly* and *Naïve* ends of these scales – relative to the corresponding means *mE*(*p,s*) across *E*s. Similar steep angles are apparent for the ellipses and distributions of *R* paintings for SD4, SD2 and SD3: that is, the *NE*s' mean ratings for these paintings varied more than from *E*s, ranging further into the *Informative, Beautiful* and *Pleasing* ends of those scales.

Figure 6 about here

Table 3 about here

In general, then, the *NE* ratings (*y*-axes) tend to separate *R* from *A* paintings more clearly than *E*s (*x*-axes), as if *E*s appraise both artwork types by a similar standard. As corollary, when both painting types are combined, *NE*s tend to spread their appraisals across a wider range than *E*s: for most SD scales, an ellipse around all 24 paintings is steeper than 45°.

3.4 Summary

Hypotheses 1 and 2, about the totality of *E* and *NE* response patterns, were affirmed in qualitative terms by our interpretation of Figs. 4 and 5. Hypothesis 3 was addressed qualitatively, though implicitly, by the ANOVA interaction effects (Tables 2 and 3). We also performed direct *t*-tests. Hypothesis 3 involved *E* and *NE* responses to *A* paintings in isolation, then *NE* responses, in isolation, to *A* and *R* paintings. In terms of the panels of Fig. 4, it predicts that the distribution of *E* respondents will be displaced rightward from their *NE* counterparts; and that the *NE* distribution will be displaced above the diagonal. Indeed, *E* scores on *mA*(*s*) were significantly higher than *NE* scores for SD2 *Ugly–Beautiful* (p = .036) and SD4 *Uninformative–Informative* (p = .005). *NE* scores on *mR*(*s*) were higher than *mA*(*s*) equivalents (p < .001 for all scales except p = .004 for SD1: repeated-measures).

4. Discussion

Stimuli were chosen to exemplify two rival modes or types of painting, *Abstract* vs. *Representational*. In addition, subjects varied in their degree of familiarity with artistic practice, and were categorized as *Expert* or *Non-expert*. Although both polar oppositions simplify a more complex situation, both pairs of groups proved to be valid, despite considerable overlap between the artwork types and between the subject groups.

An initial exploratory phase of Factor Analysis – in “subject mode”– channeled the variance among subjects into their loadings on two factors. Factor-rotated versions of these factors were treated as rival “observer prototypes” for second-order analysis, separately for *Experts* and *Non-experts* (cf. Bimler, 2013). A second FA – in “artwork mode” – focused on the variance among the stimuli, channeling it into their loadings on a different set of factors.

We had anticipated that at least three factors would be necessary to summarize the subjects’ responses on the six SD scales: versions, perhaps, of Hedonic Tone, Uncertainty/Complexity, and Arousal identified by Berlyne (1975). In practice, the collective descriptions that emerged from data analysis were unexpectedly simple. Only two factors could be justified, despite the evidence from comparisons among the panels of Figs. 4 and 5 that no two scales captured precisely the same information. In particular, despite the favorable/unfavorable connotations of many of the SD scales, there was no indication of a ‘hedonic’ quality of personal approval/enjoyment, dissociated from aesthetic appraisal.

It may be that personal preferences did influence the scale ratings, but were too variable across subjects to create a distinct factor. More generally, if some subjects attended to fine shades of meaning among the scales and used them to discriminate separate aspects of the stimuli, they did not do so systematically enough for the distinctions for find their way into the pattern of correlations and emerge from the data. This remained true when the Expert students were analyzed in isolation: their artistic training did not prompt them to single out additional aesthetic-appraisal attributes.

This may also be a consequence of the laboratory setting; subjects might have paid more attention to nuances among the scales in a naturalistic art-gallery encounter. (E.g. Brieber, Nadal, and Leder (2015) noted large differences between laboratory and museum ratings on 'Interest', 'Understanding', etc.) One might say that the art status of the 24 paintings used in this study was irrelevant; primarily they were visual stimuli, decontextualized and presented as computer-monitor reproductions, for aesthetic appraisal in the mundane setting of course-work. This is a far cry from art encounters in the valorized settings of galleries and museums. With immersion in the artworks unlikely to arouse powerful psychological mechanisms, complex models of latent personality variables (such as those considered by Pelowski et al., 2016, 2017) are not required.

