

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Psychological Aesthetic Effects of Lines and Colors:
An Interdisciplinary Inquiry”

verfasst von / submitted by

Eva Specker, Msc. Bsc.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktorin der Philosophie (Dr.phil.)

Wien, 2019 / Vienna, 2019

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on the
student record sheet:

A 796 310 298

Dissertationsgebiet lt. Studienblatt /
field of study as it appears on the student
record sheet:

Psychologie

Betreut von / Supervisor:

Univ. Prof. Dr. Helmut Leder

Univ. Prof. Dr. Raphael Rosenberg

Table of Contents

1. Introduction	p. 5 – 21
2. Published Work	p. 22 – 330
<i>2.1 Basic Psychological Research</i>	
2.1.1	p. 22 – 36
Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H. (2018). The Vienna Art Interest and Art Knowledge (VAIAK) Questionnaire: A Unified and Validated Measure of Art Interest and Art Knowledge. <i>Psychology of Aesthetics, Creativity, and the Arts</i> , http://dx.doi.org/10.1037/aca0000205	
2.1.2	p. 37 – 44
Specker, E., Leder, H., Rosenberg, R., Hegelmaier, L. M., Brinkmann, H., Mikuni, J., & Kawabata, H. (2018). The universal and automatic association between brightness and positivity. <i>Acta Psychologica</i> , 186, 47–53. https://doi.org/10.1016/j.actpsy.2018.04.007	
2.1.3	p. 45 – 53
Specker, E., & Leder, H. (2018). Looking on the Bright Side: Replicating the Association between Brightness and Positivity. <i>Collabra: Psychology</i> , 4(1), 34, 1-8. http://doi.org/10.1525/collabra.168	
<i>2.2. Implementation in the Lines & Colors Project</i>	
<u><i>Psychological</i></u>	
2.2.1	p. 54 – 78
Specker, E., Forster, M., Brinkmann, H., Boddy, J., Immelmann, B., Goller, J., Pelowski, M., Rosenberg, R., & Leder, H. (under review). Is Red Warm for Everyone? Assessing the agreement on aesthetic effects of artworks. <i>Frontiers</i> . Submitted: 22.08.2018	
2.2.2	p. 79 – 119
Specker, E., Fried, E. I., Rosenberg, R., & Leder, H. (under review). Sensual Yellows and Moving Lines: Proposing the aesthetic effect network (AEN). <i>Perspectives on Psychological Science</i> . Submitted: 12.02.2019	
<u><i>Art Historical</i></u>	
2.2.3	p. 120 – 144
Brinkmann, H., Boddy, J., Immelmann, B., Specker, E., Pelowski, M., Leder, H., & Rosenberg, R. (2018). Ferocious Colors and Peaceful Lines Describing and Measuring Aesthetic Effects. <i>Wiener Jahrbuch Für Kunstgeschichte</i> , 65, 2-26.	

2.2.4 p. 145 - 165
 Boddy, J., Brinkmann, H., Forster, M., **Specker, E.**, Leder, H., & Rosenberg, R. (manuscript in preparation). Universality, sensitivity and measurability in discourses on the aesthetic effects of lines and colors: Knowledge transfer between art and psychology.

2.3 Field Research

Articles

2.3.1 p. 166-188
 Pelowski, M., Leder, H., Mitschke, V., **Specker, E.**, Gerger, G., Tinio, P. P. L., ... Husslein-Arco, A. (2018). Capturing Aesthetic Experiences With Installation Art: An Empirical Assessment of Emotion, Evaluations, and Mobile Eye Tracking in Olafur Eliasson's "Baroque, Baroque!" *Frontiers in Psychology*, 9, 1255, 1-22.
<https://doi.org/10.3389/fpsyg.2018.01255>

2.3.2 p. 189-209
 Pelowski, M., **Specker, E.**, Gerger, G., Leder, H., & Weingarden, L. (2018). Do you feel like I do? A study of spontaneous and deliberate emotion sharing and understanding between artists and perceivers of Installation Art, *Psychology of Aesthetics, Creativity, and the Arts*. <http://dx.doi.org/10.1037/aca0000201>

2.3.3 p. 210 – 225
 Grüner, S., **Specker, E.**, & Leder, H. (2019). Effects of Context and Genuineness in the Experience of Art. *Empirical Studies of the Arts*
<https://doi.org/10.1177/0276237418822896>

2.3.4 p. 226 – 255
Specker, E., Stamkou, E., Pelowski, M., & Leder, H. (under review). Radically revolutionary or pretty flowers? An experimental museum study of the impact of curatorial narrative highlighting artistic deviance on the visitor's assessment of artist influence. *Curator*. Submitted: 05.04.2019

Related Book Chapters

2.3.5. p. 256 - 300
 Pelowski, M., & **Specker, E.** (in press). The general impact of context on Aesthetic Experience. In M. Nadal, & O. Vartanian (Eds.), *The Oxford Handbook of Empirical Aesthetics*.

2.3.6 p. 301 – 330
 Tinio, P. P. L., & **Specker, E.** (in press). The method of observation. In M. Nadal, & O. Vartanian (Eds.), *The Oxford Handbook of Empirical Aesthetics*.

- 3. Discussion p. 331-335
- 4. English Abstract p. 335
- 5. German Abstract p. 335-336

1. Introduction

During my PhD I was employed within the interdisciplinary project “*Universal aesthetics of lines and colors?*” that constituted a collaboration between psychologists (EVA-lab) and art historians (CreA-Lab) within in the University of Vienna. The association with these two labs also gave me the opportunity to participate in studies that tracked aesthetic experiences in museums. Consequently, my PhD dissertation (“*Psychological Aesthetic Effects of Lines and Colors: An Interdisciplinary Inquiry*”) can be divided up in three main sections: (1) Basic Psychological Research, (2) Implementation within the project, (3) Field Studies. As such this introduction to my published work will be structured accordingly. In each section, I will give an overview of my work and address how the individual papers relate to each other both within and across sections. To provide the appropriate context to the dissertation I will first provide an introduction to the concept of aesthetic effects.

Aesthetic Effects

In Goethe’s seminal treatise about colors, he wrote: “The colours on the plus side are yellow, red-yellow (orange), yellow-red (minium, cinnabar). The feelings they excite are quick, lively, aspiring” (Eastlake, 1840, p.306). This statement was part of a chapter in which Goethe examined the aesthetic effects of colors on perceivers, in which he also stated that “red-yellow gives an impression of warmth and gladness” (Eastlake, 1840, p.309). Because of such effects, he considered color a central aspect of visual art: “colour considered as an element of art, may be made subservient to the highest aesthetical ends” (Eastlake, 1840, p.304). In the art historical literature, attributions such as “warmth” are known as *aesthetic effects*. Aesthetic effects are seen as specific responses in viewers of art that are caused by aesthetic forms, especially lines and colors (Brinkmann et al., 2018). Thus, in art historical discourse, an artwork is seen as warm because it has warm colors in it. The idea that simple visual elements such as colors (but also lines, see e.g. Rosenberg, 2010) elicit specific effects

in humans is one of the foundations of modern discourses on art since the 18th century and is still crucial for art historians today (Brinkmann et al., 2018). Under this view, if an artwork evokes an impression of warmth, this impression can be explained by the occurrence of red tones in the artwork, and vice versa. Specifically, it assumes a cause and effect relationship between aesthetic forms and perceptions. Most commonly, aesthetic effects are related to colors and lines, since as early as the middle of the 16th century Venetian and Florentine art literature theorized that *disegno* (drawing) and *colore* (color) were seen as the central elements of painting.

Theorists describing the link between colors and forms, and their effects on humans generally conceive these effects as objective (aesthetic) properties of the stimuli. Aesthetic properties (or “aesthetic effects”) such as warmth are often treated as essentially “perceptual” (Robson, 2014), that is: we can “see” that a line is dynamic in the same way as that we can see that it is curved. From a psychological perspective, the terminology of an aesthetic effect, let alone of this effect being perceptual, is strange. In psychology we would rather speak of conceptual associations between separate concepts (i.e. the concept “red” and the concept “warmth”), which lead to an aesthetic experience (Leder, Belke, Oeberst, & Augustin, 2004). Flowing from this, many theorists claim universality of aesthetic effects either explicitly like Endell (1896) who claimed that every color and every form cause the same feeling in all people, or implicitly by assuming that people in general can see that a line is dynamic. To give a few examples, Franz Marc described colors in a letter to August Macke as follows: “Blue is the male principle, astringent and spiritual. Yellow is the female principle, gentle, gay and sensual. Red is matter, brutal and heavy and always the color to be opposed and overcome by the other two” (Macke, 1964, p. 28). Similar descriptions can be found for lines. Kandinsky, for example, postulated: “The same thing can with complete justification be said about the horizontal and the vertical - low and high. The former is lying, the latter is standing,

walking, moving about, finally climbing upward. Supporting - growing. Passive - active. Relatively: feminine – masculine” (Kandinsky, 1947, p. 63).

Interestingly, rather than strict scientific research, a major body of previous empirical work on aesthetic effects consists of artistic research, in which artists experimented with their art in an attempt to create one or the other effect. To assess this universality from a scientific point of view was therefore one of the basic premises of our interdisciplinary project and the starting point of my experimental work. In addition, the art historical work done mainly by my colleagues in our joint project, further examined both the general history and literature on aesthetic effects as well as specifically investigating the history of empirical work done on aesthetic effects.

1.1. Basic Psychological Research

Even though aesthetic effects are treated as essentially “perceptual” (Robson, 2014), it may be the case that some people are more “sensitive” observers than others. Hume (1965), for example, argued that a true judge has “a strong sense, united to delicate sentiment, improved by practice, perfected by comparison, and cleared of all prejudice”(p.17). Similarly, Sibley (1959) argues that aesthetic perception “requires the exercise of taste, perceptiveness, or sensitivity” (p.421). Though it has been argued that (very) few people would meet or come close to meeting Hume’s standards (Hume, 1965; Meskin, 2007; Robson, 2014) or perhaps those of Sibley, we felt that a true test of universality should at least attempt to take into account the notion that some observers are more sensitive than others. To this aim we wanted to include “art expertise” within our research.

However, there was no good and validated way to measure “art expertise” nor a good theoretical concept of what “expertise” was which lead to the need to develop such a measure ourselves. As we argue in the paper (Specker et al., 2018; section 2.1.1), “art expertise” is a relatively vague concept lacking a clear definition in the literature. Consequently, art interest

and art knowledge are often confounded. We argue that these are two separate concepts and should be assessed as such. Furthermore, we argue that art interest and art knowledge are the foremost dimensions of art expertise based on the fact that most measurement methods of expertise roughly follow the approach taken by Berlyne to assess interest and knowledge via self-reported behavior and education. Nonetheless, even though they follow this approach interest and knowledge are often not separated empirically. For example, one of the most widely used questionnaires by Chatterjee, Widick, Sternschein, Smith, and Bromberger (2010, for another example see Chamorro-Premuzic & Furnham, 2005) combines both items measuring art interest (e.g., how often people visit museums) as well as art knowledge (e.g., asking about the amount of art education people had). Another measurement issue is that criteria people use to measure art expertise differ widely, and they are rarely based on a theory of what art expertise is. To give a few examples, being an expert might mean any of the following: a (bachelor) art history student (Leder et al., 2014; Rosenberg & Klein, 2015), a fine art student (Chamberlain & Wagemans, 2015; Winston & Cupchick, 1992), someone who completed a year of formal training (Belke, Leder, Harsanyi, & Carbon, 2010), a professional artist (Winston & Cupchick, 1992), and people employed in an art institution with jobs ranging from a gallery director or curator to a critic (Hekkert & Van Wieringen, 1996). Alternatively, researchers make up impromptu questionnaires, sometimes using only one item. For example, Locher, Smith, and Smith (1999, 2001) asked: “Have you had training in visual arts?” rated from 1 (none) to 7 (art major). Similarly, interest may be measured with many different metrics. Whereas some (Chatterjee et al., 2010) ask for concrete behavior such as amount of museum visits, others (Pelowski, 2015) use subjective scales (“I am comfortable looking at and discussing art”), and many use a combination of both (e.g., Chamorro-Premuzic & Furnham, 2005; Laming et al., 2016). Different expert groups and different ways to measure their expertise can lead to misinterpretations of existing data and obscure interindividual differences.

As such, we developed the VAIK (Specker et al., 2018; section 2.1.1) using a three-stage validation process. We started from data (N= 740) gathered with an initial version of an “art expertise” questionnaire that had been used within the lab for a number of years. The reevaluation of this questionnaire revealed some clear points for improvement. This led us to modify the art interest scale and re-conceptualize the art knowledge scale. The final version of the art interest scale comprised both items measuring subjective interest as well as items measuring behavior (such going to the museum). The re-conceptualization of the art knowledge scale was done in close collaboration with the art history department to consider what different skills comprise knowledge of art. We identified three skills: the ability to classify artworks (both in terms of styles and artists), the ability to recognize iconography, and the ability to recognize production techniques. These then became the foundation for our knowledge scale of the VAIK. We then conducted a pre-test in which we assessed the questionnaire qualitatively and quantitatively with a sample of lay people as well as art history students and art professionals. Finally, we conducted a large-scale validation study (>600 participants) with both lay people (psychology students) and art history students, where we present evidence regarding relations to other variables and internal structure as well as the ability to discriminate between relative lay people and experts. We provided evidence for reliability of the VAIK scores as well as evidence for the validity as a measure of both art interest and art knowledge. With this new questionnaire, we offer a tool for researchers in empirical aesthetics, creativity, and also applied fields to quantify art interest and art knowledge dimensions.

Notably, this also allowed us to assess art interest and art knowledge in an adequate way throughout the rest of our research. Furthermore, we also approached this topic from an art historical perspective in later work (Boddy et al., *in prep*; section 2.2.4). A summary of this paper is given in section 2.2.4.

As noted, from a psychological perspective aesthetic effects can be seen as conceptual associations between separate concepts (i.e. the concept “red” and the concept “warmth”). To understand how colors can be related to aesthetic concepts we conducted experimental work on the association between brightness and positivity (Specker et al., 2018; Specker & Leder, 2018; section 2.1.2 and 2.1.3 respectively). This particular association is also relevant within an art historical context. For example, the art historian Gombrich (1963) wrote about the story of Theseus and the Minotaur: “Surely it was no accident that black rather than white was the sign agreed upon for failure. Black seems to us a more ‘natural’ sign for grief, and white for a ‘brighter’ mood: and even though we know that cultural conventions also play their part (and that black is not everywhere the color of mourning), the correlation makes sense in terms of expressiveness (p.61).” Interestingly, he also hints at the role that culture may play within such associations. In Specker et al. (2018) we note that though there is considerable evidence of the influence of culture on color associations connected to different hues such as the influence of language on color discrimination (Regier & Kay, 2009; Winawer et al., 2007), and the cultural differences in affective/symbolic meaning and associations related to different hues (Adams & Osgood, 1973; Saito, 1996), there seems to be less (cross-cultural) variation in the influence of brightness. Brightness seems to be associated with positivity across cultures (Adams & Osgood, 1973; Saito, 1996; Valdez & Mehrabian, 1994). We note that a remaining question is why effects of brightness seem to be consistent even across cultures when effects of hue seem more likely to vary. A way to understand this would be through the structure of our visual system. Our visual system employs two pathways, a chromatic one that processes color contrasts and an achromatic one that processes luminance contrasts (for a more detailed account see Gegenfurtner & Kiper, 2003). While the development of the achromatic response appears to be stable in infants after a few months of age (Crognale, 2002) the development of the chromatic response continues to change from birth until puberty (Boon, Suttle, & Dain, 2007). Furthermore, Page and Crognale (2005) suggest that both

pathways age differently, with the chromatic pathways being more significantly affected by age-related changes. These studies indicate that the chromatic system may be more sensitive to cultural influence than the achromatic system. Following this reasoning, we hypothesized that the effects of brightness should be universal as well as automatic. If brightness triggers an evolutionarily earlier developed system then consistent effects of brightness should be found across cultures and these effects should be automatic. In a series of three studies we investigated this in Austria and in Japan (Study 3) by use of an IAT design. We then did a pre-registered replication of Study 1 (chromatic stimuli) and 2 (achromatic stimuli) from Specker et al. (2018) in Specker and Leder (2018) which also included a meta-analysis of both papers illustrated in Figure 1.

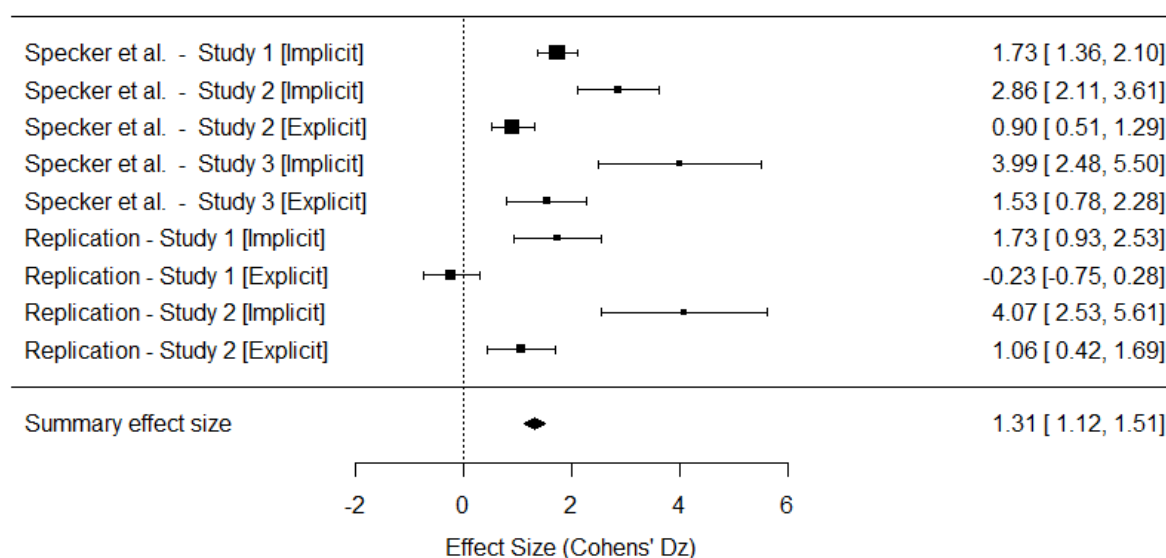


Figure 1. Forest plot of the Fixed-Effects Model. All effect sizes within the plot are Cohen's Dz values. All confidence intervals represent 95% confidence intervals. The size of the boxes represents the sample size (N) of the study.

With these studies, we provide further evidence that this association is present. Second, we provide evidence for universality of this association in Specker et al. (2018) by comparative research with a Japanese sample (Study 3). The successful pre-registered replication of Specker and Leder (2018) further supports these claims in combination with the meta-analysis that indicated a large effect 1.31 [1.12, 1.51]. In addition, we used meta-

analysis to assess potential moderators of the effect, in particular stimulus type (chromatic vs. achromatic) and measure type (implicit vs. explicit). This indicated that the effect is stronger when measured implicitly than when measured explicitly and that the effect is stronger when achromatic stimuli are used.

This research became essential to our forthcoming paper Specker, Fried, Rosenberg, and Leder (*under review*; section 2.2.2.). Within this paper, the results obtained from this previous work could be used to establish a suitable manipulation to test our statistical models (see section 1.2 and 2.2.2. for further detail).

1.2. Implementation within the Lines & Colors Project

The basic core of the project was to assess the universality of aesthetic effects. As such, the paper by Specker et al. (*under review*; section 2.2.1) can be seen as the core publication within the project. However, in order to do this empirical work it was needed to provide the necessary art historical background which was achieved by Brinkmann et al. (2018, section 2.2.3). This paper provided not only a theoretical basis on which we could base our choice of terms (i.e. aesthetic effects) to be used as dependent variables but also preliminary experimental research on which we could base our question format (bipolar scale). A summary of this paper is given in section 2.2.3.