Art and Psychology students differed in several respects, validating the labels "Experts" and "Non-experts" that amplify the comparative difference between the groups. *NE*s' responses were more 'one-dimensional', dominated by the *Abstract–Representational* distinction (increasing the VAF of the corresponding principal component), separating them by applying different criteria to the two artwork types (Fig. 2, Fig. S2 bottom). In contrast, *E*s' artwork appraisal features an appreciation of each painting's color palette and its global vibrancy-vs.-dullness (Fig. 2, Fig. S2 top). Mean *E*s' responses rated *Abstract* and *Representational* art on similar terms, and often fitted one or both types into narrower ranges of the scales (Fig. S3, Table 3). *E*s' judgments were more ordered and less dependent on the specific features of each painting, in a sense more coherent.

The ratings on the scales, averaged across stimuli for each observer (Fig. 4) and across observers for each stimulus (Fig. 5), suggest that *Experts* attend more to the formal aspects of the stimuli, whether *Representational* or *Abstract*, hence the higher salience for them of the vertical axis in Fig. 2. As corollaries, the distinction or gradient between the two extremes of painting type (the horizontal axis of Fig. 2) did not dominate their responses as it did for *Non-Experts*, and their ordering of artworks along that gradient is more nuanced. These findings confirm our Hypothesis 1, i.e. the art-training-dependent development of cognitive mastery manifested as cognitive schemata, differing qualitatively in art experts and in laypersons.

At the level of individual responses and individual loadings on the principal components however, despite the even-handedness of their appraisal of *Abstract* vs. *Representational* art, *Experts* varied more among themselves than lay observers, confirming our Hypothesis 2. This outcome is in accord with the notion of "idiosyncratic experts" (Leder et al., 2012) whereby arts training leads to divergence in the use of judgment criteria: the "additional frames of reference" and "multiple schemata" characteristic of experts need not be the same across experts.

The distributions of *Non-experts*' points in Fig. 4 are systematically displaced above the diagonals by the salience for them of the *Representational/Abstract* distinction (interaction effects, Table 2). This is also manifested in the displacement of *A* ellipses below the diagonal in the panels of Fig. 5, and in the tilting of the *A* and *R* ellipses. The implication is that the Art students viewed the *Abstract* paintings as more *Interesting*, *Informative*, *Sophisticated* etc. by employing appropriate cognitive schemata, compared to laypersons, in line with our Hypothesis 3.

In the general context as summarized in the Introduction, expertise or "cognitive mastery" is viewed as the possession of a repertoire of representational schemata and frameworks, in which the expert can assimilate novel stimuli, resolving ambiguity and encoding them in terms of hidden structural simplicity rather than surface complexity. The expertise/type interactions found here support this idea.

Muth et al. (2015, p. 206) quantified the appeal of ambiguity, but took issue with "the frequently reported idea that processing (modern) art simply equals a kind of problem-solving task". They link the aesthetic rewards of an art object not with the solution of its visual puzzle, but rather with the number of insights it provides: a resolution of its ambiguities is not crucial.

-In this formulation, expertise enhances artwork appreciation by providing viewpoints from which these insights are visible.

We envisage this application of cognitive schemata as an iterative process. In the ambiguity / insight paradigm, each solution of a perceptual puzzle or acquisition of an insight leads to reassessment, unpacking the visual meaning of the artwork in an open loop: The explicit classification stage of encoding is followed by cognitive evaluations, in a dynamic process of testing perceptual hypotheses in feedback loops (Augustin et al., 2008; Bachmann & Vipper, 1983; Belke et al., 2006; Nodine et al., 1993).

In this interpretation, aesthetic appraisal is a dynamic process of microgenesis, in which a series of perceptual hypotheses unfold and are tested against the artwork "evidence", conceivably at the stage of 'explicit classification' (Pelowski et al., 2017). The process is an interaction between bottom-up and top-down phenomena, structured in a temporal sequence: an initial sensory-based gist reaction (representational aspects; content appraisal) is followed by a scrutiny of pictorial features (style appraisal) (Augustin, Leder, Hutzler, & Carbon, 2008; Locher, Krupinski, Mello-Thoms, & Nodine, 2007). Initially tentative, diffuse, unstable, emotion-arousing percepts evolve in the direction of redundancy gain, becoming more stable, regular, differentiated and comprehensible (Bachmann & Vipper, 1983; Cupchik & Berlyne, 1979).

Compared to laypersons, perceptogenesis in art experts appears to be a more elaborated multiple-loop process, incorporating knowledge of collative (formal) properties of an artwork, scrutiny of structural relationships among pictorial elements, and interpretation (Nodine et al., 1993). The whole process may be accompanied by emotional detachment (cf. Leder et al., 2004), or by "knowledge emotions" of interest or confusion (Silvia, 2006, 2013).

Acknowledgments

The present study is based on the undergraduate final year research project of Megan Snellock. The authors thank Glen Pennington for developing an experimental program in E-Prime 2.0. Results of this study were partly presented at the *4th Visual Science of Art Conference*, 27th August 2016, Barcelona, Spain. We are also grateful to three anonymous reviewers whose constructive critiques have helped significantly to improve earlier versions of the manuscript.

References

Alluisi, E. A. (1975). Studies in the new experimental aesthetics [review]. *American Journal of Psychology, 88*, 520-523.

Augustin, M. D., Carbon, C.-C., & Wagemans, J. (2011). Measuring aesthetic impressions of visual art. *Perception, 40,* 219 Supplement.

Augustin, M. D., & Leder, H. (2006). Art expertise: A study of concepts and conceptual spaces. *Psychology Science, 48*, 133-156.

Augustin, M. D., Leder, H., Hutzler, F., & Carbon, C.-C. (2008). Style follows content: On the microgenesis of art perception. *Acta Psychologica, 128,* 127-138.

Avital, T., & Cupchik, G.C. (1998). Perceiving hierarchical structures in nonrepresentational paintings. *Empirical Studies of the Arts, 16*, 59-70.

Bachmann, T., & Vipper, K. (1983). Perceptual rating of paintings from different artistic styles as a function of semantic differential scales and exposure time. *Archives of Psychology, 135*, 149-161.

Belke, B., Leder, H., & Augustin, D. (2006). Mastering style. Effects of explicit style-related information, art knowledge and affective state on appreciation of abstract paintings. *Psychology Science, 48*, 115-134.

Berlyne, D. E. (1975). Dimensions of perception of exotic and pre-Renaissance paintings. *Canadian Journal of Psychology, 29,* 151-173.

Berlyne, D. E. (1976). Similarity and preference judgments of Indian and Canadian subjects exposed to Western paintings. *International Journal of Psychology, 11,* 43-55.

Berlyne, D. E., & Ogilvie, J. D. (1974). Dimensions of perception of paintings. In D.E. Berlyne (Ed.), *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation* (pp. 181-226). Washington, DC: Hemisphere.

Biaggio, M. K., & Supplee, K. A. (1983). Dimensions of aesthetic perception. *Journal of Psychology, 114,* 29-35.

Bilotta, J., & Lindauer, M. S. (1980). Artistic and nonartistic backgrounds as determinants of the cognitive response to the arts. *Bulletin of the Psychonomic Society, 15,* 354-356.

Bimler, D. (2013). Two applications of the Points-of-View model to subject variations in sorting data. *Quality and Quantity, 47*, 775-790.

Brieber, D., Nadal, M., & Leder, H. (2015). In the white cube: Museum context enhances the valuation and memory of art. *Acta Psychologica, 154*, 36-42.

Cela-Conde, C. J., Marty, G., Munar, E., Nadal, M, & Burges, L. (2002). The 'style scheme' grounds perception of paintings. *Perceptual & Motor Skills*, *95*, 91-100.

Chatterjee, A., Widick, P., Sternschein, R., Smith, W. B., II, & Bromberger, B. (2010). The assessment of art attributes. *Empirical Studies of the Arts, 28*, 207-222.