To test the universality of aesthetic effects we used novel statistical methods derived from generalizability theory (Brennan, 2001; Shavelson & Webb, 1991) to assess agreement. As noted also in Boddy et al (*in prep*, section 2.2.4) due to e.g. measurement error: “perfect agreement, so to speak, does not exist in experimental testing”. This creates the need to estimate how much people agree on the aesthetic effects of lines and color. The method we chose was the beholder index (“bi”) by Hönekopp (2006). Based on comparisons of variance components, this index allows to quantify the relative proportion of shared to individual variance in evaluations. Following its name, the beholder index reflects how much of the

variance is due to private evaluation. For example, if the b_i is .75, 75% of the variance is due to private evaluation and by definition 25% is due to shared evaluation. Therefore, to support the notion that aesthetic effects are universally shared, we would need to find low beholder indices where less than half of the variance is due to private evaluation, meaning that more than half of the variance is due to shared evaluation. In addition, to have the results be relevant for art historians we used real artworks and their constituting elements. To trace back shared aesthetic effects, elicited by artworks, to shared aesthetic effects, elicited by their constituting elements, we tested the associations elicited by 1) the artwork as a whole, 2) the combination of all lines of the image, 3) the combination of all colors of the image, 4) the single lines of the image, and 5) the single colors of the image. We manipulated images digitally to separate these elements from the artwork as a whole. This is illustrated in Figure 2. This allowed us to investigate the single elements in isolation while retaining stimulus material derived from real artworks. As noted above, a true test of universality should at least attempt to take into account the notion that some observers are more sensitive than others. We supposed that people who have a higher knowledge of art and interest in art may represent a more sensitive viewer. If these viewers are more sensitive perceivers, and there is a “true” effect, then they should agree more with each other on the aesthetic effects of an artwork than a group of less sensitive viewers. Therefore, all participants also filled out the VAIK (Specker et al., 2018; section 2.1.1).

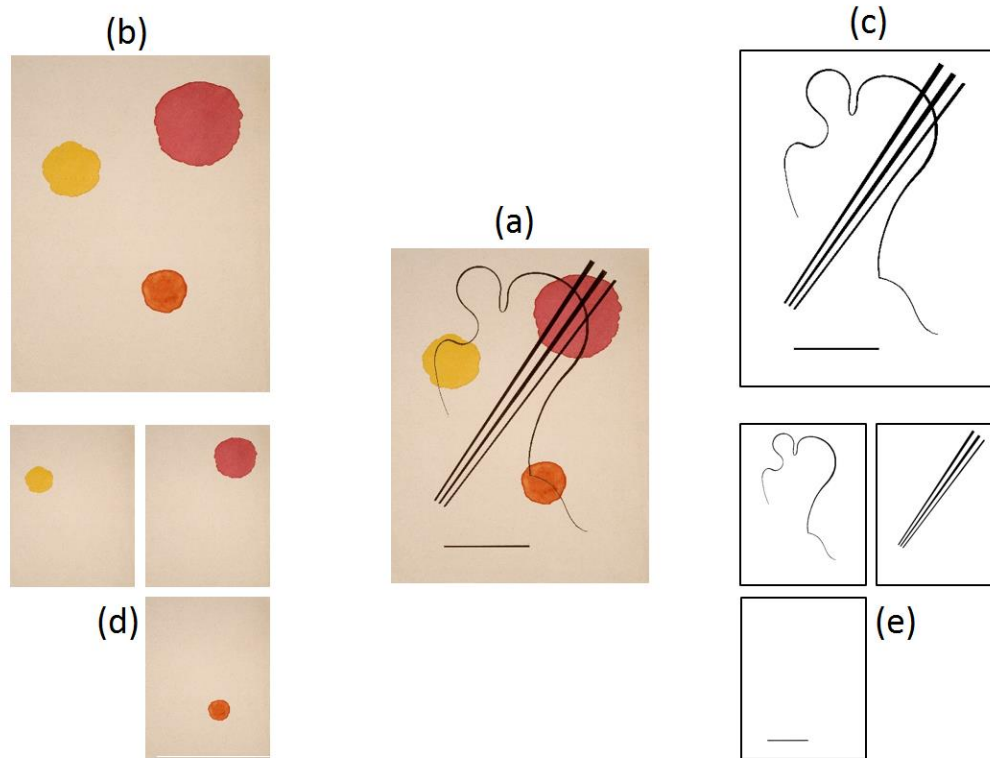


Figure 2. Separation from lines and colors from the artwork Wassily Kandinsky, *Untitled*, 1934, watercolor, ink, 31.6 × 24.6 cm, Paris, Centre Georges Pompidou. **(A)** Whole artwork, **(B)** the combination of colors, **(C)** combination of lines, **(D)** single colors and, **(E)** single lines.

In this research (one original study, N=100, and a replication, N= 50) we found significantly lower agreement in the effects of artworks and their single elements than expected. For the whole artworks, participants agreed on the effects of warm-cold, heavy-light, and happy-sad, but not on the 11 other scales. Further, we found that the image type (artwork or its constituting elements) was a major factor influencing agreement. Art expertise did not play a significant role and agreement was low on dimensions usually of interest in empirical aesthetics (e.g., like-dislike). Our results pose a challenge to the practice of interpreting artworks based on their aesthetic effects, as these effects may not be as universal as previously thought.

These results are addressed from an art historical perspective in the forthcoming (Boddy et al., *in prep*; section 2.2.4) which, as noted, also investigates the notion of expertise

in or sensitivity to art from a historical perspective in correspondence to Specker et al. (2018; section 2.1.1) which provides a psychological and measurement perspective.

Specker, Fried, Rosenberg, and Leder (*under review*, section 2.2.2.) presents an extension on theory of aesthetic effects from a methodological perspective. It takes aesthetic effects and the experience of them as a specific cognitive process present in aesthetic experience. It proposes an aesthetic effect network (AEN) where the experience of aesthetic effects is seen as an associative process where having one association leads to the next association. Through a network of associations, this generates an overall aesthetic experience.

Previous work typically views relationships between aesthetic effects as due to a common cause (e.g., Takahashi, 1985). That is, if active and lively are correlated this is because they are both caused by the latent factor of “Activity” (since most are based on the Evaluation, Potency, and Activity model (EPA model) of Osgood et al. 1957). In our view, relationships between aesthetic effects should be seen as direct associations between one aesthetic effect and another. We propose that aesthetic effects can directly influence each other; for example, if a viewer sees an artwork as active, this will lead the viewer to also perceive it higher in liveliness. We conceptualize these effects as an associative process that proposes direct relationships between aesthetic effects using a network approach. The difference between these perspectives is illustrated in Figure 3.

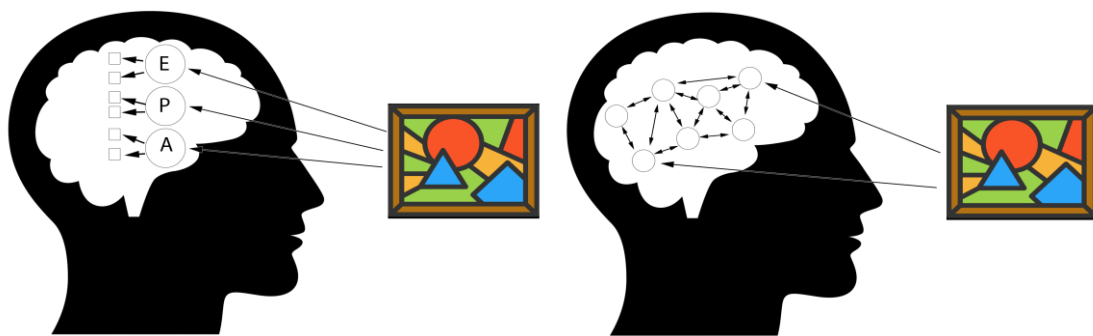


Figure 3. Conceptual model of how an artwork can influence a person in the EPA model (left) and a network model (right).

In this paper we first did an exploratory empirical study (Study 1, N=255) that explored what can be gained from this approach. This allowed us to formulate a preliminary account of the general structure of the relationships between the different aesthetic effects within the AEN. A limitation to Study 1 is that theoretical common cause models (like the EPA model) and network models can give rise to the same correlations among items (Fried & Cramer, 2017; Kruis & Maris, 2016; Van Der Maas et al., 2006). It is not possible to figure out which true model underlies the data based on estimating models from merely observing participants, but it is possible to distinguish between the models by experimental manipulation. This thus led to Study 2 (N=133) where we manipulated the brightness of artworks as a means to activate the node of positivity. From our AEN model, this manipulation should lead the node of positivity to be activated, and from this node the activation can then spread throughout the network. In contrast, this should not happen if the data are generated by a common cause model, due to the causal direction the model proposes: the underlying factor of Evaluation is the cause and positivity is the effect. Thus changing the effect should not change the cause, much like with illnesses such as measles: treating the symptoms may alleviate suffering but does not cure the disease. Here we utilized our own previous work (Specker et al., 2018; Specker & Leder, 2018; section 2.1.2 and 2.1.3 respectively) to be able to create a manipulation—overall brightness of images—that allowed us to differentiate between our proposed AEN model and other models (such as a common cause model).

Study 2 provided empirical support in favor of the AEN model over other potential alternatives. Which strengthened our newly proposed AEN model. More importantly, the AEN model presents a new perspective that changes the way we, at least formally, conceptualize aesthetic effects and to a larger degree aesthetic experience. Conceptualizing aesthetic experience (at least in part) as an associative process provides an explicit account of a specific cognitive process involved in aesthetic experience, that extends beyond research on

aesthetic effects, that can be applied to aesthetic experience in general, and that provides a basis for future studies. This facilitates our understanding of why aesthetic experience is both somehow different and also somehow the same for all of us and is thus intimately paired with the differences found in experience in Specker et al. (*under review*; section 2.2.1).

Finally, as noted it can be applied to research in aesthetic experience in general including field research. As such, I have already conducted some exploratory work with application of networks of aesthetic experience within a field setting (Pelowski et al., 2018; section 2.3.1). Since the application of networks to aesthetics is in its infancy, the Specker, Fried, Rosenberg, and Leder (*under review*, section 2.2.2) paper can provide a strong theoretical basis to guide future studies.

2.3 Field Research

To contextualize the individual papers I will briefly summarize the main theoretical frameworks of each empirical paper and then address how they fit within the general framework of my thesis. In addition, I have been involved in writing two book chapters, one on the general influence of context on aesthetic experience and one from a methodological perspective on how to use observation methods within a museum context (Pelowski & Specker, *in press*; Tinio & Specker, *in press*; sections 2.3.5 and 2.3.6 respectively). Both the empirical papers as well as the book chapters allowed me to continue previous work (Specker, Tinio, & Van Elk, 2017; Van Elk et al., 2016). Each book chapter is accompanied by a summary in section 2 since they do not include an abstract. It is particularly worth noting that since Pelowski and Specker (*in press*) takes a wider perspective on context and also considers context in the form of learning or education it corresponds to Specker et al. (2018, section 2.1.1.) and through its inclusion of a historical perspective also connects to Boddy et al. (*in prep*, section 2.2.4). In addition, Tinio and Specker (*in press*) is based both on own previous

work (Specker, Tinio, & Van Elk, 2017) as well as work done within the thesis notably Pelowski et al. (2018; section 2.3.1).

Pelowski et al. (2018; section 2.3.1) represents a first exploratory study of installation art within a museum setting with the aim to capture the eclectic, the embodied, and the emotionally-charged viewing experience. These were assessed by pre- and post-viewing questionnaires focusing on emotion, meaning-making, and appraisals, in tandem with mobile eye tracking to consider viewers' attention to both installed artworks and/or to the museum environment. People's experiences were assessed with regard to two pieces from the Olafur Eliasson's exhibition "Baroque, Baroque" in the Belvedere Museum in Vienna. Results showed how viewers' appraisal strategies, meaning making, and physical actions facilitated relatively more or less deep engagement with, and enjoyment of, the art. With respect to the exploratory aesthetic emotion networks analyzed by me, we found that across artworks positive and negative emotions tended to cluster together but on opposite sides of the network from each other and with negative connections to each other implying a low likelihood of experiencing mixed emotional states. At the same time, mediating these groups were more cognitive responses—confusion, insight, and especially self-awareness. We also found differences in network structure across artworks, suggesting that the relationship between different emotions is partly influenced by which artwork people are looking at.

In Pelowski, Specker, Gerger, Leder, and Weingarden (2018; section 2.3.2) the main aim was to see if interpersonal communication can happen through artworks. Research in the area of empathy or emotion contagion have routinely documented our ability to respond to others' affective states, often via faces or bodies. This can occur on an intellectual level of perceiving emotion signs (or "cognitive empathy") or can involve actually feeling how others appear to be feeling or mirroring responses with our own brains and physiology (or "emotional empathy"). Here, we introduce a paradigm wherein we assessed working artists'

emotional experiences and intentions as they produced installation artworks and then assessed resulting emotions and understanding of viewers. We find that in most cases (2 of 3 artworks) viewers did consistently feel more emotions that the artist had intended, viewers also spontaneously reported feeling similar emotion patterns while viewing as those reported by the artists while making, regardless of whether these had been intended. With all 3 artworks, viewers showed ability to guess the emotion-transmission intentions of artists. Success at emotion sharing and at feeling intended emotions also positively related to ratings (i.e., goodness) of the art.

In Grüner, Specker, and Leder (2019; section 2.3.3) we studied the effects of genuineness (real artwork vs. a reproduction) and context (museum vs. laboratory) on art experience. Previous studies (e.g., Brieber, Nadal, & Leder, 2015; Locher, Smith, & Smith, 1991, 2001; Specker, Tinio, & Van Elk, 2017) found enhanced aesthetic experience and higher ratings in aesthetic value, liking, and interest in museums as compared with laboratories. In addition, people look longer at artworks in a museum context than in a laboratory context (Brieber, Nadal, Leder & Rosenberg, 2014). A current issue in the field is that in these studies context and genuineness are confounded as identified by Brieber, Leder, and Nadal (2015). Brieber, Leder, and Nadal (2015) consequently presents the first empirical study where these factors are systematically differentiated, however, they did not find any effects of context or genuineness. To substantiate these results, we setup a conceptual replication by using the same 2 (museum/laboratory) by 2 (genuine/reproduction) between-subjects design with improved manipulations of context and genuineness. We found an effect of context, as artworks presented in a museum were liked more and rated more interesting than in the laboratory. We did not find effects of genuineness. Exploratively, we found that art style had a big impact on how artworks were rated regardless of context and genuineness, indicating that this may be a more important factor for aesthetic experience.

In Specker, Stamkou, Pelowski, and Leder (*under review*; section 2.3.4) we translate a lab-based paradigm into a museum context. Previous work in the lab (Stamkou et al. 2018) showed that artists are seen as more influential if they are seen as more deviant. In the lab participants were shown a series of artworks, the last artwork being the focal artwork. The focal artwork either deviated from the artworks (in style: figurative/abstract) before or it was the same. For example, participants saw three figurative artworks and then a focal abstract artwork (deviance condition). We saw these findings as potentially highly relevant for curators, though found in a lab-based paradigm and focused on testing general principles in social psychology. This is because curatorial narratives—especially in temporary exhibitions focused on one artist—often include deviance. For example, temporary exhibitions that focus on one artist—or one man shows—classically follow the same curatorial narrative: this narrative tends to be historical, one starts with the early career of the artist and so follows the artist throughout his or her lifetime. This constitutes interindividual deviance: we see an artist’s style change in comparison to their own earlier style. The findings of Stamkou et al. (2018) would predict that these curatorial narratives would lead visitors to see the artist as more influential and thus provide a basis for predicting how curatorial narratives influence visitor experience. We tested this in the Albertina Museum in Vienna: we used an artwork placed within a temporary exhibition on Claude Monet—that showed his stylistic progression throughout his life—as a real-life equivalent of the deviance condition from the lab, and the same artwork placed within a permanent exhibition—that showed the artwork amongst other examples of impressionism—as a real-life equivalent of a no deviance lab condition. We hypothesized that the artist is seen as more influential in the temporary exhibition (interpersonal deviance) than in the permanent exhibition (no deviance). Our results confirm this hypothesis. These findings support the notion that curatorial narratives influence visitor experiences. Specifically, it can explain how curatorial narratives influence if and when artists are seen as influential and offers guidelines for evidence-based curatorial practices.

Though not a main focus of the project, conducting research from an interdisciplinary point of view brings the need to assess people in ecologically valid settings to understand their aesthetic experiences. As noted, Pelowski et al. (2018; section 2.3.1) includes an exploratory network analysis which is related to my theoretical work in Specker, Fried, Rosenberg, and Leder (*under review*; section 2.2.2.). Both Grüner, Specker, and Leder (2019; section 2.3.3) and Pelowski, Specker, Gerger, Leder, and Weingarden (2018; section 2.3.2) allowed for an extension of previous work. In addition, Grüner, Specker, and Leder (2019) allowed for the testing of the VAIK (Specker et al., 2018; section 2.1.1) within a field setting (the same goes for Specker, Stamkou, Pelowski, & Leder, *under review*; section 2.3.4). As such, both provide important data on the validity of the questionnaire outside of the lab. Even though the time frame of the PhD did not allow for further exploration of this in the form of a paper, follow-up research on the validity of the VAIK within field settings (and other conditions influencing measurement such as language) is ongoing. Thus these studies provide data that will be essential to further establishing the output of the project even if this was not the main goal of the individual studies.

The next sections contain my published (and submitted articles). Each article includes a complete reference as well as a description of my contribution to this paper. Both art historical articles are provided with a summary of the paper before the full-text article since neither contains an abstract and have only briefly been touched upon in the introduction. In addition, both book chapters are accompanied by a short summary. Afterwards a discussion will follow that focuses on the interrelations of the papers and what story they can tell together. In-depth discussion of specific papers can be found in the discussion section of each respective paper.

2. Published Work

2.1 Basic Psychological Research

2.1.1

Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H. (2018). The Vienna Art Interest and Art Knowledge (VAIAK) Questionnaire: A Unified and Validated Measure of Art Interest and Art Knowledge. *Psychology of Aesthetics, Creativity, and the Arts*, <http://dx.doi.org/10.1037/aca0000205>

My Contribution:

My contributions to this paper are manifold. I came up with the core concept of the paper. Together with MF, I collected the previous data (“Initial Questionnaire”) section. Based on that we developed the pre-test; new items were created with MF, HB, JB and with feedback of RR, HL, and MP. Based on these results we again collaboratively came to a final version of the test. I was responsible for most of the data collection, for all of the analyses, and for the literature review the paper is based on. I was the main writer of the article but received feedback from all co-authors on different versions of the manuscript. In addition, MF also helped in interpretation of the data.

The Vienna Art Interest and Art Knowledge Questionnaire (VAIAK): A Unified and Validated Measure of Art Interest and Art Knowledge

AQ:1-2

AQ: au

Eva Specker, Michael Forster, Hanna Brinkmann, Jane Boddy, Matthew Pelowski,
Raphael Rosenberg, and Helmut Leder
University of Vienna

AQ: 3

Being interested in art and having knowledge about art are arguably central dimensions in art experience and two of the most important individual differences when assessing how people process or respond to art. Nonetheless, there is to date no reliable and validated measure of these dimensions. In this paper, we present the Vienna Art Interest & Art Knowledge Questionnaire (VAIAK) as a tool for researchers to measure both art interest and knowledge of their participants in studies involving visual art. In a 3-stage validation process, we first developed the questionnaire. Second, we pretested it qualitatively and quantitatively with a sample of lay people as well as art history students and art professionals. Finally, we conducted a large-scale validation study (>600 participants) with both lay people (psychology students) and art history students, where we present evidence regarding relations to other variables and internal structure as well as the ability to discriminate between relative lay people and experts. Our results show evidence for the reliability of the VAIAK scores as well as evidence for the validity as a measure of both art interest and art knowledge in (scientific) research, while at the same time also providing important clarification and differentiation between the two dimensions. With this new questionnaire, we offer a tool for researchers in empirical aesthetics, creativity, and also applied fields to quantify one or both of the art interest and art knowledge dimensions.

Keywords: art expertise, art interest, art knowledge, questionnaire, VAIAK

AQ: 5

Supplemental materials: <http://dx.doi.org/10.1037/aca0000205.supp>

Numerous studies have shown that art expertise significantly influences how we engage with art in terms of perception, understanding, and evaluation of visual art (Augustin & Leder, 2006; Brieber, Leder, & Nadal, 2015; Chamberlain & Wagemans, 2015; Chirumbolo, Brizi, Mastandrea, Mannetti, & Martens, 2014; Cupchick & Gebotys, 1988; Leder, Gerger, Brieber, & Schwarz, 2014; Winston & Cupchick, 1992). However, despite considerable research effort, up to now there has been no unified and validated measure to test for art expertise, or even an agreement on the dimensions of art expertise.