Cupchik, G. C. (1974). An experimental investigation of perceptual and stylistic dimensions of paintings suggested by art history. In D.E. Berlyne (Ed.), *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation* (pp. 235-257). Washington, DC: Hemisphere.

Cupchik, G. C., & Berlyne, D. E. (1979). The perception of collative properties in visual stimuli. *Scandinavian Journal of Psychology, 20*, 93-104.

Frechtling, J. A., & Davidson, P. W. (1970). The development of the concept of artistic style: A free classification study. *Psychonomic Science*, *18,* 79-81.

Friendly, M. (2006). Data ellipses, HE plots and reduced-rank displays for multivariate linear models: SAS software and examples. *Journal of Statistical Software, 17(6)*. http://www.jstatsoft.org

Goude, G. (1972). A multidimensional scaling approach to the perception of art: I and II. *Scandinavian Journal of Psychology*, *13*, 258-284.

Greenberg, C. (1960). *Modernist painting*. Forum Lectures. Washington, DC: Voice of America.

Hekkert, P., & van Wieringen, P. C. W. (1996). The impact of level of expertise on the evaluation of original and altered versions of post-impressionistic paintings. *Acta Psychologica, 94*, 117-131.

Kroonenberg, P. M. (1983). *Three-mode principal component analysis: Theory and applications*. Leiden: DWSO Press.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology, 96*, 489-508.

Leder, H., Gerger, G., Brieber, D., & Schwarz, N. (2014). What makes an art expert? Emotion and evaluation in art appreciation. *Cognition and Emotion, 28*, 1137-1147.

Leder, H., Gerger, G., Dressler, S. G., & Schabmann, A. (2012). How art is appreciated. *Psychology of Aesthetics, Creativity, and the Arts, 6*, 2-10.

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode – Developments and challenges in empirical aesthetics. *British Journal of Psychology*, *105*, 443-464.

Lindell, A. K., & Mueller, J. (2011). Can science account for taste? Psychological insights into art appreciation. *Journal of Cognitive Psychology, 23*, 453-475.

Locher, P.J. (2003). An empirical investigation of the visual rightness theory of picture perception. *Acta Psychologica, 114*, 147-164.

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

Marković, S. (2011). Perceptual, semantic and affective dimensions of experience of abstract and representational paintings. *Psihologija*, *44*, 191-210.

Marković, S., & Radonjić, A. (2008). Implicit and explicit features of paintings. *Spatial Vision*, *21*, 229-259.

Mullennix, J. W., & Robinet, J. (2018). Art expertise and the processing of titled abstract art. *Perception, 47*, 359-378.

Muth, C., Hesslinger, V. M., & Carbon, C.-C. (2015). The appeal of challenge in the perception of art: How ambiguity, solvability of ambiguity, and the opportunity for insight affect appreciation. *Psychology of Aesthetics, Creativity, and the Arts, 9,* 206-216.

Neperud, R. W. (1970). Towards a structure of meaning in the visual arts: A three-mode factor analysis of non-art college student responses to selected art forms. *Studies in Art Education*, *12*, 40-49.

Neperud, R. W. (1988). A propositional view of aesthetic experiencing for research and teaching in art education. In F. Farley & R.W. Neperud (Eds.), *The foundations of Aesthetics, art and art education* (pp. 273-319). New York: Praeger.

Neperud, R. W., & Jenkins, H. C. (1982). Ethnic aesthetics: Blacks' and nonblacks' aesthetic perception of paintings by Blacks. *Studies in Art Education*, *23*, 14-21.

Neperud, R. W., & Marschalek, D. G. (1988). Informational and affect bases of aesthetic response. *Leonardo*, *21*, 305-312.

Nodine, C. F, Locher, P. J., & Krupinski, E. A. (1993). The role of formal art training on perception and aesthetic judgement of art compositions. *Leonardo, 26,* 219-227.

O'Hare, D. (1976). Individual differences in perceived similarity and preference for visual art: A multidimensional scaling analysis. *Perception & Psychophysics, 20,* 445-452.

O'Hare, D., & Gordon, I. E. (1977). Dimensions of the perception of art: Verbal scales and similarity judgements. *Scandinavian Journal of Psychology*, *18,* 66-70.