We argue that art interest and art knowledge are the foremost dimensions of art expertise. Given this, we aimed to develop a tool

to measure those dimensions. This paper first outlines general problems about defining and measuring art expertise and second introduces the Vienna Art Interest and Art Knowledge questionnaire (VAIAK). This new questionnaire aims to measure art interest as well as art knowledge. This will allow researchers interested in empirical aesthetics, creativity, and also applied fields to gain a more comprehensive and accurate understanding of their participants, enable them to measure art interest and art knowledge, as well as gain a better understanding as to how art interest and knowledge influence, or combine to influence, people engaging with visual art.

Eva Specker, Faculty of Psychology, Department of Basic Psychological Research and Research Methods, and Faculty of Historical and Cultural Studies, Department of Art History, University of Vienna; Michael Forster, Faculty of Psychology, Department of Basic Psychological Research and Research Methods, University of Vienna; Hanna Brinkmann and Jane Boddy, Faculty of Historical and Cultural Studies, Department of Art History, University of Vienna; Matthew Pelowski, Faculty of Psychology, Department of Basic Psychological Research and Research Methods, University of Vienna; Raphael Rosenberg, Faculty of Historical and Cultural Studies, Department of Art History, and Cognitive Science Platform, University of Vienna; Helmut Leder, Faculty of Psychology, Department of Basic Psychological Research and Research Methods, and Cognitive Science Platform, University of Vienna.

AQ: 12

The writing of this paper was supported by a grant to Raphael Rosenberg and Helmut Leder by Wiener Wissenschafts-, Forschungs-, und Technologiefonds (WWTF: Projectnumber: CS15-036). We thank Lisa Hegelmaier for her help in collecting data for this project, data entry, and the formatting of the tables in the manuscript. We thank Maximilian Douda for help with data entry. We also thank all lecturers who freed time in their busy teaching schedule to allow us to collect data from their students. Finally, we thank everyone in both the CreA and EVA lab who have helped us with their constructive and insightful comments that initially inspired this research and helped us refine the research along the way.

Correspondence concerning this article should be addressed to Eva Specker, Faculty of Psychology, Department of Basic Psychological Research and Research Methods, University of Vienna, Liebiggasse 5, A-1010 Vienna, Austria. E-mail: eva.specker@univie.ac.at

Background: Defining and Assessing What It Means to Be Interested or an Expert in Art

The importance given to interest in art and knowledge about art can be traced back as far as the ancient Greeks. Aristotle (1908), for example, emphasized the importance of interest. However, only around the 17th century, when the observer of art came into play as a new paramount player of the art world in addition to artists and committers (Fumaroli, 1994), did the concepts “interest” and “knowledge” become especially important. The first explicit discrimination of expertise levels—that of connoisseurs, amateurs, and *curieux*—also dates from this time (Olivier, 1976). In empirical study, the importance of these constructs can be traced back to Berlyne (e.g., Berlyne, Ogilvie, & Parham, 1968), who introduced a means of separating participants based on interest and knowledge by asking them questions such as how often they visited galleries (interest) or how many art classes they took (knowledge). The importance of interest and knowledge as key concepts affecting art evaluation is also reflected in current theories and models of aesthetic appreciation (Leder, Belke, Oeberst, & Augustin, 2004; Martindale, 1988; Pelowski, Markey, Luring, & Leder, 2016).

Before moving on to the current state of research it is worth mentioning another dimension of art expertise that has both historically and empirically played a role, namely taste and aesthetic sensitivity (two similar, but not quite identical terms). As a consequence of the introduction of lay people as observers in the art discourse in the 17th century (Fumaroli, 1994), the idea arose that aesthetic perception “requires the exercise of taste, perceptiveness, or sensitivity” (Sibley, 1959, p. 421). Experimentally, this notion of taste or aesthetic sensitivity was investigated mainly by Eysenck, Child, and Götz. Recently, taste has been of renewed research interest (Myszkowski, Çelik, & Storme, 2016; Myszkowski & Zenasni, 2016). Research done in this direction typically investigated taste along the lines of the definition of Child (1964), where taste is seen as the ability to respond to aesthetic stimuli in agreement with external standards. Several tests have been designed, some examples are the Visual Aesthetic Sensitivity Test (VAST I and VAST II; Götz, Lynn, Borisy, & Eysenck, 1979; Iwawaki, Eysenck, & Götz, 1979), the Meier Art Test (I: Art Judgment and II: Aesthetic Perception; Meier, 1942, 1963), and the Maitland Graves Design Judgment Test (Graves, 1946). For a methodological evaluation and critique of the latter three see Liu (1990). This includes a critique of the theoretical assumption inherent in all of the above mentioned tests that the agreement with “expert” choices or judgments can be used to assess taste/aesthetic sensitivity.

This definition of taste/aesthetic sensitivity relies on criteria about what is good, correct, and valid art. But these are concepts which are neither agreed upon nor stable (see Carroll, 2000, for an overview). Thus, we focus on art interest and art knowledge: art interest, because it is plausible to assume that the amount of time and money spent on art-related behaviors represents the amount of interaction with art; art knowledge, because it allows creating questions that can be reliably coded as right or wrong. There is less subjectivity in the decision whether an artwork is an example of impressionism than whether that artwork is good.

Current Issues in Defining and Measuring Expertise

Though most current measurement methods of expertise roughly follow the approach taken by Berlyne to assess interest and knowledge via self-reported behavior and education, a critical approach to the current research identifies two main issues. The first issue is a lack of a clear conceptualization or definition of art expertise, which leads to the second issue: a lack of a unified way of measurement.

As it stands, art expertise is a vague concept that lacks a clear definition. As a consequence, art interest and art knowledge are often confounded, even though these are, in our opinion, two different things. Conceptually it is clear that being interested in something does not necessarily imply that someone also knows a lot about the topic of interest. For example, many people are interested in outer space, but that does not mean that they have extensive knowledge about astronomy or astrophysics. Similarly, it is not hard to imagine a person that enjoys visiting museums on a Sunday afternoon, but that still knows rather little about art. In addition, it may be possible for a person with deep knowledge as a result of several years of school or study and training, to, along the line, lose their intrinsic interest in art (see Bell, 1914 or more recently Elkins, 2004).

Conceptually, this argument for an interest/knowledge decoupling seems relatively straightforward. However, the concepts of art interest and art knowledge are rarely separated empirically. For example, one of the most widely used questionnaires by Chatterjee, Widick, Sternschein, Smith, and Bromberger (2010, for another example see Chamorro-Premuzic & Furnham, 2005) combines both items measuring art interest (e.g., how often people visit museums) as well as art knowledge (e.g., asking about the amount of art education people had). Such mixtures are common, and these are just two examples. This common practice leads to a lack of clarity as to what is meant by art expertise. This also means that it is impossible to tell whether the influence of art expertise is attributable to differences in interest or differences in knowledge.

The second issue concerns measurement; the criteria people use to measure art expertise differ widely, and they are rarely based on a theory of what art expertise is. To give a few examples, being an expert might mean any of the following: a (bachelor) art history student (Leder et al., 2014; Rosenberg & Klein, 2015), a fine art student (Chamberlain & Wagemans, 2015; Winston & Cupchick, 1992), someone who completed a year of formal training (Belke, Leder, Harsanyi, & Carbon, 2010), a professional artist (Winston & Cupchick, 1992), and people employed in an art institution with jobs ranging from a gallery director or curator to a critic (Hekkert & Van Wieringen, 1996). Alternatively, researchers make up impromptu questionnaires, sometimes using only one item. For example, Locher, Smith, and Smith (1999, 2001) asked: “Have you had training in visual arts?” rated from 1 (*none*) to 7 (*art major*). Similarly, interest may be measured with many different metrics. Whereas some (Chatterjee et al., 2010) ask for concrete behavior such as amount of museum visits, others (Pelowski, 2015) use subjective scales (“I am comfortable looking at and discussing art”), and many use a combination of both (e.g., Chamorro-Premuzic & Furnham, 2005; Luring et al., 2016).

Different expert groups and different ways to measure their expertise can lead to misinterpretations of existing data and obscure interindividual differences. For example, even though the

AQ: 6

AQ: 7

term art expert may apply equally to all people mentioned above, we can assume that an artist will have a different kind of expertise than an art historian. Some of this knowledge may overlap (both may know about production methods) and some may not (only the art historian may have knowledge on the historical context of an artwork while only the art practitioner may intuitively appreciate how or why an artist generated a certain effect). Even more, because of differential underlying metrics and implicit definitions by different research teams, studies may actually lead to conflicting findings.

To our knowledge, no systematic attempt has been made up until now to develop a tool to comprehensively measure art interest. For art knowledge the only published scale that takes a systematic approach is the Aesthetic Fluency Scale, developed by Smith and Smith (2006). Even though this scale has the benefit of providing a continuous measure, it only measures art knowledge and not art interest. Crucially, it relies exclusively on self-report. Self-report measures are always prone to socially desirable answering. Even though the prevalence of social desirable answering on the Aesthetic Fluency Scale may be low in psychology samples (McKibben & Silvia, 2017), it may be more prevalent in a sample of art history students or others active in an art-related field. For these samples knowledge of art might be more relevant to their self-concept and/or status among peers. A similar argument can perhaps be made for studies that use the questionnaire in a museum context. The context of the museum may trigger a higher rate of social desirable answering, because the context makes people feel as if they should know about art.

In addition considering a classical definition of expertise, the 10-years-rule (Ericsson, 1996; Hayes, 1989), most of the samples considered experts in empirical studies, would not be candidates for such a label. According to this rule, it takes about 10 years of continued practice to become an expert in a field. Even if we do not apply this rule in its strictest sense and grant that expertise can be developed quicker, we do not believe—nor do we think Hayes or Ericsson would—that art history students or frequent museum goers count as art experts. Rather the label of “art expert” would only apply to professors in art history or mid to late career art professionals. However, most researchers looking at art “expertise” are not interested in the expertise of a small group of specialists, but rather how general differences in art interest and knowledge influence people engaging with art. Therefore, we focus on developing a validated measure of general art interest and art knowledge rather than an adequate measure of art expertise in the fullness of the term. Unfortunately a thorough discussion of the concept of expertise is beyond the scope of the current paper, but we refer the interested reader to Chi, Glaser, and Farr (1988) and Ericsson and Smith (1991) for more detailed accounts on this topic. We do not wish to propose that art interest and art knowledge represent the concept of art expertise in its entirety. What we aim to illustrate is that most authors to the present date, when talking about art expertise, have actually talked about either art interest, art knowledge, or a mixture of both. In addition, we argue that most researchers are, in fact, interested in differences in general art interest and/or art knowledge and how these influence people engaging with art, rather than expertise in the strict sense.

Development of the Vienna Art Interest & Art Knowledge (VAIAK) Questionnaire

To develop and validate the questionnaire we went through a three-part process. As a starting point we used a preexisting questionnaire that had been in use for a number of years at the University of Vienna and has been used in publications such as Augustin and Leder (2006). As a starting point for a new measurement tool, we reevaluated this questionnaire both conceptually as well as empirically. In a second step, we aimed to conceptually improve the questionnaire based on a solid theory of what constitutes interest and knowledge in art. Therefore, the VAIAK has been developed in a cooperation between academics from psychology and art history. As a third and final step we conducted a large-scale validation study with more than 600 participants. We will now first discuss the first two steps in development of the questionnaire. Afterward we will discuss our approach to validation and our findings.

Initial Questionnaire

The initial questionnaire consisted of three parts. Part A aimed to measure art interest, and Part B and C together aimed to measure art knowledge. Part A consisted of 15 questions on art interest (see Table 1), all rated on a 7-point scale. Part B consisted of 10 multiple choice questions (Table 4, Supplementary Material). Part C consisted of eight open-ended questions (Table 5, Supplementary Material), where people were asked to identify the artworks' style and artist.

For the analysis we compiled a dataset with a total of 740 participants (91% lay people, 9% art history students; 30.2% male, 69.8% female, $M_{\text{age}} = 24.3$ years, $SD = 5.3$, $\text{range} = 18\text{--}55$). These data were collected over four years (2013–2016) during different experimental procedures.

Results

Art interest. When evaluating the questionnaire conceptually we identified four items (Item 1, 5, 6, and 8) that would rather measure attitudes toward art (e.g., Item 8: “Art should first and foremost be decorative”) rather than a pure measure of art interest. When initially developing the questionnaire these questions were put in to test the conceptual notion that these items would measure attitudes about art rather than interest. To investigate this idea further we performed an exploratory factor analysis with varimax rotation over all 15 items. As can be seen in Table 1, this clearly showed a two-factor solution where Items 1, 5, 6, and 8 load on a second factor while the other items all cluster together on the first factor. The reliability of scores in Part A including all 15 items was $\omega = .93$, when all four items are removed the scores in Part A had a reliability of $\omega = .94$. Here and in all further text omega always refers to McDonald's omega. We report omega instead of Cronbach's alpha due to the numerous deficiencies in Cronbach's alpha outlined in detail in Dunn, Baguley, and Brunsden (2014). All calculations of omega were carried out with the “psych” package in R (Revelle, 2018). In sum, both the reliability analysis and factor analysis show support for the conceptual notion that Items 1, 5, 6, and 8 should be excluded because they measure attitude toward art rather than interest in art. We found that art history

Table 1
Art Interest Items and Factor Loadings

Item	Factor 1	Factor 2
1. An artwork has to primarily be beautiful for me to like it.		.805
2. I enjoyed art classes in school.	.561	
3. I like to talk about art with others.	.831	
4. I have many friends/acquaintances that are interested in art.	.720	
5. I cannot stand ugly artworks.		.777
6. Art has to be about an exact representation of the world.		.621
7. I'm interested in art.	.842	
8. Art should first and foremost be decorative.		.727
9. I'm always looking for new artistic impressions and experiences.	.797	
10. In everyday life I routinely see art objects that fascinate me.	.703	
11. I come from an art interested family.	.583	
12. How often do you visit art museums and/or galleries?	.749	
13. How often do you read books, magazines or catalogs about art?	.754	
14. How often do you look at images of artworks (catalogs, internet, etc.)?	.735	
15. How often do you visit talks about art or art history?	.698	

Note. Coefficients smaller than .300 are not shown. Bold font is used to indicate the items that were dropped.

students ($M = 63.24$, $SD = 8.71$) were more interested in art than lay people ($M = 37.91$, $SD = 12.89$), $t(81.85) = 14.79$, $p < .001$, Hedges' $g = 2.01$. In all the following text we use g to refer to Hedges' g . We report Hedges' g instead of Cohen's d because it takes into account unequal sample sizes; this was especially important because in some cases the assumption of equality of variances was violated.

Art knowledge. When reevaluating the questionnaire we were curious as to how well the multiple-choice questions in Part B captured art knowledge. There did not seem to be a unified type of knowledge that formed the basis of the different questions in Part B. That is, this part seemed to lack conceptual clarity. Even though it can be considered art knowledge that it was Vincent van Gogh rather than Edvard Munch who cut his ear off, it is hard to see how this relates to knowing where "Die Brücke" was founded. That is, some questions rather captured trivia knowledge that not necessarily translates to knowledge in the arts. The reliability of scores in Part B and C together was $\omega = .86$. However, when looking at the parts separately scores in Part B turned out to be rather unreliable, $\omega = .42$. In Part C, reliability was $\omega = .88$. We therefore dropped Part B in subsequent analyses.

Second, we were interested to know whether Part C was able to discriminate between art history students and lay people. We found that art history students and lay people differed in how well they performed in Part C, $t(64.22) = 13.97$, $p < .001$, $g = 2.21$. Art history students on average gave 9.34 ($SD = 3.49$) and lay people gave on average 2.83 ($SD = 2.71$) correct answers out of a possible 16. The density plots (see Figure 1) clearly show differentiated distributions between lay people and art history students. In addition, the distributions of lay people showed a clear floor effect, with many people not answering any question correct at all.

Finally, participants in Part C were also asked to indicate whether they knew the artwork. We wanted to know whether this self-reported, subjective knowledge (saying you know the artwork) was correlated with objective knowledge (classifying it the right way). This was of interest because most current measures that aim to measure knowledge use self-report. We found that subjective art knowledge was positively correlated with objective art knowledge (Part C), $r = .63$, $p < .001$ (Figure 8, Supplementary Material).

This indicates that there was a convergence between subjective (self-reported) knowledge and actual knowledge. It must be noted that Part C was constructed to make it possible for people who do not know the exact image shown to give the correct answer using their knowledge about art. For example, someone may not know the artwork *The Water Lily Pond* by Claude Monet. However, if they are familiar with other works by the artist they would be able to correctly identify the artist of the image. Correspondingly, if they are familiar with Impressionist painting they would also be able to correctly identify the style of the image. In both these cases, someone may indicate that they do not know the image and at the same time give the correct answer(s). This will lead to a reduction in the correlation between the two different measures. In addition, one may have a subjective feeling of knowing (as with a song on the radio) without any objective knowledge (e.g., not knowing the artist or title of the song), which would also lead to a reduction in the correlation.

Relationship between art interest and art knowledge. We were interested to see how art interest and art knowledge relate.

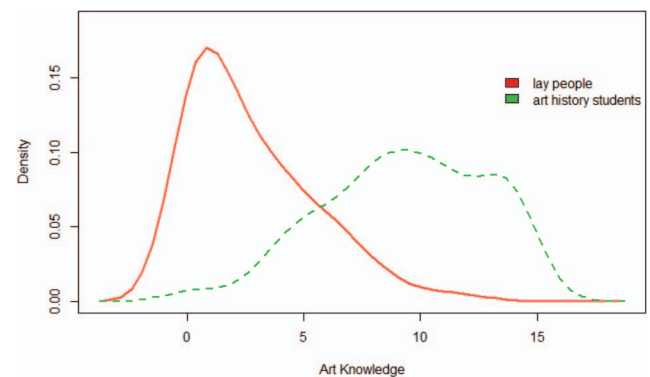


Figure 1. Density distribution of lay people and art history students on art knowledge. The solid (red) line represents lay people, and the dotted (green) line represents art experts. The density distribution is an estimate of the population distribution rather than a histogram of the sample; it uses a fitted curve that is not cut off at the actual minimum (0) or maximum (26) of the scale. See the online article for the color version of this figure.

Earlier we argued for a clear distinction between these two concepts. Even though we acknowledge (and assume) that these concepts are related to each other, we have argued for them being separated concepts. If art interest and art knowledge are separate yet related dimensions they should be moderately correlated to each other. Based on the previous results we used all items of Part A except 1, 5, 6, and 8 to measure art interest and we used Part C to measure art knowledge. The correlation between art knowledge and art interest was $r = .64$, $p < .001$, which we interpret to be moderate (see Figure 2).

F2

Discussion

The reevaluation of the initial questionnaire revealed some clear points for improvement. For art interest, there were clear empirical and conceptual grounds to drop Items 1, 5, 6, and 8. This led us to modify the art interest scale by dropping these items and keeping the rest. The remaining items showed a high reliability as well as a clear one-factor solution, both indications that Part A measured one unified dimension, namely art interest. For art knowledge, we found that Part B lacked both conceptual clarity and empirical support. This led us to decide to conceptually reconsider this part

(see below). Part C was both conceptually and empirically a good measure of art knowledge. However, as can be seen from Figure 1 there was a clear floor effect for lay people. This would make the art knowledge scale very difficult for lay people, which in turn leads to low variation and consequently reduces the information and reliability of this scale when testing a sample of lay people only. This indicates that even though Part C is good at differentiating between lay people and experts, it cannot discriminate well among lay people with different degrees of (low) art knowledge. In other words, it cannot differentiate between people who know little and people who know nothing at all. Therefore, we decided to keep the structure of Part C as it was, but included a wider range of artworks; more items that are easy and more that are difficult. The latter was done to avoid a ceiling effect. This is crucial for a questionnaire that is applicable both in mixed samples and in homogenous samples of only lay people or only experts.

In addition, we wanted to include different artworks to gain a higher level of content coverage. The current items had a strong focus on modern art and as a consequence had a very limited range of art styles and time periods. In our new selection, we therefore

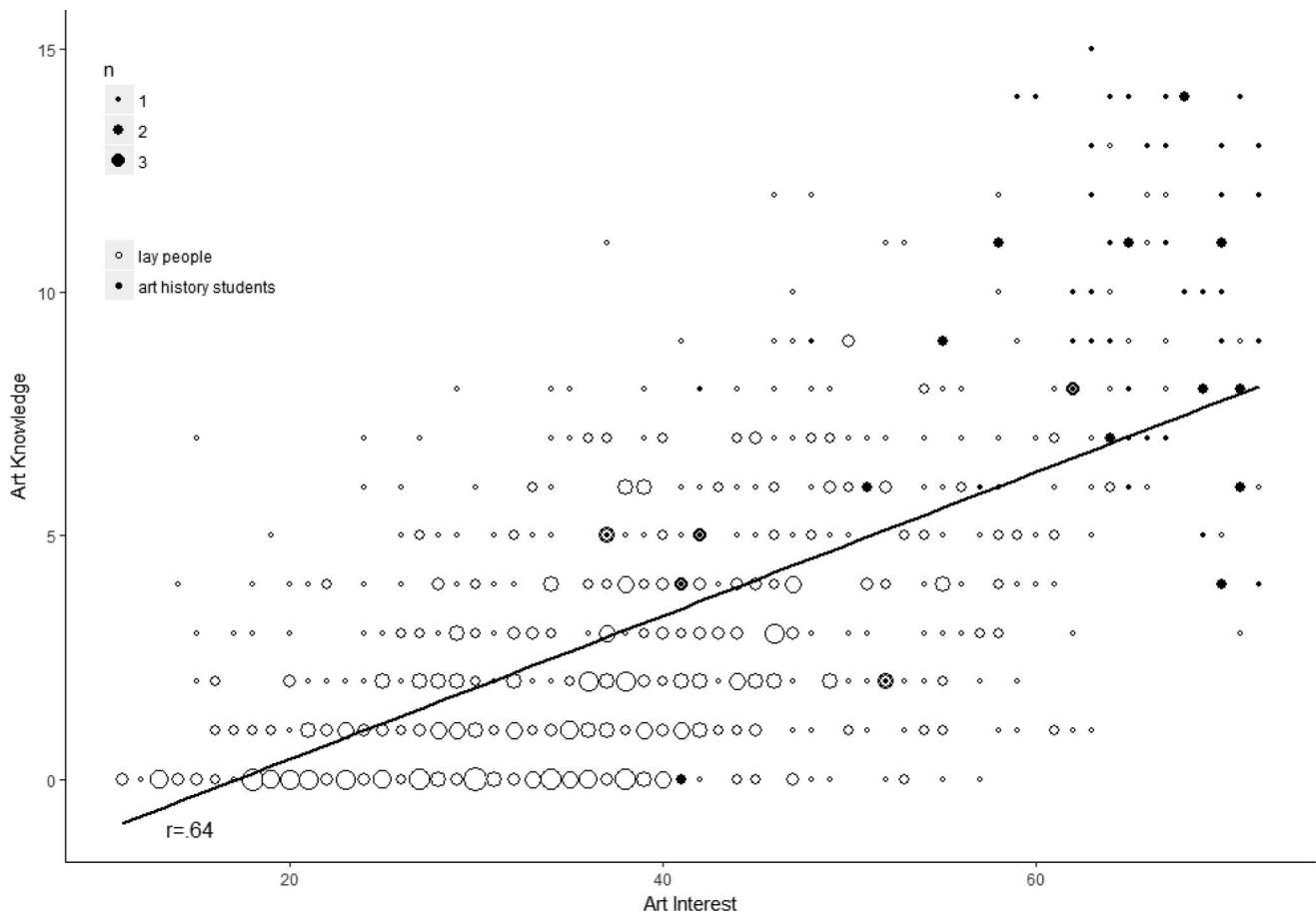


Figure 2. Scatterplot of the relationship between art knowledge and art interest. This is a count plot, which is a type of scatterplot where the size of the dots represents the amount of people (or counts). Empty circles represent lay people, filled circles represent art history students. Dot-filled circles represent both groups simultaneously.

included items from the early middle ages to contemporary art. In addition, we also decided to drop some items that were suboptimal in coding. For example, it was hard to discriminate between a typo, sloppy writing, or a incorrect answer when someone wrote “Ma-net” where the correct answer was “Monet.”

Finally, we decided to introduce a new Part B with which we hoped to increase content coverage by including items that measures other skills than classification. Content coverage is linked to content validity, which was defined by Messick (1989) as follows: “Content validity is based on professional judgments about the relevance of the test content to the content of a particular behavioral domain of interest and about the representativeness with which item or task content covers that domain” (p. 17). Essentially this means that the content of the test is representative for the domain it aims to measure. We chose to include iconographical knowledge as well as knowledge about production methods. Both of these types of knowledge are a standard part of Western art history education, but can also be gained independently. During art education this type of knowledge would have been acquired through standard textbooks such as *Janson’s History of Art* (Davies et al., 2010, first published in 1960) or the *Story of Art* by Ernst Gombrich (2006, first published in 1950). Within their studies students learn how to recognize iconographical conventions of depicting Christian or mythological themes. *Birth of Venus*, for example, can be recognized by either portraying Venus on a seashell or portraying her on the wave that she was born from. Being able to recognize such themes helps in understanding what is depicted (the content) and what is intended with the work (the meaning). Despite the high prevalence of themes in premodern art, this type of knowledge is also applicable to contemporary art. Many contemporary artists still use age-old iconographical themes in their work or refer to them. Therefore, we see knowledge of iconography as an important part of general art knowledge and as a skill that can be applied to all art forms.

The same argument can be made for knowledge of production methods. The method used influences the way the final work looks. Low-level features of artworks, like texture, are quickly processed and recognized when looking at original artworks, even by lay viewers (Specker, Tinio, & van Elk, 2017). However, those with more expert knowledge may show a deeper appreciation or give more emphasis to these features (Graham, Hughes, Leder, & Rockmore, 2012). In addition, the production method can be informative as to what the artwork is trying to convey.

Pretest

Method

Participants. The pretest was conducted as a second step in development and last step before validation. This was done to select items based on response processes as well as item difficulty. We used quantitative and qualitative measures to evaluate our new version of the questionnaire. We had a total sample of 62 participants in the quantitative part (25 lay psychology students, 29 art historians, eight art professionals; 49 females, 11 males, one person identifying as other, one person with missing data; $M_{\text{age}} = 27.28$, $SD = 6.72$, $\text{range} = 21\text{--}55$) and eight in the qualitative part (three lay people, two art history students, three art professionals; five females, three males; $M_{\text{age}} = 29.86$, $SD = 4.95$, $\text{range} =$

26–40). We especially wanted to include a sample of art professionals: people who were working in the arts and had an art history or related background. Per definition of art professional, those individuals should unanimously score higher than lay people in the questionnaire. Including art professionals allowed us to evaluate whether the questions also can discriminate on the higher end of the knowledge range.

Materials. Based on the initial questionnaire, we kept the items that loaded on the first factor (above, Table 1). Thus, the art interest scale (Part A), had a total of 11 questions. As in the initial questionnaire the art knowledge scale consisted of two subscales. Part B consisted of 18 multiple choice questions asking about iconography and art production methods. For each question people saw an image and were asked to identify the scene or the technique used. One question asked about the relation between figures in the image, and two questions did not have an image but only asked about technique. Part C consisted of 19 images; for each image participants were asked to identify the artist and style, resulting in a total of 38 items. Pretest questions for Part B and C can be found in the Supplementary Material, Tables 6 and 7. In addition, as before, we included a measure of subjective knowledge (asking people to indicate whether they knew the image or not).

Procedure. For the quantitative part, lay people and art history students participated in seminars. The images were projected and people used a smartphone, tablet, computer, or pen and paper to answer the questions. The art professionals completed the questionnaire at home, they were sent a pdf of the images to be able to complete the questionnaire.

For the qualitative part all participants came to the lab where they viewed the images on a laptop. They were asked to talk out loud about their answer process to the questions, while viewing the images with the experimenter (E.S.). Their sessions were audio recorded. Thus, they provided feedback on how they interpreted the questions or answering tactics (e.g., elimination of unlikely answer options in the multiple choice part).

Results

Art interest. Reliability of the scores of the art interest scale (Part A) was $\omega = .95$. We conducted a Parallel Analysis with the R package “paran” (Dinno, 2012) using 5000 iterations, to test for unidimensionality. This indicated a one-factor solution with an adjusted eigenvalue of 5.50. We took this as an indication that our scale shows the intended unidimensionality.

There was no significant difference between art history students and art professionals on art interest, $t(35) = -1.157$, $p = .255$, $g = .46$. We therefore grouped them together in a group of art experts to test whether there was a significant difference between experts and lay people. This was the case, $t(60) = -7.31$, $p < .001$, $g = 1.89$. Art experts had a higher level of art interest ($M = 65.76$, $SD = 7.27$) than lay people ($M = 46.56$, $SD = 13.35$).

Art knowledge. The art knowledge scale consisted of two subparts. We first assessed each part on its own and then as a combined scale. In a first step, we tested whether the different groups (lay people, art history students, art professionals) differed on the sum score of Part B when all items were included. A one-way ANOVA showed significant differences, $F(2, 59) = 32.65$, $p < .001$, partial $\eta^2 = .53$. Post hoc pairwise comparisons showed no significant difference between art history students and

art professionals, $t(35) = .261, p = .796, g = .10$. Therefore, we again grouped art history students and art professionals together in one art expert group. In a second step we did a reliability analysis for scores in Part B to see which items to retain and which to drop. For each item we calculated the reliability (omega) of the whole scale if this item was deleted. This allowed to evaluate the importance of each item for the overall reliability. This was done separately for art experts and lay people. This approach was taken to ensure that future use of a subsample of either group would still lead to reliable results. This was an important caveat because using a subsample typically leads to a reduction in reliability because of the reduction in sample size. Given our aim to provide a questionnaire that at the same time can distinguish among and between lay people and art experts, it was important to ensure that the scale still works adequately in subsamples. When selecting items we consistently compared how dropping this item would influence reliability in both groups. If dropping an item would increase reliability in both groups, this item was dropped. If dropping an item would increase reliability in one group but not in the other we compared the relative gains and losses. Based on this analysis the optimal solution was one with six items (Items 2, 3, 5, 6, 7, and 13)¹ that had a reliability of $\omega = .84$. Reliability for lay people was $\omega = .82$ and for art experts it was $\omega = .79$. The final items can be found in the Supplementary Material, Table 8. Note that, because Part B was also planned to be combined with Part C to represent art knowledge, we expected further increased overall reliability as well. For Part B we found that art experts gave more correct answers than lay people, $t(60) = -6.44, p < .001, g = 1.67$. Art history students and art professionals on average correctly answered 4.62 of six questions ($SD = 1.46$), whereas lay people on average correctly answered 2.04 ($SD = 1.67$).

The qualitative study indicated that people approached the questionnaire as expected. People in all groups attempted to answer the question by using “signs” in the image. For example, for the first image that represented a Venus a lay person said: “I thought maybe Venus but from the pose and the way she covers her intimate parts maybe Eve. I would choose between those two answers, the beauty points to Venus, but from the pose probably Eve.” In comparison, an art history student said: “Art history has specific ways to portray people by marks or poses. Here the woman is scarcely clad and her pose indicates Venus Pudica.” These examples clearly show that both participants used signs in the image, in this case even the same signs (pose of the woman), to arrive at an answer. However, because of a difference in knowledge the art history student arrived at the correct answer while the lay person arrived at the wrong conclusion. This pattern was similar for the other questions.

We then followed the same procedure for Part C. We tested whether the groups differed on their sum scores when all items were included. The groups differed on the sum score of Part C, $F(2, 59) = 19.21, p < .001$. There was a difference between art history students and art professionals, $t(35) = -2.41, p = .02, g = .96$, and both groups differed from lay people; for art history students: $t(52) = -4.71, p < .001, g = 1.29$, and for art professionals: $t(31) = -5.32, p < .001, g = 2.16$. In a second step, we conducted a reliability analysis of the scores in Part C to see which items to retain and which to drop. We first did a reliability analysis including all questions, the reliability was $\omega = .92$. We then conducted reliability analyses for the three groups separately.

These did not give a clear indication of which items to drop.² To still make the questionnaire shorter and more efficient, we took a conceptual approach. In a first step we took out items that were difficult to code (Cindy Sherman, Francis Bacon, Piet Mondriaan, Pablo Picasso). In these cases it was easy to classify the artworks by artist but difficult by style. This idea was supported by the qualitative data. For example, all art history students and art professionals mentioned that Bacon’s work is expressionistic; however, because of the time that it was painted in it was “too late” to be typically classified this way. In a second step we went deeper into the qualitative data to see whether this could give us further insight on which items would be good to drop (or keep). This rendered two obvious candidates: Dürer and Vermeer. For the Dürer image most participants (especially non lay people) mentioned that the style could not be derived from the image alone. They argued that to classify it correctly one needed to recognize the artist and to be able to put it in an historical context. The Vermeer had the same problem, however in this case some argued that the artist cannot easily be derived from the image either. In addition, the Fragonard and Friedrich paintings had the opposite issue. In this case participants said that the style of the image was very clear, however they found that the artist could not be easily derived from the work. Finally, we dropped the Leonardo da Vinci item because overall it was a very difficult item to answer. Only 23% of people were able to correctly identify the artist and 34% of people were able to correctly identify the style. Table 2 shows the item difficulties for the final set of items including the Leonardo da Vinci both for the different groups and overall for all participants together. The style of the Van Gogh item (post impressionism) was difficult as well, but we retained it for two reasons: first, identifying the artist was relatively easy for both lay people and experts, and second, the difficult style helps to avoid ceiling effects.

This reduction led to a final set of 10 images, corresponding to 20 items (two questions per image), that had a reliability of $\omega = .92$. The final items can be found in Supplementary Material, Table 9. For this final version of Part C, art professionals and art history students did not significantly differ in how many they got correct, $t(35) = -1.77, p = .084, g = .74$. Therefore, we combined art professionals and art history students together in one group of art experts. Reliability was $\omega = .94^3$ for lay people and $\omega = .87$ for art experts. Experts got more answers correct than lay people, $t(60) = -4.91, p < .001, g = 1.27$. Art history students and art professionals on average answered 12.78 ($SD = 4.33$) of 20 questions correctly while lay people only answered 6.84 ($SD = 5.14$).

The intention of the questionnaire was to sum up Part B with Part C to come to a sum score of art knowledge. To see whether this was justified we did a reliability analysis over all items of Part B and C; the reliability of this was $\omega = .93$ for lay people and $\omega =$

¹ The interested reader can find tables reporting the relevant statistics related to the different selections split up by subsample in the Supplementary Material Tables 10 and 11.

² For both the reliability analysis over all participants as well as the three groups separately, deleting any item would decrease the reliability. However, these decreases were very small (second decimal point).

³ The item pertaining to the “post-impressionism” (item 6) was excluded from this analysis, because there was no variance for this item in our sample of lay people (i.e., none of the lay people answered this item correctly).

Table 2
Item Difficulty for Artist and Style

Item	Lay people		Experts		All	
	<i>n</i>	<i>M</i> (<i>SD</i>)	<i>n</i>	<i>M</i> (<i>SD</i>)	<i>n</i>	<i>M</i> (<i>SD</i>)
Artist: Munch	25	.64 (.49)	37	.86 (.35)	62	.77 (.42)
Style: Expressionism	25	.28 (.46)	37	.59 (.50)	62	.47 (.50)
Artist: Giotto	25	.08 (.28)	36	.53 (.51)	61	.34 (.48)
Style: Late middle Ages	25	.08 (.28)	37	.41 (.50)	62	.27 (.45)
Artist: Van Gogh	25	.48 (.51)	37	.76 (.44)	62	.65 (.48)
Style: Post-Impressionism	25	.00 (.00)	37	.05 (.30)	62	.03 (.18)
Artist: Klimt	25	.84 (.37)	37	.95 (.30)	62	.90 (.30)
Style: Jugendstil	25	.20 (.41)	37	.68 (.48)	62	.48 (.50)
Artist: Michelangelo	25	.64 (.49)	37	.78 (.42)	62	.73 (.45)
Style: High Renaissance	24	.33 (.48)	37	.68 (.48)	62	.54 (.50)
Artist: Rubens	25	.08 (.28)	37	.54 (.51)	62	.35 (.48)
Style: Baroque	25	.12 (.33)	37	.49 (.51)	62	.34 (.48)
Artist: Renoir	25	.20 (.41)	37	.32 (.48)	62	.27 (.45)
Style: Impressionism	25	.32 (.48)	37	.68 (.48)	62	.53 (.50)
Artist: Dali	25	.64 (.49)	37	.76 (.44)	62	.71 (.46)
Style: Surrealism	25	.40 (.50)	37	.84 (.37)	62	.66 (.48)
Artist: Duchamp	25	.24 (.44)	37	.62 (.49)	62	.47 (.50)
Style: Dada	25	.24 (.44)	37	.65 (.48)	62	.48 (.50)
Artist: da Vinci	25	.12 (.33)	37	.30 (.46)	62	.23 (.42)
Style: Renaissance	25	.16 (.37)	37	.46 (.51)	62	.34 (.48)
Artist: Warhol	25	.60 (.50)	37	.84 (.37)	62	.74 (.44)
Style: Pop Art	25	.44 (.50)	37	.78 (.42)	62	.65 (.48)

Note. Lay people = psychology students; Experts = used to denote the combined group of art history students and art professionals; All = lay people and experts; *n* = number of participants; *M* = mean; *SD* = standard deviation. Items scores can range from 0–1; the mean expresses the percentage of people who answered correctly.

.73 for art experts, indicating that a sum score of both parts is justified.

We assessed whether our art knowledge scale (Part B and C together) was able to discriminate between experts and lay people. We found that this was the case. Art experts and lay people differed in art knowledge, $t(60) = -6.109, p < .001, g = 1.58$, with art history students and art professionals giving 16.62 ($SD = 4.59$) correct answers (of 26) on average and lay people giving 8.36 ($SD = 6.07$) correct answers on average. To put this in context, the density plots (see Figure 3) clearly show differentiated distributions between lay people and experts without any floor or ceiling effects. In fact, none of the participants scored 0, the minimum being 2 points. No one scored the maximum of 26 points.

Relationship between subjective and objective art knowledge. As before, we included a measure of subjective art knowledge in Part C. We found that subjective knowledge was positively correlated with objective knowledge, $r = .730, p < .001$ (Figure 9, Supplementary Material). This indicates that there was a convergence between subjective (self-reported) art knowledge and actual art knowledge.

Relationship between art interest and art knowledge. To get insight into the relation between art interest and art knowledge we computed a correlation between the sum scores for the subscales. The correlation between art interest and art knowledge was $r = .737, p < .001$ (see Figure 4). Inspection of the scatterplot indicates that there were more instances of high art interest with a low knowledge than instances of high art knowl-

edge with low interest. This tentatively indicates that measuring interest alone does not suffice to measure knowledge in art.

Discussion

For art interest, the parallel analysis showed evidence for the unidimensionality of the scale. In addition, we found that scores on our scale were reliable. For art knowledge, we first assessed the two subparts of our scale separately. The qualitative analysis of response styles showed that Part B and Part C worked as intended. Additionally, quantitative analyses gave clear results regarding which items to retain and which to drop for Part B. This led to a final set of six items in Part B that would be suitable for a larger validation study. In Part C we could reduce the set of items to 10 images that were both easy to code and interpret, and varied in difficulty. This final set of images did not show a floor effect. Together, test scores on Part B and C were reliable.

We found that art professionals and art history students did not differ in this questionnaire on art interest and art knowledge. This lends support to the idea that the questionnaire tests for general art interest and art knowledge. The higher level of specialized knowledge in art professionals and generally lacking in art history students cannot be detected by the questionnaire. It has to be noted, however, that our power to detect difference between these groups was low. However, and crucially, the questionnaire could clearly distinguish between art experts (art history students and art professionals) and lay people in both art interest and art knowledge.

Furthermore, we found that subjective art knowledge is correlated with objective art knowledge. This indicates that there was a convergence between subjective (self-reported) art knowledge and actual art knowledge.

Finally, art interest and art knowledge showed a moderate correlation with each other, supporting the notion that even though they are related concepts they are also clearly two separate dimensions. Whether interest and knowledge differentially influence art

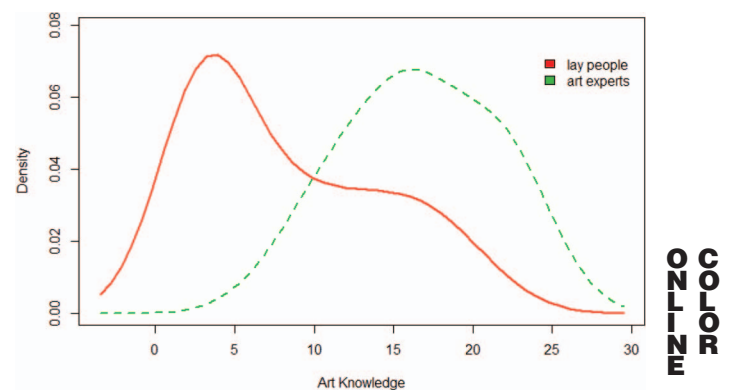


Figure 3. Density distribution of lay people and art experts (that is the combined group of art history students and art professionals) on art knowledge. The solid (red) line represents lay people and the dotted (green) line represents art experts. The density distribution is an estimate of the population distribution rather than a histogram of the sample; it uses a fitted curve that is not cut off at the actual minimum (0) or maximum (26) of the scale. See the online article for the color version of this figure.