Osgood, C. E., May, W., & Miron, M. (1975). *Cross-cultural universals of affective meaning.* Urbana, IL: University of Illinois Press.

van Paasschen, J., Bacci, F., & Melcher, D. P. (2015). The influence of art expertise and training on emotion and preference ratings for representational and abstract artworks. *PLoS ONE, 10(8):e0134241*. doi: 10.1371/journal.pone.0134241

Parsons, M. J. (1987). *How we understand art: A cognitive developmental account of aesthetic experience*. Cambridge: Cambridge University Press.

Pelowski, M., Markey, P.S., Lauring, J. O., & Leder, H. (2016). Visualizing the impact of art: An update and comparison of current psychological models of art experience. *Frontiers in Human Neuroscience*, *10:160*. doi: 10.3389/fnhum.2016.00160

Pelowski, M., Markey, P.S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish me… delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews, 21*, 80-125.

Pihko, E., Virtanen, A., Saarinen, V.-M., Pannasch, S., Hirvenkari, L., Tossavainen, T., … Hari, R. (2011). Experiencing art: The influence of expertise and painting abstraction level. *Frontiers in Human Neuroscience*, *5:94*. doi: 10.3389/fnhum.2011.00094

Schepman, A., Rodway, P., Pullen, S. J., & Kirkham, J. (2015). Shared liking and association valence for representational art but not abstract art. *Journal of Vision*, *15(5):11*, 1-10. doi: 10.1167/15.5.11

Silvia, P. J. (2006). Artistic training and interest in visual art: Applying the appraisal model of aesthetic emotions. *Empirical Studies of the Arts, 24*, 139-161.

Silvia, P. J. (2013). Interested experts, confused novices: Art expertise and the knowledge emotions. *Empirical Studies of the Arts*, 31, 107-115.

Stojilović, I. Z., & Marković, S. (2014). Evaluation of paintings: Effects of lectures. *Psihologija*, *47*, 415-432.

Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception, 36*, 91-100.

Table 1. Paintings used as stimuli.

Representational		Abstract	
Artist	*Title*. Year	Artist	*Title*. Year
Bellow	*Dempsy and Firpo.* 1924	Braque	*Bottle and fishes.* 1910-1912
Clausen	*A frosty March morning.* 1904	Delaunay	*Champs de Mars: La Tour Rouge*. 1911-1923
Dix	*Metropolis*. 1928	Dove	*Cow.* 1914
Friesz	*Roofs and cathedral in Rouen*. 1908	Jawlensky	*Meditation. (The prayer)*. 1922
Gauguin	*The ford.* 1901	Klee	*Senecio*. 1922
Hopper	*Railroad sunset*. 1929	Leger	*The railway crossing*. 1919
Lempicka	*Portrait of Miss Allan Bott*. 1930	Marc	*Fighting forms*. 1914.
Magritte	*The portrait*. 1935	Matisse (1)	*Le rideau jaune*. 1915
Matisse (2)	*Dishes and fruit*. 1901	Metzinger	*Danseuse au café.* 1912
Petrov-Vodkin	*Portrait of Anna Akhmatova.* 1922	Picabia	*Tarantelle*. 1912
Roerich	*Lord of the Night*. 1918	Picasso	*Bowl of fruit, violin and bottle*. 1914
van Dongen	*The violin player*. 1920	Villon	*Girl at the piano.* 1912

Table 2. Outcomes of the 2 x 2 mixed ANOVA for six SD scales, with mean descriptions per subject across Representational vs. Abstract stimuli [Artwork Type], $mR(i,s)$ and $mA(i,s)$, comparing *Experts* and *Non-experts* [Expertise]; df=(1, 61) throughout. No main effects of Expertise reached significance. * = Marginally significant.