VIENNA ART INTEREST AND ART KNOWLEDGE QUESTIONNAIRE

9

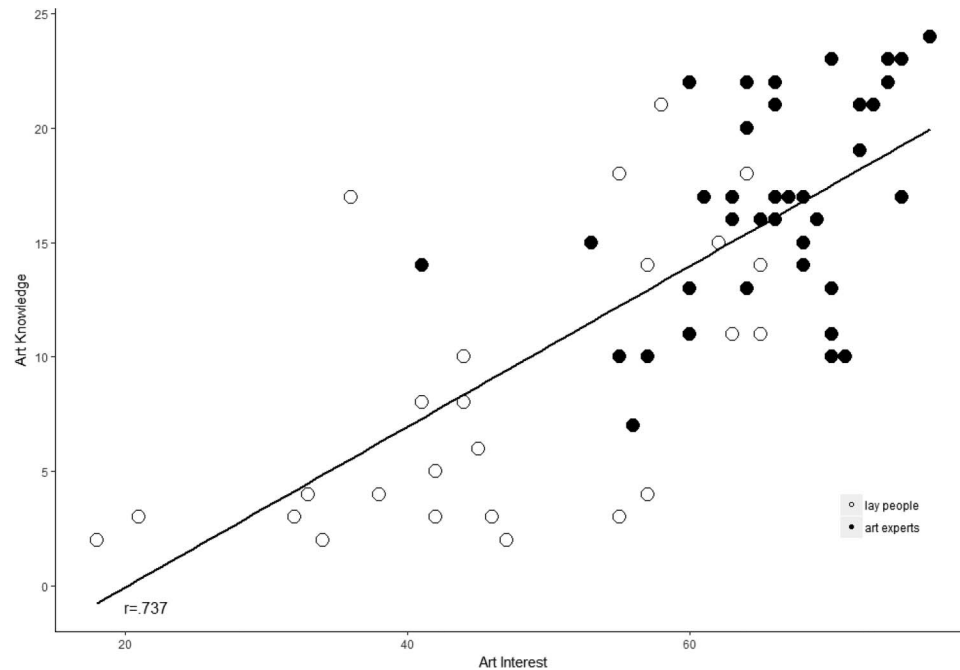


Figure 4. Scatterplot of the relationship between art knowledge and art interest. Like Figure 2, this is a count plot. Empty circles represent lay people, filled circles represent art history students. As can be seen, all dots are the same size, indicating that it didn't occur that people had the exact same value on both scales. However, for the sake of unity we still opted to present a counts plot rather than a scatterplot.

experiences is, however, an empirical question. However, with this questionnaire we hope to provide a tool to answer this question.

From this process of development, we ended up with a final version of the VAIK questionnaire. The final version of the VAIK has two scales: one measuring art interest (Part A) and one measuring art knowledge (with two subscales; Part B and C). A German and an English paper-and-pencil version of the VAIK can be found in the [Supplementary Material](#).⁴ Full materials (including a syntax for coding of the answers) can be found on the Open Science Framework: <https://osf.io/88d2d>.

The first scale of the VAIK (Part A) aims to measure art interest (see Table 1). It consists of 11 questions, all rated on a 7-point scale, with the subjective interest questions having anchors from 1 (*not at all*) to 7 (*completely*) and the questions about concrete behaviors are rated on a scale from 1 (*less than once a year*) to 7 (*once a week or more often*). This means that scores can range from 11 to 77. The second scale of the VAIK aims to measure art knowledge and consists of two subsections: Part B (see Table 8) and Part C (see Table 9). In Part B, questions concern knowledge of iconography and production methods. In Part C, participants are presented with a selection of 10 artworks and are asked to name the artists and styles. Completion of the VAIK questionnaire as a whole (that is both the art interest and the art knowledge scale) takes about 10 minutes.

Validation Study

The approach we took to validate the VAIK is an argument-based approach to validation as proposed by Kane (1992, 2006, 2009). In this approach, one builds an argument that focuses on the

use of a test for a particular purpose and bases this argument on empirical evidence to support the particular use. In short, this approach requires three sources of evidence: (a) evidence that the test measures what it claims to measure, (b) evidence that the test scores display adequate reliability, and (c) evidence that test scores display relationships with other variables in a manner congruent with its predicted properties. This approach is congruent with the current Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014) and other accounts of validity such as the one by Messick (1989).

Following this approach, in our discussion of the development of the questionnaire we have tried to show evidence for (a); *the test measures what it claims to measure*. The art interest scale of the VAIK covers a wide range of items that assess interest in art not only by asking for graded (dis)agreement to statements but also by asking for concrete behaviors such as the amount of museum visits. The art knowledge scale of the VAIK assesses different kinds of knowledge, not only does it assess knowledge about art styles and artists but it also tests for knowledge about techniques and iconography.

In addition, the prestudy has provided preliminary evidence for (b); *adequate reliability of the test scores*. Finally, by analyzing the relationship between subjective and objective art knowledge we

⁴ The questionnaire was developed in German, however an English translation was created through the method of back-translation. Both can be found in the [Supplementary Material](#).

Fn4

have tried to show preliminary evidence for (c); *that the test scores display relationships with other variables congruent with its predicted properties*. The validation study aimed to extend the evidence for (b) and (c).

Method and Participants

We tested 617 people: 157 art historians (25.4%) and 460 psychology students (74.6%). The sample consisted of 424 females (68.7%) and 185 males (30%); six participants identified as “other” (1%) and for two participants data on gender were missing. The average age was 23.85 years ($SD = 9.2$, $range = 17-81$).

The testing procedure was the same as in the pretest. That is, participation took place in seminars and the images were projected. We employed this method to economically collect data of a large sample. In addition to the VAIK in full (that is both the art interest and the art knowledge scale), we also administered the Aesthetic Fluency Scale (AFS; [Smith & Smith, 2006](#)), the Biographical Inventory of Creative Behaviors (BICB; [Batey, 2007](#); [Silvia, Wigert, Reiter-Palmon, & Kaufman, 2012](#)), and the Creative Achievement Questionnaire (CAQ; [Carson, Peterson, & Higgins, 2005](#)) as criterion validity evidence. The AFS aims to measure art knowledge, the BICB aims to test everyday creativity, and the CAQ aims to test creative achievement. Participants first completed the VAIK, second the AFS, third the BICB, and finally the CAQ.⁵

We expected a high and positive correlation between the VAIK art knowledge scale and the AFS, because they both aim to measure the same dimension, art knowledge. In contrast, we expected low correlations with the BICB and the CAQ. This is because neither of the scales aim to measure art knowledge or art interest, therefore they should measure something different from both scales included in the VAIK questionnaire.

Results and Discussion

We will first discuss the results on internal properties of the final version of the VAIK. We then consider the results regarding the relationships between the VAIK scales and the other questionnaires. Because of our large sample size, we had a high level of power to detect even very small effects; we therefore used effect sizes as a point of inference rather than p -values.

Art interest. The reliability of the scores of art interest was $\omega = .94$ for the total sample, $\omega = .92$ for psychology students, and $\omega = .82$ for art history students. Again we found that art history students and lay people differed in art interest (see [Figure 5](#)); art history students ($M = 62.53$, $SD = 8.14$) were more interested in art than lay people ($M = 42.46$, $SD = 13.14$), $t(615) = -17.99$, $p < .001$, $g = 1.66$.

Art knowledge. The reliability of the scores of art knowledge was $\omega = .89$ for the total sample, $\omega = .77$ for lay people, and $\omega = .85$ for art history students. Again, we found that art history students and lay people differed in art knowledge (see [Figure 6](#)); art history students correctly answered 15.29 ($SD = 5.45$) questions (of 26) on average and lay people correctly answered 6.26 ($SD = 3.14$) questions on average, $t(615) = -23.43$, $p < .001$, $g = 2.16$. As expected in this larger sample, some participants did not answer any question correctly and thus scored 0, however the low frequency suggested that there was no floor effect. Also, none

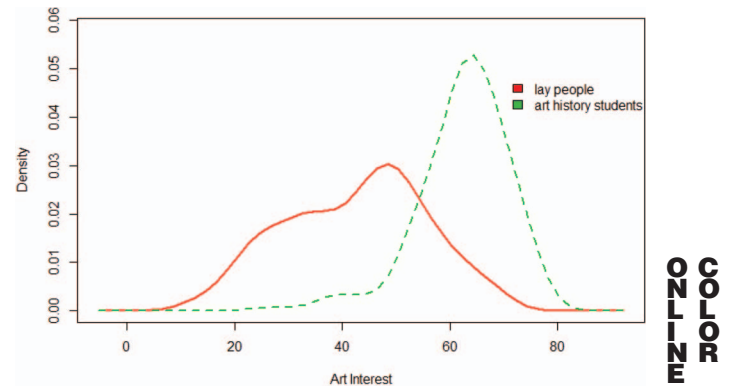


Figure 5. Density distribution of lay people and art history students on art interest. The solid (red) line represents lay people, and the dotted (green) line represents art history students. The density distribution is an estimate of the population distribution rather than a histogram of the sample; it uses a fitted curve that is not cut off at the actual minimum (11) or maximum (77) of the scale. See the online article for the color version of this figure.

of our participants answered all questions correctly, suggesting that there was no ceiling effect.

Relationship between art interest and art knowledge. The correlation between art interest and art knowledge was $r = .624$, $p < .001$ (see [Figure 7](#)). In addition, a comparison of [Figure 5](#) and [Figure 6](#) clearly showed the differences between these two measures. Art history students had a very narrow distribution on art interest, meaning that all art history students had a relatively high level of interest in art; however, they had a wide distribution on art knowledge, meaning that art history students varied more in how much they knew about art. For lay people we see the opposite pattern: they varied a lot in how interested they were in art. In contrast, what they shared is a collective low knowledge of art. In sum, in general art history students were highly interested in art and lay people in general knew little about art. This is in line with the findings of the pretest and emphasizes, in our opinion, the need to clearly differentiate these dimensions.

Relationships with other variables. We aimed to show consistency of art knowledge scores with other measures of knowledge by use of the AFS and the subjective art knowledge ratings. For both we assumed high correlations. For other constructs measured by the BICB and CAQ we assumed low correlations with art knowledge.

The correlation between the AFS and the VAIK art knowledge scale was $r = .618$, $p < .001$ ([Figure 10](#), Supplementary Material). We interpret this as a moderate correlation, though this is consistent with our expectations; the correlation was smaller than one would have theorized and therefore does not provide clear evidence. From the figure we can see that there are some people who scored very high on the aesthetic fluency scale but very low on the VAIK art knowledge scale; these participants probably influenced the correlation strongly making it smaller than expected.

⁵ All three scales (AFS, BICB, CAQ) were translated from English to German by a native German speaker and then back-translated by a native speaker of both English and German. Afterwards, differences between the translations were solved by discussion between the two translators.

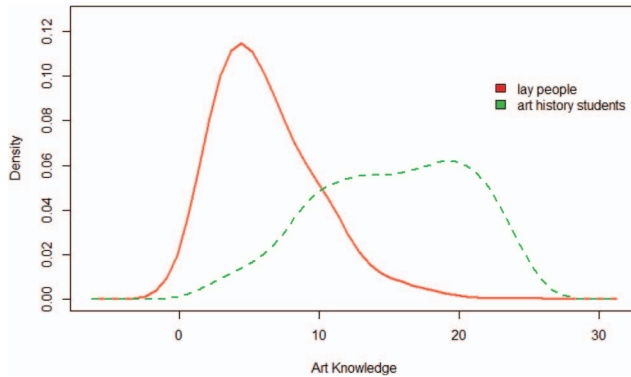


Figure 6. Density distribution of lay people and art history students on art knowledge. The solid (red) line represents lay people, and the dotted (green) line represents art history students. The density distribution is an estimate of the population distribution rather than a histogram of the sample; it uses a fitted curve that is not cut off at the actual minimum (0) or maximum (26) of the scale. See the online article for the color version of this figure.

A possible explanation of these findings may be the difference in measurement method. The AFS uses self-report of knowledge whereas the VAIK scale uses a measure of predefined correct and incorrect answers. The subjective, self-reported artwork knowledge measure (i.e., “Do you know this artwork?”) offers an opportunity to compare the two different measurement methods more directly. The correlation between the subjective art knowledge ratings and objective art knowledge (correct artist/style in

Part C of the VAIK art knowledge scale) was $r = .758, p < .001$ (Figure 11, Supplementary Material).

When comparing the subjective art knowledge measure with Part C of the VAIK, one is directly assessing the difference between subjective knowledge and objective knowledge of the exact same image. When comparing this correlation to the correlation with the AFS, we can clearly see that the two different measurements were more strongly correlated when they were both measured by use of the same image ($r = .758$ vs. $r = .618$). This means that there is an influence of measurement method, which shows evidence for criterion validity because subjective and objective knowledge were more strongly correlated when the influence of measurement method was reduced.

In addition, both the AFS and the VAIK art knowledge scale have an item measuring knowledge of impressionism; the AFS asks people to indicate how much they know about impressionism and Part C of the VAIK art knowledge scale includes the image *Bal du Moulin de La Gallette* by Renoir, one of the most famous Impressionists. This led to the opportunity to also do a more direct comparison between the AFS and the VAIK on this item. In this case we wanted to assess whether self-reported knowledge on the AFS would predict actual knowledge on the VAIK. That is, does stating that you know a lot about impressionism mean that you are also more likely to correctly identify the artist and style of a (very famous) impressionist artwork? To do this we first centered the impressionism item of the AFS and then ran two separate logistic regressions, one for style and one for artist. We found that self-reported knowledge on the AFS did predict being able to correctly identify the style, $\chi^2(1) = 131.35, p < .001$, Cox & Snell $R^2 =$

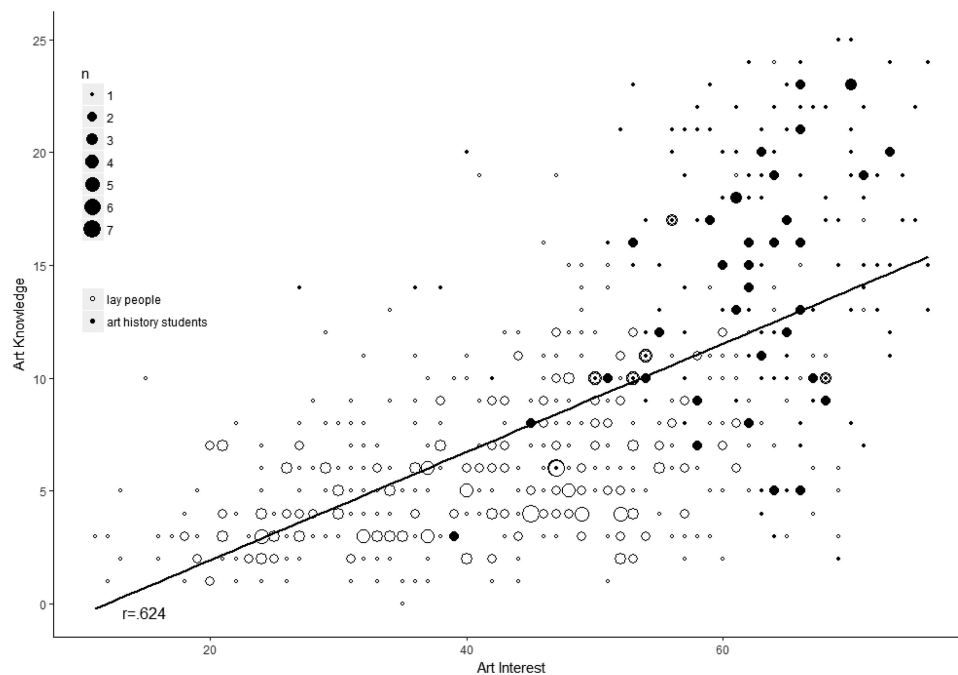


Figure 7. Scatterplot of the relationship between art knowledge and art interest. Like previous figures, this is a count plot, which is a type of scatter plot where the size of the dots represents the amount of people (or counts). Empty circles represent lay people, filled circles represent art history students. Dot-filled circles represent both groups simultaneously.

.193, $B = 1.33$, $p < .001$, $OR = 3.77$. However, it did less so on being able to correctly identify the artist, $\chi^2(1) = 23.83$, $p < .001$, Cox & Snell $R^2 = .038$, $B = .934$, $p < .001$, $OR = 2.55$. These results make it likely that people were overestimating their knowledge of art. The fact that self-reported knowledge predicted being able to correctly identify the style indicates that people were not purposely lying about how much they know about art; however, the fact that it did not predict correctly identifying the artist clearly indicates that people were overestimating themselves.

This interpretation is especially likely because the image used in the VAIK is an artwork made by Renoir, who is probably the most famous impressionist artist after Monet. It would therefore be highly unlikely that a person who knows a lot about impressionism would not be able to recognize the art of one of the style's most prominent artists. These findings also stress the need for a measure like the VAIK art knowledge scale that does not rely on self-report.

T3 We expected low correlations for BICB and the CAQ. As can be seen in Table 3, this is exactly what we found. There was one outlier on the CAQ with a score of 1090—the next highest score was 198. Scores on the CAQ do not follow a normal distribution, because in principle only a minority of individuals within a population will exhibit high levels of creative achievement (Carson et al., 2005). To have outliers in the data is therefore not unlikely or surprising, however the outlier influenced the results to a large extent. We therefore report the results both with and without the outlier. Neither the art interest nor the art knowledge scale of the VAIK showed a correlation higher than .30 with the BICB or the CAQ.

The highest correlations between the two VAIK scales and the other measures (excluding the AFS) were found with the visual art subscale of the CAQ. This is consistent, as creative achievement in the visual arts would be more related to knowing about art or being interested in art than achievement in another domain; our results support this idea. We do want to stress that the low correlation between the CAQ visual art and the VAIK scales does not mean that people who have high creative achievement in the visual arts do not have a high level of interest in art and knowledge about art. Rather, it would be the other way around, namely that there are any number of people who have a high level of interest in art and knowledge about art but do not have a high level of creative achievement in this domain. In sum, our results show evidence for validity by consistency in how the VAIK scales relate to other variables.

General Discussion

Our aim with the VAIK was to provide the field of empirical aesthetics with a good and validated measure of both art interest and art knowledge as well as a conceptualization of both concepts. This, we argue, is necessary to allow for a more comparable investigation of the influence of art interest and knowledge. It should furthermore be a useful tool for all researchers interested in these concepts. The primary purpose of measuring art interest and art knowledge is a better understanding of study participants and how these factors influence people engaging with art. The questionnaire does not aim to measure any specific field of interest in the arts nor does it measure specialized knowledge of any theme, artist, or period. The questionnaire is furthermore limited in a focus exclusively on visual art and does not cover knowledge or interest in other domains.

Another limitation is that it has a focus on Western art history and attendance at typical Western art settings. The questionnaire aims to measure variation in knowledge about art that is familiar to people who had a standard Western art education or frequently visit museums in a Western context. We assume that this type of knowledge could be gained independently from formal education in art, for example by visiting museums or by reading books on art. Because of the focus on Western art history the questionnaire is probably less suitable to be used in non-Western samples. Even though this is an open and empirical question, researchers should be aware of this issue when using the questionnaire in a different cultural context. Finally, the questionnaire is not meant as an assessment tool that would be appropriate to use in art history exams or the like.

Following the argument-based approach (Kane, 1992, 2006, 2009) we have shown (a) evidence that the questionnaire measures what it claims to measure by way of face validity and content validity when constructing the questionnaire and the consequent process of questionnaire development. The art interest scale of the VAIK covers a wide range of items that assess interest in art not just subjectively but also more objectively by asking for concrete behaviors such as the amount of museum visits. The art knowledge scale of the VAIK assesses different kinds of knowledge. Not only does it assess knowledge about art styles and artists but it also tests for knowledge about techniques and iconography. All of these would be included in a typical Western art education. In addition, we have consistently shown that the VAIK scales can discriminate between lay people and art experts (i.e., art history students and art professionals).

Table 3
Correlations Between the VAIK and the AFS, BICB, and CAQ

Measure	AFS	BICB	CAQ total	CAQ total without outlier	CAQ visual art subscale	CAQ visual art subscale without outlier
VAIK: Art interest	.559	.201	.075	.192	.291	.288
VAIK: Art knowledge	.618	.033	.041	.196	.253	.008
AFS		.158	.057	.293	.265	.063
BICB			.237	.258	.215	.232

Note. AFS = Aesthetic Fluency Scale (Smith & Smith, 2006); BICB = Biographical Inventory of Creative Behaviors (Batey, 2007); CAQ = Creative Achievement Questionnaire (Carson, Peterson, & Higgins, 2005). Bolded correlations were significant. All were significant at $p < .01$ except the correlation between Art interest and CAQ total, which was significant at $p < .05$.

We have shown (b) evidence that the test scores display adequate reliability. The reliability of the scores of the VAIK scales was taken into consideration when developing the questionnaire. The pretest provided preliminary evidence for reliability of scores on both scales of the VAIK, and the final validation study also showed reliability of scores on both subscales of the VAIK.

Finally, we have shown (c) evidence that the test scores display relationships with other variables in a manner congruent with its predicted properties. The pretest already provided preliminary evidence by showing a convergence between self-reported art knowledge and art knowledge as assessed by the VAIK art knowledge scale. In the validation study we presented additional evidence by replication of the previous findings as well as the use of the AFS, BICB and CAQ to establish criterion validity.

The evidence presented shows preliminary evidence for the validity of the use of the VAIK as a measure of art interest and art knowledge in research as well as reliability of the test scores. Nonetheless, as ever, ongoing validation efforts are recommended. Further validation efforts, for example, could investigate the VAIK in different cultural contexts (and different languages) as well as in different testing situations (e.g., lab vs. online).

We hope that the VAIK will prove a useful tool for all researchers interested in the concepts of art interest and art knowledge. As we have stressed in the introduction, we hope that the new measure inspires researchers to more clearly differentiate between interest in art and knowledge about art. We hope that research will become more comparable as a result of this distinction. In addition to this clearer conceptualization of these two dimensions, the VAIK also offers a more objective way to measure art knowledge. Our findings in the validation study indicate that self-report measures overestimate the knowledge that people have about art. The findings furthermore stress the importance of a better and more objective measure. In sum, we feel and hope that the measure will allow us to gain a better and more comprehensive understanding of how interest and knowledge influence people engaging with art.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *The standards for educational and psychological testing*. Washington, DC: Author.
- Aristotle. (1908). *The works of Aristotle translated into English* (W. D. Ross, Ed.). London, UK: Clarendon.
- Augustin, D., & Leder, H. (2006). Art expertise: A study of concepts and conceptual spaces. *Psychology Science*, 48, 135–156.
- Batey, M. (2007). *A psychometric investigation of everyday creativity*. Unpublished doctoral dissertation. University College, London.
- Belke, B., Leder, H., Harsanyi, G., & Carbon, C. C. (2010). When a Picasso is a "Picasso": The entry point in the identification of visual art. *Acta Psychologica*, 133, 191–202. <http://dx.doi.org/10.1016/j.actpsy.2009.11.007>
- Bell, C. (1914). *Art*. London, UK: Chatto & Windus.
- Berlyne, D. E., Ogilvie, J. C., & Parham, L. C. (1968). The dimensionality of visual complexity, interestingness, and pleasingness. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 22, 376–387. <http://dx.doi.org/10.1037/h0082777>
- Brieber, D., Leder, H., & Nadal, M. (2015). The experience of art in museums. *Empirical Studies of the Arts*, 33, 95–105. <http://dx.doi.org/10.1177/0276237415570000>
- Carroll, N. (2000). *Theories of art today*. Madison, WI: University of Wisconsin Press.
- Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. *Creativity Research Journal*, 17, 37–50. http://dx.doi.org/10.1207/s15326934crj1701_4
- Chamberlain, R., & Wagemans, J. (2015). Visual arts training is linked to flexible attention to local and global levels of visual stimuli. *Acta Psychologica*, 161, 185–197. <http://dx.doi.org/10.1016/j.actpsy.2015.08.012>
- Chamorro-Premuzic, T., & Furnham, A. (2005). Art judgement: A measure related to both personality and intelligence? *Imagination, Cognition and Personality*, 24, 3–24. <http://dx.doi.org/10.2190/U4LW-TH9X-80M3-NJ54>
- 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. <http://dx.doi.org/10.2190/EM.28.2.f>
- Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. London, UK: Routledge.
- Child, I. L. (1964). Observations on the meaning of some measures of esthetic sensitivity. *The Journal of Psychology*, 57, 49–64. <http://dx.doi.org/10.1080/00223980.1964.9916671>
- Chirumbolo, A., Brizi, A., Mastandrea, S., Mannetti, L., & Martens, A. (2014). 'Beauty is no quality in things themselves': Epistemic motivation affects implicit preferences for art. *PLoS ONE*, 9, e110323. <http://dx.doi.org/10.1371/journal.pone.0110323>
- Cupchick, G. C., & Gebotys, R. J. (1988). The search for meaning in art: Interpretive styles and judgments of quality. *Visual Arts Research*, 14, 38–50.
- Davies, P. J. E., Denny, W. B., Hofrichter, F. F., Jacobs, J. F., Roberts, A. S., & Simon, D. L. (2010). *Janson's history of art: The Western tradition* (8th ed.). London, UK: Pearson Education Limited.
- Dinno, A. (2012). *paran: Horn's test of principal components-factors*. Retrieved from <https://cran.r-project.org/package=paran>
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 399–412. <http://dx.doi.org/10.1111/bjop.12046>
- Elkins, J. (2004). *Pictures & tears: A history of people who have cried in front of paintings*. New York, NY: Routledge.
- Ericsson, K. A. (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*. New York, NY: Erlbaum.
- Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. New York, NY: Cambridge University Press.
- Fumaroli, M. (1994). Rome 1630: Entree en scene du spectateur [Rome 1630: The spectator enters on stage]. In O. Bonfait (Ed.), *Rome 1630. Il trionfo del pennello* [Rome 1630. The triumph of the brush] (pp. 53–82). Milan, Italy: Electa.
- Gombrich, E. (2006). *The story of art* (16th ed.). London, UK: Phaidon Press.
- Götz, K. O., Lynn, R., Borisy, A. R., & Eysenck, H. J. (1979). A new Visual Aesthetic Sensitivity Test: I. Construction and psychometric properties. *Perceptual and Motor Skills*, 49, 795–802. <http://dx.doi.org/10.2466/pms.1979.49.3.795>
- Graham, D. J., Hughes, J. M., Leder, H., & Rockmore, D. N. (2012). Statistics, vision, and the analysis of artistic style. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4, 115–123. <http://dx.doi.org/10.1002/wics.197>
- Graves, M. (1946). *Design judgement test*. New York, NY: The Psychological Corporation.
- Hayes, J. R. (1989). *The complete problem solver* (2nd ed.). Hillsdale, NJ: Erlbaum.

AQ: 10

- Hekkert, P., & Van Wieringen, P. C. W. (1996). Beauty in the eye of expert and nonexpert beholders: A study in the appraisal of art. *The American Journal of Psychology*, 109, 389. <http://dx.doi.org/10.2307/1423013>
- Iwawaki, S., Eysenck, H. J., & Götz, K. O. (1979). A new Visual Aesthetic Sensitivity Test (VAST): II. cross-cultural comparison between England and Japan. *Perceptual and Motor Skills*, 49, 859–862. <http://dx.doi.org/10.2466/pms.1979.49.3.859>
- Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527–535. <http://dx.doi.org/10.1037/0033-2909.112.3.527>
- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Washington, DC: American Council on Education/Praeger.
- Kane, M. (2009). Validating the interpretations and uses of test scores. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 39–64). Charlotte, NC: Information Age Publishing Inc.
- Lauring, J. O., Pelowski, M., Forster, M., Gondan, M., Pito, M., & Kupers, R. (2016). Well, if they like it . . . : Effects of social groups' ratings and price information on the appreciation of art. *Psychology of Aesthetics, Creativity, and the Arts*, 10, 344–359. <http://dx.doi.org/10.1037/aca0000063>
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95, 489–508. <http://dx.doi.org/10.1348/0007126042369811>
- 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. <http://dx.doi.org/10.1080/02699931.2013.870132>
- Liu, F.-J. (1990). Critique of three tests of aesthetic judgment: Maitland Graves Design Judgment Test; The Meier Art Tests: I, Art Judgment; and The Meier Art Tests: II, Aesthetic Perception. *Visual Arts Research*, 16, 90–99.
- Locher, P. J., Smith, J. K., & Smith, L. F. (2001). The influence of presentation format and viewer training in the visual arts on the perception of pictorial and aesthetic qualities of paintings. *Perception*, 30, 449–465. <http://dx.doi.org/10.1068/p3008>
- Locher, P., Smith, L. F., & Smith, J. K. (1999). Original paintings versus slide and computer reproductions: A comparison of viewer responses. *Empirical Studies of the Arts*, 17, 121–129. <http://dx.doi.org/10.2190/R1WN-TAF2-376D-EFUH>
- Martindale, C. (1988). Aesthetics, psychobiology, and cognition. In F. H. Farley & R. W. Neperud (Eds.), *The foundations of aesthetics, art, & art education* (pp. 7–42). New York, NY: Praeger.
- McKibben, W. B., & Silvia, P. J. (2017). Evaluating the distorting effects of inattentive responding and social desirability on self-report scales in creativity and the arts. *The Journal of Creative Behavior*, 51, 57–69. <http://dx.doi.org/10.1002/jocb.86>
- Meier, N. C. (1942). *The Meier Art Tests: I, Art judgement*. Iowa City, IA: Bureau of Educational Research and Service, State University of Iowa.
- Meier, N. C. (1963). *The Meier Art Tests: II, Aesthetic perception*. Iowa City, IA: Bureau of Educational Research and Service, State University of Iowa.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). Washington, DC: American Council on Education. **AQ: 11**
- Myszkowski, N., Çelik, P., & Storme, M. (2016). A meta-analysis of the relationship between intelligence and visual “taste” measures. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication.
- Myszkowski, N., & Zenasni, F. (2016). Individual differences in aesthetic ability: The case for an aesthetic quotient. *Frontiers in Psychology*, 7, 750. <http://dx.doi.org/10.3389/fpsyg.2016.00750>
- Olivier, L. A. (1976). *Curieux, amateurs and connoisseurs: Laymen and the fine arts in the Ancien Regime*. Baltimore, MD: John Hopkins University Press.
- Pelowski, M. (2015). Tears and transformation: Feeling like crying as an indicator of insightful or “aesthetic” experience with art. *Frontiers in Psychology*, 6, 1006. <http://dx.doi.org/10.3389/fpsyg.2015.01006>
- 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. <http://dx.doi.org/10.3389/fnhum.2016.00160>
- Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research*. Retrieved from <https://cran.r-project.org/package=psych>
- Rosenberg, R., & Klein, C. (2015). The moving eye of the beholder. Eye-tracking and the perception of paintings. In J. Huston, M. Nadal, F. Mora, L. F. Agnati, & C. J. Cela-Conde (Eds.), *Art, aesthetics and the brain* (pp. 79–96). Oxford, UK: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199670000.003.0005>
- Sibley, F. (1959). Aesthetic Concepts. *The Philosophical Review*, 68, 421–450.
- Silvia, P. J., Wigert, B., Reiter-Palmon, R., & Kaufman, J. C. (2012). Assessing creativity with self-report scales: A review and empirical evaluation. *Psychology of Aesthetics, Creativity, and the Arts*, 6, 19–34. <http://dx.doi.org/10.1037/a0024071>
- Smith, L. F., & Smith, J. K. (2006). The nature and growth of aesthetic fluency. In P. Locher, C. Martindale, & L. Dorfman (Eds.), *New directions in aesthetics, creativity and the arts* (pp. 47–58). Amityville, NY: Baywood.
- Specker, E., Tinio, P. P. L., & van Elk, M. (2017). Do you see what I see? An investigation of the aesthetic experience in the laboratory and museum. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 265–275. <http://dx.doi.org/10.1037/aca0000107>
- Winston, A. S., & Cupchick, G. C. (1992). The evaluation of high art and popular art by naive and experienced viewers. *Visual Arts Research*, 18, 1–14.

Received January 26, 2018

Revision received September 1, 2018

Accepted September 5, 2018 ■

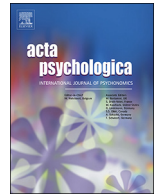
2.1.2

Specker, E., Leder, H., Rosenberg, R., Hegelmaier, L. M., Brinkmann, H., Mikuni, J., & Kawabata, H. (2018). The universal and automatic association between brightness and positivity. *Acta Psychologica*, 186, 47–53. <https://doi.org/10.1016/j.actpsy.2018.04.007>

My Contribution:

Being the first author, I came up with the core concept of the paper.

LMH, JM, HB, and HK assisted in data collection. I was responsible for all of the analyses and for the literature review, the paper is based on. I was responsible for the design of the study but received feedback from HL. I was the main writer of the article but received feedback from all co-authors on different versions of the manuscript.



The universal and automatic association between brightness and positivity[☆]

Eva Specker^{a,*}, Helmut Leder^a, Raphael Rosenberg^b, Lisa Mira Hegelmaier^a, Hanna Brinkmann^b, Jan Mikuni^c, Hideaki Kawabata^c

^a Faculty of Psychology, Department of Basic Psychological Research and Research Methods, University of Vienna, Austria

^b Faculty of Historical and Cultural Studies, Department of Art History, University of Vienna, Austria

^c Faculty of Psychology, Keio University, Japan

ARTICLE INFO

Keywords:

Color
Color association
Brightness
Positivity
Cross-cultural psychology

ABSTRACT

The present study investigates the hypothesis that brightness of colors is associated with positivity, postulating that this is an automatic and universal effect. The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) was used in all studies. Study 1 used color patches varying on brightness, Study 2 used achromatic stimuli to eliminate the potential confounding effects of hue and saturation. Study 3 replicated Study 2 in a different cultural context (Japan vs. Austria), both studies also included a measure of explicit association. All studies confirmed the hypothesis that brightness is associated with positivity, at a significance level of $p < .001$ and Cohen's D varying from 0.90 to 3.99. Study 1–3 provided support for the notion that this is an automatic effect. Additionally, Study 2 and Study 3 showed that people also have an explicit association of brightness with positivity. However, as expected, our results also show that the implicit association was stronger than the explicit association. Study 3 shows clear support for the universality of our effects. In sum, our results support the idea that brightness is associated with positivity and that these associations are automatic and universal.

1. Introduction

The notion that brightness is positive and darkness is negative can be traced back to mythical time. The famous story of Theseus and the Minotaur features the agreement between Theseus and his father, Aegeus, to hoist a white flag for success (defeat of the Minotaur) and a black one for defeat. As the art historian Gombrich (1963) mentions: “Surely it was no accident that black rather than white was the sign agreed upon for failure. Black seems to us a more ‘natural’ sign for grief, and white for a ‘brighter’ mood: and even though we know that cultural conventions also play their part (and that black is not everywhere the color of mourning), the correlation makes sense in terms of expressiveness (p.61).”

Historically there has been great interest in studying the aesthetic effects of different colors in literature on art as well as in psychology (Bullough, 1907; De Camp, 1917; Fechner, 1876). Artists and art historians often use the term “effect” or “aesthetic effect” (e.g. Gombrich, 1963; Itten, 1961; Kandinsky, 1911; Schwitters, 1918), whereas in psychology it is more common to speak of associations. Basically both terms refer to the same experience. Kandinsky (1946) for example,

specified different associations for the color black: “Black is something extinguished like a burned pyre, something immobile, corpse-like, which has no connection with any occurrences, and accessible to all things. It is like the silence of the body after death, the end of life (p.68).”

Even though early psychological studies (e.g., Alexander & Shansky, 1976; Bullough, 1907; De Camp, 1917; Mogensén & English, 1926; Monroe, 1925; Pinkerton & Humphrey, 1974; Tinker, 1938) show a great interest in color associations, the methodological quality is poor (Gelineau, 1981; Valdez & Mehrabian, 1994). More recent work has neglected associations somewhat and has a greater focus on color preferences (Hurlbert & Ling, 2007; Osborne et al., 2016; Palmer & Schloss, 2010; Wei, Houser, Allen, & Beers, 2014) despite this there have been studies conducted that looked at associations (Adams & Osgood, 1973; Albertazzi et al., 2013; Chen, Tanaka, & Watanabe, 2015; Dael, Persegues, Marchand, Antonietti, & Mohr, 2016; Lazreg & Mullet, 2001; Saito, 1996).

Most color conceptualizations define color as having 3 features: hue (i.e. wavelength), saturation (i.e. vividness, with lower saturation containing more grey), and brightness (i.e. black-to-white quality).

[☆] The writing of this paper was supported by a grant to RR and HL by Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF: Project number: CS15-036). We wish to acknowledge Simon Schreiberlmayr for his help with data collection and programming.

* Corresponding author at: Faculty of Psychology, Department of Basic Psychological Research and Research Methods, University of Vienna, Liebiggasse 5, A-1010 Vienna, Austria. E-mail address: eva.specker@univie.ac.at (E. Specker).

<https://doi.org/10.1016/j.actpsy.2018.04.007>

Received 9 May 2017; Received in revised form 8 March 2018; Accepted 16 April 2018
0001-6918/ © 2018 Elsevier B.V. All rights reserved.

Psychologists studying color and its associations, have almost exclusively focused on the effects of hue. In addition, results vary greatly between as well as within studies. Variation in associations connected to hue are often due to cultural variation, such as the influence of language on color discrimination (Regier & Kay, 2009; Winawer et al., 2007), cultural differences in affective/symbolic meaning and associations related to different hues (Adams & Osgood, 1973; Saito, 1996). In contrast, there seems to be less (cross-cultural) variation in the influence of brightness. Brightness seems to be associated with positivity across cultures (Adams & Osgood, 1973; Saito, 1996; Valdez & Mehrabian, 1994). In relation to the emotional effect of color, Gao et al. (2007) even concluded that influences of hue and cultural background are limited, whereas brightness was seen as an important influence. It has to be mentioned though that the majority of these studies used color words (“Blue”, “Red”, etc.) rather than color stimuli (i.e. an image of a color), however, these studies are not strictly about hue. Nevertheless, this research suggest that an association between brightness and positivity exists. There are only a handful of studies (Hemphill, 1996; Lakens, Fockenberg, Lemmens, Ham, & Midden, 2013; Lakens, Semin, & Foroni, 2012; Meier, Fetterman, & Robinson, 2015; Meier, Robinson, & Clore, 2004) that tested this association directly. In the studies by Meier et al. (2004, 2015) positive and negative words were printed in either black or white font. In the study by Lakens et al. (2012) participants had to indicate the correct translation of a word into Chinese by selecting a Chinese character printed in either black or white. Both found an association between brightness and positivity. The other two studies, Lakens et al. (2013) and Hemphill (1996), didn't use font colors (black vs. white) as the basis for associations.

The study by Hemphill (1996) asked participants to associate emotional responses with a various range of hues, in this case people tended to associate bright colors (e.g., white, pink) with positive emotions and dark colors (e.g. black, brown) with negative emotion. The Lakens et al. (2013) study looked at overall brightness of naturalistic pictures. They found an overall “brightness bias”, meaning that brighter pictures were evaluated as more positive.

We use color patches as stimuli in the current study. Using different stimuli can provide evidence for the stimulus independence of the effect, that is, the notion that the effect is there and can consistently be found when using a variety of stimuli.

One remaining question is why effects of brightness seem to be consistent even across cultures when effects of hue seem more likely to vary. A way to understand this would be through the structure of our visual system. Our visual system employs two pathways, a chromatic one that processes color contrasts and an achromatic one that processes luminance contrasts (for a more detailed account see Gegenfurtner & Kiper, 2003). While the development of the achromatic response appears to be stable in infants after a few months of age (Crognale, 2002) the development of the chromatic response continues to change from birth until puberty (Boon, Suttle, & Dain, 2007). Furthermore, Page and Crognale (2005) suggest that both pathways age differently, with the chromatic pathways being more significantly affected by age-related changes. These studies indicate that the chromatic system may be more sensitive to cultural influence than the achromatic system.