Semantic differential scale	Artwork Type main effect			Artwork Type–Expertise interaction effect		
	F	p	η_p^2	F	p	η_p^2
SD1 *Boring–Interesting*	3.51	0.066*	0.054	6.50	0.013	0.096
SD2 *Ugly–Beautiful*	28.28	< 0.001	0.317	3.73	0.058*	0.058
SD3 *Annoying–Pleasing*	24.56	< 0.001	0.287	4.01	0.050	0.062
SD4 *Uninformative–Informative*	77.24	< 0.001	0.559	12.97	0.001	0.175
SD5 *Cool–Warm*	31.80	< 0.001	0.343	4.09	0.048	0.063
SD6 *Naïve–Sophisticated*	26.14	< 0.001	0.300	4.27	0.043	0.065

Table 3. Outcomes of the 2 x 2 mixed ANOVA for six SD scales, with mean descriptions per painting from *Experts* and *Non-experts* [Expertise]), *mE*(*p*,*s*) and *mNE*(*p*,*s*), comparing *Abstract* vs. *Representational* [Artwork Type]; df=(1, 23) throughout. Only significant effects shown.

Semantic differential scale	Expertise main effect			Artwork Type main effect			Artwork Type–Expertise interaction effect		
	F	p	h_p^2	F	p	h_p^2	F	p	h_p^2
SD1 *Boring–Interesting*							9.16	0.006	0.294
SD2 *Ugly–Beautiful*	9.10	0.006	0.293	5.51	0.028	0.200	6.07	0.022	0.216
SD3 *Annoying–Pleasing*				6.55	0.018	0.229	6.59	0.018	0.230
SD4 *Uninformative–Informative*	6.83	0.016	0.237	19.84	< 0.001	0.474	18.12	< 0.001	0.452
SD5 *Cool–Warm*							4.31	0.050	0.164
SD6 *Naïve–Sophisticated*				6.68	0.017	0.233	5.58	0.027	0.202

Figure legends

Figure 1. Loadings for 63 observers on the first two components of unrotated subject-mode PCA solution with painting-scale combinations as rows (PC1 = "Consensus", PC2 = "Distinction"). Points colored to distinguish *Experts* (Art students;•) and *Non-experts* (Psychology students; •).

Figure 2. Scores for 24 paintings in second-order PCA of "Consensus" components from separate PCA of *Expert*-only and *Non-expert*-only data ($PC1^{(E)}$, $PC1^{(NE)}$). Lines link the two points representing each painting, for *E* and *NE* appraisal; artist's name identifies the *NE* version. Points colored to distinguish *R* (•) and *A* (•) paintings.

Figure 3. Loadings for 24 paintings on the first two components of rotated artwork-mode PCA solution with subject-scale combinations as rows (PC1 = "Representational", PC2 = "Abstract"). PC3 ("Color") is orthogonal. Points colored to distinguish *R* (•) and *A* (•) paintings.

Figure 4. Scatter-plots of mean scores per subject on each of six SD scales, averaged across *Abstract* vs. *Representational* stimuli, $mA(i,s)$ and $mR(i,s)$ (horizontal and vertical axes). Points colored to distinguish *E* (•) and *NE* subjects (•), with ellipses enclosing separate distributions.

Figure 5. Scatter-plots of mean scores per painting on each of six SD scales, averaged across Experts vs. Non-Experts, $mE(p,s)$ and $mNE(p,s)$ (horizontal and vertical axes). Points colored to distinguish *R* (•) and *A* (•) paintings, with ellipses enclosing separate distributions.

Figure 6. Mean scores on six SD scales within *Artwork Type* and art *Expertise*, showing interaction. Error bars show standard errors.

Supplementary Material

Figure S1. Scores for 24 paintings in second-order PCA of the “Consensus” component ($PC1^{(C)}$) from unrotated subject-mode PCA of combined *Expert* and *Non-expert* data. ● Representational paintings; ● Abstract paintings.

Figure S2. Scores for 24 paintings in second-order PCA of *Expert* (top) and *Non-expert* (bottom) exaggerated, prototype components from unrotated subject-mode PCA of combined data. Identification of components only tentative. ● Representational paintings; ● Abstract paintings.

Figure S3. Bar graphs showing mean scores for each painting on six SD scales, separately across *Experts* (Art students; ■) and *Non-experts* (Psychology students; ■). Paintings arranged in order of ascending means, separately for Representational and Abstract artworks. Error bars show 95% CIs.