This earlier biological onset of stability of achromatic responses might be related to cultural development in language. Anthropological research has shown that in Indo-European cultures (old English for example) there used to be a brightness-oriented color paradigm. In such a paradigm, color discrimination is done based on brightness rather than hue (Biggam, 2014; Borg, 2014). In this case all the bright colors would group together (bright blue, bright yellow, bright red) and all the dark colors would group together (dark blue, dark yellow, dark red). Rather than grouping based on hue where, for example, bright and dark blue together. Later, many of these languages shifted to a hue-based paradigm of color discrimination, like we all know from Modern English today. Interestingly, there are still modern languages such as Russian (Winawer et al., 2007) that use a brightness-oriented color

discrimination paradigm (Biggam, 2014; Borg, 2014). It is notable, however, that large-scale studies such as the World Color Survey (Kay & Cook, 2015) have found that industrial cultures have very similar chromatic discrimination.

Following this reasoning, we hypothesize that the effects of brightness should be universal as well as automatic. If brightness triggers an evolutionarily earlier developed system then consistent effects of brightness should be found across cultures and these effects should be automatic. Based on previous research, one potential effect of brightness seems to be that brightness is associated with positivity. Saito (1996) found cross-cultural similarity for a preference of white amongst East Asians. Valdez and Mehrabian (1994) found that white (in comparison to grey and black) was seen as more positive in an American sample. Similarly, Adams and Osgood (1973) found a high-level agreement on ratings of black. Research by Meier et al. (2004, 2015) and Lakens et al. (2012, 2013) have found support for the idea that brightness and positivity are associated.

In this study, we therefore want to test the hypothesis that brightness is associated with positivity and, in addition, test the claim for universality and automaticity. To test for automaticity we used an Implicit Association Test (hereafter: IAT, Greenwald, McGhee, & Schwartz, 1998) to measure the implicit association between brightness and positivity. The IAT is a well-known measure of social attitudes and has also been used in previous studies to investigate color associations (e.g. Soriano & Valenzuela, 2009). In Study 2 and 3 we additionally included an explicit measure. In Study 1 we use chromatic stimuli in order to see if the brightness-positivity association can be found in chromatic stimuli. In order to exclude alternative explanations based on hue or saturation we use achromatic stimuli in Study 2 and Study 3. All studies used images to represent the color stimuli. In order to test for universality we used an Austrian sample in Study 1 and 2 and a Japanese sample in Study 3.

2. Study 1

2.1. Participants

We used a small to medium effect size in our power analysis. This indicated that we needed 71 participants to achieve a power of 0.80 with an error rate probability of 0.05. We therefore collected data from 71 participants. All participants were psychology students of the University of Vienna and received course credit for their participation. Fifty-three participants were female and the mean age was 22.11 (SD = 3.5). The experiment was carried out in accordance with the Declaration of Helsinki and in a procedure which was approved by the local ethical committee of the University of Vienna.

2.2. Method

We used a within-subjects design. All participants completed an IAT task and then reported demographic information. The IAT is a reaction based task where two concepts (in our case bright and positive) either share a response key (congruent block) or have different response keys (incongruent block). The idea is that if two concepts are implicitly associated reaction time will be faster in congruent blocks as opposed to incongruent blocks. If there is no implicit association, people should perform equally well in both blocks. For a detailed description of the task see Table 1. The IAT is a widely used test, it is robust and easy to administer, and has good reliability compared to other implicit measures (Fazio & Olson, 2003; Nosek, Greenwald, & Banaji, 2005). Explicit ratings of the positivity of the stimuli were also collected but due to a data logging error cannot be reported here.

Table 1
Implicit association test task sequence.

Block	Type	Amount of trials	Items assigned to “E” key response	Items assigned to “I” key response
1	Practice	20	Bright colors	Dark colors
2	Practice	20	Positive adjectives	Negative adjectives
3 (6)	Practice	20	Bright colors + positive adjectives	Dark colors + negative adjectives
4 (7)	Test	40	Bright colors + positive adjectives	Dark colors + negative adjectives
5	Practice	20	Dark colors	Bright colors
6 (3)	Practice	20	Dark colors + positive adjectives	Bright colors + negative adjectives
7 (4)	Test	40	Dark colors + positive adjectives	Bright colors + negative adjectives

Note: the numbers between brackets () indicate the counterbalancing of the blocks. That is to say for half of the people block 3 was block 6 and block 4 was block 7.

2.3. Materials & procedure

2.3.1. IAT task

We used the standard IAT design, see Table 1 for a detailed description of the task. We applied the standard version of the IAT that includes error feedback for incorrect responses. In this case, the error message remains on the screen until the correct key is pressed. Key assignments were visible throughout the task as recommended by Gawronski, Deutsch, and Banse (2011). We opted to counterbalance the task in order to be able to eliminate block order as potential confound. Therefore, half of the participants followed the structure as outlined in Table 1 and for half of the participants block 6 and 7 were switched with 3 and 4, indicated by the brackets in Table 1.

2.3.2. Stimuli

We created 18 color patches differing on brightness as stimuli for the IAT task. All color patches were blue. Nine color patches represented the bright category and nine the dark category. The bright category ranged from 10 to 30% brightness with increasing steps of about 2–3%. The dark category ranged from 70 to 90% brightness with increasing steps of about 2–3%. All images can be found here: <https://osf.io/gsnf4/>. Color can be described in different models, some models are perceptually uniform and resemble human color perception, such as the Lab (Luminance, Red-Green, Blue-Yellow) system (CIE; McLaren, 1976), others are perceptually not uniform and mostly depend on hardware used to produce the color such as the RGB (Red, Blue Green) system. For completeness we report both LAB and RGB values in Table 2. As a manipulation check all stimuli were measured by use of a colorimeter (i1Xtreme by X-Rite) this was done by presenting the stimuli full screen on a CRT monitor and then taking the measurements. Measurement was carried out on the monitor that was used by the participant during the test. As noted above all values can be found in Table 2. For this manipulation check we focused on the luminance measures of the LAB color model. This showed that the dark category was lower in luminance (and therefore darker) than the bright category, $t(16) = -15.49$, $p < .001$. All stimuli were presented against a grey background (RGB: 128, 128, 128; LAB: 64.3, -4.4, -23.1). To ensure consistency of color presentation across subjects the IAT task was conducted in a windowless room with constant lighting conditions and on a CRT Monitor.

To represent the positive and negative category, we selected words from the Berlin Affective Word List Reloaded (BAWL-R, Vö et al., 2009), the words used can be found in Table 3. The BAWL-R is a cross-validated resource of affective words in German that have been rated for valence and arousal. In the BAWL-R valence is rated on a 7-point scale from -3 (very negative) through 0 (neutral) to +3 (very positive). Arousal on the other hand is measured on a 5-point scale ranging from 1 (low arousal) to 5 (high arousal). The words included in the BAWL-R are a combination of emotion-laden and emotion-words. Our selection consisted of emotion-laden words with the possible exception of “sad (traurig)”. These two word types are processed differently (e.g. Kazanas & Altarriba, 2015; Sutton & Altarriba, 2016). For example, Kazanas and Altarriba (2015) showed that priming effects are greater for emotion-

Table 2
Image stimuli.

Stimuli brightness	Stimulus categorization	R B G	L A B
10%	Dark	11, 0, 51	1.2, 8.9, -31.5
13%	Dark	14, 0, 66	5.2, 22.9, -48.1
15%	Dark	17, 0, 77	7.2, 27.2, -58.7
17%	Dark	19, 0, 87	11.9, 31.8, -60.8
20%	Dark	22, 0, 102	15.9, 37.3, -70.6
23%	Dark	25, 0, 117	20.2, 43.1, -81.8
25%	Dark	28, 0, 128	23.3, 46.7, -88.4
27%	Dark	30, 0, 138	25.9, 49.6, -93.8
30%	Dark	33, 0, 153	29.9, 54.5, -102.9
70%	Bright	135, 102, 255	73.9, 61.7, -132.5
73%	Bright	142, 117, 255	79.1, 54.6, -124.5
75%	Bright	155, 128, 255	82.3, 49.5, -118.9
77%	Bright	163, 138, 255	85.8, 42.4, -112
80%	Bright	175, 153, 255	91.5, 34.9, -103.4
83%	Bright	187, 168, 255	96.4, 28.0, -94.4
85%	Bright	195, 179, 255	98.7, 22.1, -86.0
87%	Bright	203, 189, 255	101.1, 18.2, -80
90%	Bright	215, 204, 255	103.6, 11.3, -70.2

Table 3
Word stimuli.

Stimulus word German (English translation)	Stimulus categorization	Valence	Arousal
Asozial (<i>asocial</i>)	Negative	-2.4	2.5
Bankrott (<i>bankrupt</i>)	Negative	-2.3	2.8
Traurig (<i>sad</i>)	Negative	-2.1	2.8
Machtlos (<i>powerless</i>)	Negative	-2	3
Einsam (<i>lonely</i>)	Negative	-2.2	3.1
Liebslos (<i>loveless</i>)	Negative	-2.3	3.27
Trostlos (<i>hopeless</i>)	Negative	-2.1	3.3
Verlogen (<i>mendacious</i>)	Negative	-2.3	3.5
Leblos (<i>lifeless</i>)	Negative	-2.3	3.5
Beliebt (<i>loved</i>)	Positive	2.4	2.8
Vital (<i>vital</i>)	Positive	2.2	2.9
Toll (<i>great</i>)	Positive	2	3.05
Lebendig (<i>lively</i>)	Positive	2.5	3.05
Brilliant (<i>brilliant</i>)	Positive	2.2	3.16
Perfekt (<i>perfect</i>)	Positive	2.4	3.16
Mutig (<i>courageous</i>)	Positive	2.2	3.3
Super (<i>super</i>)	Positive	2.5	3.3
Grandios (<i>grandiose</i>)	Positive	2.2	3.4

word pairs than emotion-laden word pairs. Since the majority of the words used are emotion-laden we do not presume that this should have influenced our results to a great extent. However, we do see the mixture

as a limitation and would recommend further research to limit itself to only one type of emotion words. The words selected to be in the positive and negative category differed in valence, $t(16) = 63.41$, $p < .001$, but not in arousal, $p = .77$. We matched for word length, on average negative words had a length of 7.22 characters and positive words of 6.44 characters, this difference in length was not significant, $p = .244$. Table 3 summarizes all stimuli information pertaining to the word stimuli.

2.3.3. Procedure

Upon arrival in the lab participants received and signed the informed consent form. They were instructed by the experimenter to start the IAT task when they were finished with the consent form. After completing the IAT task participants reported demographic information (age and gender), which concluded the experiment.¹ The block order of the IAT was counterbalanced, this is also designated in Table 3, here it shows that for half of the people block 3 was block 6 and that block 4 was block 7.

2.4. Results

The average response time was 799.8 ms with a standard deviation of 916.5 ms. A D-score was calculated as recommended by Greenwald, Nosek, and Banaji (2003) for every person. The formula essentially deduces the mean reaction time of the incongruent practice block from the mean reaction time of the congruent practice block and divide that by the pooled standard deviation of the practice blocks. It then does the same for the test blocks, finally it takes the mean of these two. A nice illustration of the formula can be found in Pavlović & Marković (2012, p. 384). The D-score ranges from -1.5 to $+1.5$. Typically, D-scores of > 0.65 are seen as large effects. Since the IAT is a relative measure a positive D-score indicates a larger association between brightness and positivity than brightness and negativity. As a consequence this also means a larger association between darkness and negativity than darkness and positivity. This indicated a large effect with the mean D-score in our sample being 0.75 ($SD = 0.44$). In a second step a one-sided one sample t -test was conducted. This showed a significant difference from 0 in the expected (positive) direction, $t(70) = 14.57$, $p < .001$, Cohen's $D = 1.73^2$, $CI [1.36, 2.10]$. To ensure that the found effect was not due to semantic relationships with any of the words used in the IAT task, we excluded trials that included the word “brilliant”. This exclusion led to a mean D-score of 0.75 ($SD = 0.44$), as can be seen this is identical to the results including the “brilliant” trials. For completeness we also conducted a one-sided one sample t -test, in this case the t -value changed from 14.57 to 14.56 none of the other statistics were affected.

2.5. Discussion

Our results indicated that people showed a strong implicit association between brightness and positivity which supports our hypothesis. However, due to the fact that we did not perfectly control for hue variation when varying brightness it is possible that the effects found are due to hue or saturation, which in this case varied structurally with brightness as can be seen from Table 2, rather than brightness. To exclude the alternative hypotheses that the effects are caused by hue and/or saturation rather than brightness, we used achromatic stimuli in Study 2.

¹ Participants also completed an explicit rating task, unfortunately, due to logging problems this data was not stored correctly and therefore will not be discussed here.

² In this case and all following instances Cohen's D and its confidence interval is calculated by using the method outlined here: <http://daniellakens.blogspot.co.at/2014/06/calculating-confidence-intervals-for.html>.

3. Study 2

3.1. Participants

Based on the findings of Study 1, we updated our estimate of the effect size from a small to medium effect (used in Study 1) to a medium estimated effect (0.5) for our power analysis. Even though this effect would be 3 times as small as the effect found in Study 1, we decided to be conservative in our estimation in case Study 1 reflected a Type 2 error. This power analysis indicated that 36 participants should be tested to achieve a power of 0.90 with an error rate probability of 0.05 . We therefore collected data from 36 participants. All participants were psychology students of the University of Vienna and received course credit for their participation. Twenty-nine participants were female and the mean age was 21.06 ($SD = 3.3$). The experiment was carried out in accordance with the Declaration of Helsinki and in a procedure which was approved by the local ethical committee of the University of Vienna.

3.2. Method, materials and procedure

The method, materials, and procedure were identical to Study 1 with the exception of the color stimuli. Rather than having variations of blue we constructed achromatic stimuli. These employed a ratio from black-to-white. The bright category consisted of 8 color patches which were filled 10 or 20% black (with 90 or 80% white respectively). The dark category consisted of 8 color patches which were filled 10 or 20% white (with 90 or 80% black respectively). As in Study 1, explicit ratings were also collected, in this case the data logging error was solved and this data could be analyzed. In the explicit rating task people were asked to rate each stimulus on positivity. People were asked: “How positive is this color patch?” and answered on a 7 point scale ranging from 1 (negative) to 7 (positive). The IAT and the explicit rating task were counterbalanced. Meaning that half of the participants first completed the IAT and then gave explicit ratings and the other half first gave explicit ratings and then completed the IAT. This was done in line with the recommendations made by Greenwald et al. (2003). Even so, we checked for potential task order effects but did not find a significant effect, $t(34) = -1.598$, $p = .119$. Again, the IAT task and the explicit rating task were conducted in a windowless room with constant lighting conditions on a CRT monitor.

3.3. Results

The average response time was 1231.5 ms with a standard deviation of 1543.8 ms. As in Study 1, a D-score was calculated for each participant. This indicated a large effect with the mean D-score in our sample being 0.75 ($SD = 0.26$). In a second step, a one-sided one sample t -test was conducted. This showed a significant from 0 in the expected (positive) direction, $t(35) = 17.19$, $p < .001$, Cohen's $D = 2.86$, $CI [2.12, 3.60]$. As in Study 1 we wanted to ensure that the found effect was not due to semantic relationships with any of the words used in the IAT task, thus we excluded trials that included the word “brilliant”. As in Study 1 this did not affect any of the above reported statistics with the exception of the t -value that changed from 17.19 to 17.24 . For the explicit association, a difference score was created by subtracting the standardized mean rating of the dark category from the standardized mean rating of the bright category. We then conducted a one-sided one sample t -test over the difference score. This showed a significant difference from 0 in the expected (positive) direction, $t(35) = 5.37$, $p < .001$, $CI [0.50, 1.28]$, Cohen's $D = 0.90$. Finally, we computed the correlation between implicit and explicit association. As can be seen in Fig. 1, there was no correlation between implicit and explicit association, $r = -0.140$, $p = .207$.

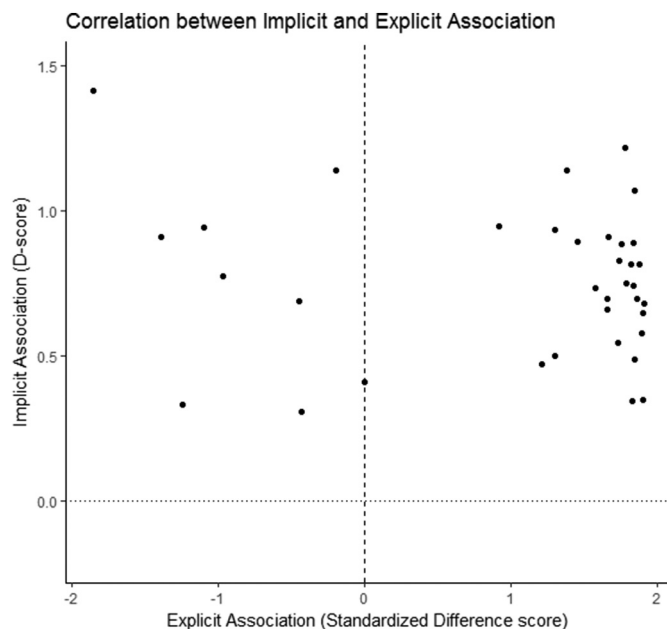


Fig. 1. Scatterplot of the correlation between Explicit and Implicit (D score of the IAT) rating. To aid interpretation lines are drawn on the 0 point of both axis. Because both measures represent a difference score, a positive score, thus above the 0 point on the horizontal axis and right from the 0 point on the vertical axis, represents an association between brightness and positivity.

3.4. Discussion

Our results indicated that people have a strong implicit association between brightness and positivity, which supports our hypothesis. In addition, our results indicated that people show a strong explicit association which, again, supports our hypothesis. Finally, as hypothesized, the implicit association between brightness and positivity was stronger than the explicit association. No correlation was found between the explicit and implicit association. This suggests that the implicit and explicit associations represent two separate cognitive processes that do not necessarily lead to the same outcome. This is in line with previous research, that found that a two-factor solution in which implicit and explicit associations were represented by one factor each fitted better than a one-factor solution (Greenwald & Farnham, 2000; Nosek, 2005; Nosek & Smyth, 2007). Another possibility for the absence of a correlation may be the low level of variance in the current data.

Study 1 and 2 together make a strong case for the existence of an association between brightness and positivity, implicitly as well as explicitly. However, these studies do not allow us to test the hypothesis that these effects are universal and would be also present in other cultures. In order to test the hypothesis that the association between brightness and positivity is universal we repeated Study 2 with a Japanese sample in Study 3.

4. Study 3

4.1. Participants

Based on the findings of Study 1 and 2 we again updated our estimated effect size, we now estimated a large effect (0.8) in our power analysis. Even though this effect is still considerably smaller than the effects found in Study 1 and 2, we again decided to be conservative in our estimation especially to account for potential cultural variation in the size of the effect. This power analysis indicated that we needed 15 participants to achieve a power of 0.90 with an error rate probability of 0.05. We therefore collected data from 15 participants. All participants were psychology students of the Keio University in Tokyo and received

1000 yen for their participation. Five participants were female and the mean age was 23.67 (SD = 2.7). The experiment was carried out in accordance with the Declaration of Helsinki and in a procedure which was approved by the local ethical committee of Keio University.

4.2. Method, materials and procedure

All methods, materials, and procedure were identical to Study 2 including the counterbalancing of task order. We checked for potential task order effects and found no effect of task order, $t(13) = 1.43$, $p = .176$. The emotion words were translated by a professional translator from German to Japanese and checked by the Japanese authors (Author 6 and 7). This approach was taken since there is, to our knowledge, no Japanese equivalent of the BAWL-R. The limitation of this approach is that the emotionality of the words used may vary dependent on culture. However, we assume that even though this leads to a lack of experimental control, the words should still (in general) represent the categories of positivity and negativity adequately. Again, the IAT task and the explicit rating task were conducted in a windowless room with constant lighting conditions. Due to the use of different labs the conditions could not be identical. However, due to the achromatic nature of the stimuli employed in Study 2 and Study 3 we feel that it is likely that the viewing conditions were near identical.

4.3. Results

The average response time was 1310.3 ms with a standard deviation of 2134.7 ms. As in Study 1 and 2, a D-score was calculated for each participant. This indicated a large effect, with the mean D-score in our sample being 1.02 (SD = 0.26). In a second step a one-sided one sample t -test was conducted. This showed a significant difference from 0 in the expected (positive) direction, $t(14) = 15.47$, $p < .001$, Cohen's $D = 3.99$, CI [2.44, 5.53]. As in the previous studies we wanted to ensure that the found effect was not due to semantic relationships with any of the words used in the IAT task. In this case, two words had a “brilliant” connotation: “素敵な” which was the translation for the German “brilliant” but also “素晴らしい” which was the translation of “grandiose”. This led to a minimal reduction of the mean D-score from 1.02 to 1.00 and a minimal increase in the standard deviation from 0.26 to 0.27. It did not influence the results of the t -test with the exception of an increase in the t -value from 15.47 to 14.52. For the explicit association a difference score was created the same way as in Study 2, by subtracting the standardized mean rating of the dark category from the standardized mean rating of the bright category. We then conducted a one-sided one sample t -test over the difference score. This showed a significant difference from 0 in the expected (positive) direction, $t(14) = 5.92$, $p < .001$, Cohen's $D = 1.53$, CI [0.76, 2.27]. We also conducted a two-sided t -test in which we compared the D-score of Austrians (Study 2) with that of the Japanese participants of this study. The t -test was significant, $t(49) = -3.317$, $p < .05$, Cohen's $D = 0.47$, CI [0.175, 0.759]. This indicated that Japanese ($M = 1.02$) have a stronger implicit association than Austrians ($M = 0.75$). Finally, we computed the correlation between implicit and explicit association. As can be seen in Fig. 2, there was no correlation between implicit and explicit association, $r = 0.210$, $p = .235$.

4.4. Discussion

Our results indicated that people show a strong implicit association between brightness and positivity. In addition, our results indicated that people show a strong explicit association. Finally, the implicit association between brightness and positivity was stronger than the explicit association. All of these results support our hypotheses. No correlation was found between explicit and implicit association. These results are very similar to the results obtained in Study 2. Therefore these results clearly support the notion of universality.

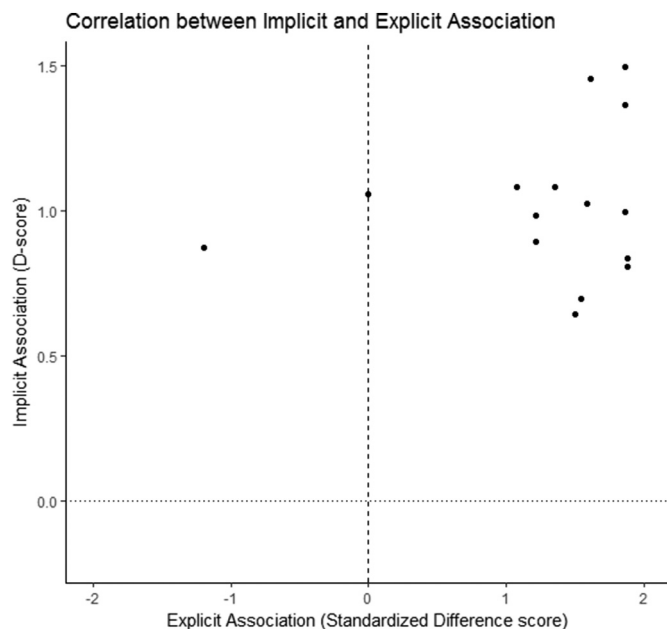


Fig. 2. Scatterplot of the correlation between Explicit and Implicit (D score of the IAT) rating. To aid interpretation lines are drawn on the 0 point of both axis. Because both measures represent a difference score, a positive score, thus above the 0 point on the horizontal axis and right from the 0 point on the vertical axis, represents an association between brightness and positivity.

5. Conclusions

The results of three studies show a clear support for the notion that the association between brightness and positivity is strong, automatic and universal. Due to the lack of control for hue variation when varying brightness in Study 1, alternative hypotheses that the effects are caused by hue and/or saturation rather than brightness could not be excluded based on this study alone. However, the use of achromatic stimuli in Study 2 and 3 suggest that the effects are due to brightness. In addition, Study 2 and 3 show that these effects are also present in more elaborate and conscious processing. The lack of a correlation between explicit and implicit association suggests that these are two separate processes. However, as noted, the absence of a correlation may also be due to very small variance in the current data.

It is worth mentioning that the effect size of the explicit association was markedly smaller than the effect size of the implicit association. This suggests that the association is mainly due to automatic processes and that when more complex processing is involved the association is influenced. Differences between explicit and implicit association are often reported in research on stereotypes using an IAT paradigm or another implicit measure (Dunton & Fazio, 1997; Gawronski, Ehrenberg, Banse, Zukova, & Klauer, 2003; Gawronski & LeBel, 2008). We refer the interested reader to Gawronski and De Houwer (2014) for a detailed overview of the IAT in relation to other implicit measures. In the case of research on stereotypes, findings are often explained in terms of social desirability. People don't want to be racist and therefore claim not to have a negative association with a marginalized group such as African-Americans. However, when using an implicit measure people cannot influence the results (as much as in the explicit measure) and then the implicit association is found. Within the current context it seems unlikely that participants felt a need to hide their association with brightness and positivity because of social desirability, since this is not typically an attitude that is socially undesirable.

Another possibility is that in the explicit rating task people confounded “positivity” with “liking”. In other words, if someone saw an image they liked that was dark they may have perceived it as more positive, similarly, when they saw a bright image they disliked they

may have perceived it as more negative. This influence of preference on positivity judgements in the explicit association seems to us a more likely explanation for the difference found. As noted in the discussion of Study 2, this finding is in line with previous research that found that a two-factor solution in which implicit and explicit associations were represented by one factor each fitted better than a one-factor solution (Greenwald & Farnham, 2000; Nosek, 2005; Nosek & Smyth, 2007). It may therefore be that these are two separate cognitive processes.

Our results are in line with the idea that responses to brightness may be more stable because they have developed earlier in our evolution both biologically and culturally. Our finding that the association between brightness and positivity is automatic is in line with previous findings (Hemphill, 1996; Lakens et al., 2013, 2012; Meier et al., 2004, 2015). Lakens et al. (2012) point out that what may be driving the association between brightness and positivity is the strong association between black and negativity rather than that of white and positivity. Due to the relative nature of the IAT (association between positive words and bright pictures necessarily also means an association between negative words and dark images), we cannot exclude this explanation. It bears mention though that in our study we used both achromatic as well as chromatic stimuli. For the achromatic stimuli the account of Lakens et al. (2012) holds up, however, for the chromatic stimuli it may be more difficult to see how this would work. The hue we used in Study 1, blue, is typically reported as the most preferred color (e.g. Hemphill, 1996). In contrast, blue is also typically associated with sadness (e.g. Soriano & Valenzuela, 2009) as also evidenced by common phrases such as “feeling blue”. Therefore blue can be said to be typically associated with positivity (because it's highly liked) as well as negativity (because it is associated with sadness). Since these associations are typically reported as associated with “blue” in general (rather than dark or light blue specifically) it is hard to see how they could lead our findings. We think it is more likely that people have a general “brightness bias” as later shown by Lakens and colleagues in their 2013 study. Even though people may see white as neutral and black as bad, the overall brightness of a stimulus gives it an implicit marker of positivity (high overall brightness) or negativity (low overall brightness).

We hope that these findings will inspire other researchers to look beyond hue and to deeper investigate the effects of brightness in particular in the domain of color associations. To facilitate future research efforts of our effects we have made the experimental scripts as well as the syntax used for analysis available on the open science framework: <https://osf.io/gsnf4/>. Taking everything together we feel that our results clearly show a strong and universal effect for brightness.

References

- Adams, F. M., & Osgood, C. E. (1973). A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology*, 4(2), 135–156.
- Albertazzi, L., Da Pos, O., Canal, L., Micciolo, R., Malfatti, M., & Vescovi, M. (2013). The hue of shapes. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 37–47. <http://dx.doi.org/10.1037/a0028816>.
- Alexander, K. R., & Shansky, M. S. (1976). Influence of hue, value, and chroma on the perceived heaviness of colors. *Perception & Psychophysics*, 19(1), 72–74. <http://dx.doi.org/10.3758/BF03199388>.
- Biggam, C. P. (2014). Prehistoric colour semantics. In W. Anderson, C. P. Biggam, C. Hough, & C. Kay (Eds.). *Colour studies: A broad spectrum* (pp. 18–43). John Benjamins Publishing Company.
- Boon, M. Y., Suttle, C. M., & Dain, S. J. (2007). Transient VEP and psychophysical chromatic contrast thresholds in children and adults. *Vision Research*, 47(16), 2124–2133. <http://dx.doi.org/10.1016/j.visres.2007.04.019>.
- Borg, A. (2014). Towards a historical and cultural atlas of colour terms in the near east. In W. Anderson, C. P. Biggam, C. Hough, & C. Kay (Eds.). *Colour studies: A broad spectrum* (pp. 46–67). John Benjamins Publishing Company.
- Bullough, E. (1907). On the apparent heaviness of colours: A contribution to the aesthetics of colour. *British Journal of Psychology*, 2(2), 111–152.
- Chen, N., Tanaka, K., & Watanabe, K. (2015). Color-shape associations revealed with Implicit Association Tests. *PLoS One*, 10(1), 1–15. <http://dx.doi.org/10.1371/journal.pone.0116954>.
- Crognale, M. A. (2002). Development, maturation, and aging of chromatic visual pathways: VEP results. *Journal of Vision*, 2(6), 438–450. <http://dx.doi.org/10.1167/2.6.2>.
- Dael, N., Perseguers, M.-N., Marchand, C., Antonietti, J.-P., & Mohr, C. (2016). Put on

- that colour, it fits your emotion: Colour appropriateness as a function of expressed emotion. *The Quarterly Journal of Experimental Psychology*, 69(8), 1619–1630. <http://dx.doi.org/10.1080/17470218.2015.1090462>.
- De Camp, J. E. (1917). The influence of color on apparent weight. A preliminary study. *Journal of Experimental Psychology*, 2(5), 347. <http://dx.doi.org/10.1037/h0075903>.
- Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, 23(3), 316–326. <http://dx.doi.org/10.1177/0146167297233009>.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54(1), 297–327. <http://dx.doi.org/10.1146/annurev.psych.54.101601.145225>.
- Fechner, G. T. (1876). *Vorschule der ästhetik*. Breitkopf & Härtel.
- Gao, X. P., Xin, J. H., Sato, T., Hansuebsai, A., Scalzo, M., Kajiwar, K., ... Billger, M. (2007). Analysis of cross-cultural color emotion. *Color Research and Application*, 32(3), 223–229. <http://dx.doi.org/10.1002/col.20321>.
- Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.). *Handbook of research methods in social and personality psychology* (pp. 283–310). (2nd ed.). New York, NY: Cambridge University Press.
- Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology*, 44(5), 1355–1361. <http://dx.doi.org/10.1016/j.jesp.2008.04.005>.
- Gawronski, B., Deutsch, R., & Banse, R. (2011). Response interference tasks as indirect measures of automatic associations. *Cognitive methods in social psychology* (pp. 78–123).
- Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (2003). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology*, 39(1), 16–30. [http://dx.doi.org/10.1016/S0022-1031\(02\)00517-6](http://dx.doi.org/10.1016/S0022-1031(02)00517-6).
- Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience*, 26, 181–206. <http://dx.doi.org/10.1146/annurev.neuro.26.041002.131116>.
- Gelineau, E. P. (1981). A psychometric approach to the measurement of color preference. *Perceptual and Motor Skills*, 53(1), 163–174.
- Gombrich, E. H. (1963). Expression and communication. *Meditations on a hobby horse: and other essays on the theory of art* (pp. 56–69). Phaidon Press.
- Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, 79(6), 1022–1038. <http://dx.doi.org/10.1037/0022-3514.79.6.1022>.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. <http://dx.doi.org/10.1037/0022-3514.74.6.1464>.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197–216. <http://dx.doi.org/10.1037/0022-3514.85.2.197>.
- Hemphill, M. (1996). A note on adults' color-emotion associations. *The Journal of Genetic Psychology*, 157(3), 275–280. <http://dx.doi.org/10.1080/00221325.1996.9914865>.
- Hurlbert, A. C., & Ling, Y. (2007). Biological components of sex differences in color preference. *Current Biology*, 17(16), R623–R625. <http://dx.doi.org/10.1016/j.cub.2007.06.022>.
- Itten, J. (1961). *Kunst der Farbe. Subjektives Erleben und objektives Erkennen als Weg zur Kunst*. Ravensburg: Otto Maier.
- Kandinsky, W. (1911). *Über das Gesichtige in der Kunst. Insbesondere in der Malerei*. München: Piper und Co.
- Kandinsky, W. (1946). *On the spiritual in art (first comp)*. New York: Solomon R. Guggenheim Foundation.
- Kay, P., & Cook, R. S. (2015). World color survey. *Encyclopedia of color science and technology* (pp. 1–8). Berlin, Heidelberg: Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-27851-8_113-10.
- Kazanas, S. A., & Altarriba, J. (2015). The automatic activation of emotion and emotion-laden words: Evidence from a masked and unmasked priming paradigm. *The American Journal of Psychology*, 128(3), 323. <http://dx.doi.org/10.5406/amerjpsyc.128.3.0323>.
- Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J., & Midden, C. J. H. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion*, 27(7), 1225–1246. <http://dx.doi.org/10.1080/02699931.2013.781501>.
- Lakens, D., Semin, G. R., & Foroni, F. (2012). But for the bad, there would not be good: Grounding valence in brightness through shared relational structures. *Journal of Experimental Psychology: General*, 141(3), 584–594. <http://dx.doi.org/10.1037/a0026468>.
- Lazreg, C. K., & Mullet, E. (2001). Judging the pleasantness of form-color combinations. *The American Journal of Psychology*, 114(4), 511–533.
- McLaren, K. (1976). XIII-The development of the CIE 1976 (L* a* b*) uniform colour space and colour-difference formula. *Journal of the Society of Dyers and Colourists*, 92(9), 338–341. <http://dx.doi.org/10.1111/j.1478-4408.1976.tb03301.x>.
- Meier, B. P., Fetterman, A. K., & Robinson, M. D. (2015). Black and white as valence cues. *Social Psychology*, 46(3), 174–178. <http://dx.doi.org/10.1027/1864-9335/a000236>.
- Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white: Automatic inferences about stimulus valence based on brightness. *Psychological Science*, 15(2), 82–87. <http://dx.doi.org/10.1111/j.0963-7214.2004.01502002.x>.
- Mogensen, M. F., & English, H. B. (1926). The apparent warmth of colors. *The American Journal of Psychology*, 37(3), 330–344. <http://dx.doi.org/10.1163/156853109X436801>.
- Monroe, M. (1925). The apparent weight of color and correlated phenomena. *The American Journal of Psychology*, 36(2), 192–206.
- Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, 134(4), 565–584.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31(2), 166–180. <http://dx.doi.org/10.1177/0146167204271418>.
- Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the implicit association test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, 54(1), 14–29. <http://dx.doi.org/10.1027/1618-3169.54.1.14>.
- Osborne, L. A., Romano, M., Re, F., Roaro, A., Truzoli, R., & Reed, P. (2016). Evidence for an internet addiction disorder: Internet exposure reinforces color preference in withdrawn problem users. *The Journal of Clinical Psychiatry*, 77(2), 269–274. <http://dx.doi.org/10.4088/JCP.15m10073>.
- Page, J. W., & Crognale, M. A. (2005). Differential aging of chromatic and achromatic visual pathways: Behavior and electrophysiology. *Vision Research*, 45(11), 1481–1489. <http://dx.doi.org/10.1016/j.visres.2004.09.041>.
- Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences of the United States of America*, 107(19), 8877–8882. <http://dx.doi.org/10.1073/pnas.0906172107>.
- Pavlović, M., & Marković, S. (2012). Automatic processes in aesthetic judgment: Insights from the Implicit Association Test. 45(4), 377–393. <http://dx.doi.org/10.2298/PSI1204377P>.
- Pinkerton, E., & Humphrey, N. K. (1974). The apparent heaviness of colours. *Nature*, 2(2), 164–165. <http://dx.doi.org/10.1111/j.2044-8295.1907.tb00236.x>.
- Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, 13(10), 439–446. <http://dx.doi.org/10.1016/j.tics.2009.07.001>.
- Saito, M. (1996). A comparative study of color preferences in Japan, China and Indonesia, with emphasis on the preference for white. *Perceptual and Motor Skills*, 83(1), 115–128. <http://dx.doi.org/10.2466/pms.1996.83.1.115>.
- Schwitters, K. (1918). *Mainfeste und kritische Prosa*. In F. Lach (Ed.). *Nachträge zu den Bänden*. DuMont: Köln.
- Soriano, C., & Valenzuela, J. (2009). Emotion and colour across languages: Implicit associations in Spanish colour terms. *Social Science Information*, 48(3), 421–445. <http://dx.doi.org/10.1177/0539018409106199>.
- Sutton, T. M., & Altarriba, J. (2016). Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection. *Behavior Research Methods*, 48(2), 686–728. <http://dx.doi.org/10.3758/s13428-015-0598-8>.
- Tinker, M. A. (1938). Effect of stimulus-texture upon apparent warmth and affective value of colors. *The American Journal of Psychology*, 51(3), 532–535.
- Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, 123(4), 394–409. <http://dx.doi.org/10.1037/0096-3445.123.4.394>.
- Vö, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534–538. <http://dx.doi.org/10.3758/BRM.41.2.534>.
- Wei, M., Houser, K. W., Allen, G. R., & Beers, W. W. (2014). Color preference under LEDs with diminished yellow emission. *LEUKOS*, 10(3), 119–131. <http://dx.doi.org/10.1080/15502724.2013.865212>.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19), 7780–7785. <http://dx.doi.org/10.1073/pnas.0701644104>.

2.1.3

Specker, E., & Leder, H. (2018). Looking on the Bright Side: Replicating the Association between Brightness and Positivity. *Collabra: Psychology*, 4(1), 34.
<http://doi.org/10.1525/collabra.168>

My Contribution:

Again, being first author, my contributions to this paper are manifold. I came up with the core concept of the paper. This allowed us not just to replicate our results but also for me to gain experience in the pre-registration process which eventually led to us being one of the prize winners of the Pre-registration Challenge of the Open Science Framework.

I conducted the analyses and literature search the paper is based on. I was the main writer of the article but received feedback from HL on different versions of the manuscript.

ORIGINAL RESEARCH REPORT

Looking on the Bright Side: Replicating the Association between Brightness and Positivity

Eva Specker^{*,†} and Helmut Leder^{*,‡}

The present study is a pre-registered replication of a study by Specker et al. (2018) that tests the hypothesis that brightness of colors is associated with positivity. Our results showed an implicit association between brightness and positivity in both Study 1 and Study 2, however, an explicit association between brightness and positivity was only found in Study 2, thereby replicating 3 out of 4 effects. To investigate these effects in more detail, we present a meta-analysis of both the original and the replication study. This indicated a large effect 1.31 [1.12, 1.51]. In addition, we used meta-analysis to assess potential moderators of the effect, in particular stimulus type (chromatic vs. achromatic) and measure type (implicit vs. explicit). This indicated that the effect is stronger when measured implicitly than when measured explicitly and that the effect is stronger when achromatic stimuli are used. In sum, we take these findings to indicate that there is a strong and replicable association between brightness and positivity. These findings offer researchers interested in the effect concrete tools when designing a study investigating the effect with regard to effect size estimates for power analysis as well as stimulus and measurement design.

Keywords: color; color association; brightness; positivity; replication

The idea that the brightness of colors is associated with positivity is not new in terms of everyday experience nor in terms of scientific thinking. Numerous studies have indicated a brightness-positivity association that applies cross-culturally (Adams & Osgood, 1973; Saito, 1996; Valdez & Mehrabian, 1994). This notwithstanding, only a handful of studies (Hemphill, 1996; Lakens, Fockenberg, Lemmens, Ham, & Midden, 2013; Lakens, Semin, & Foroni, 2012; Meier, Robinson, & Clore, 2004; Meier, Fetterman, & Robinson, 2015) tested this association directly.

Different theoretical accounts have been proposed to explain the brightness-positivity association. One account, supported by Meier and colleagues (2004, 2015) and Huang, Tse, and Xie (2017) bases itself on conceptual metaphor theory (Lakoff & Johnson, 1980). According to conceptual metaphor theory people can represent abstract concepts (such as valence) in terms of concrete concepts (brightness) through metaphoric associations. Thus, since brightness is metaphorically connected to valence a brightness-positivity association arises. This

account suggests that the association between brightness and positivity is a learned association. In contrast, based on the indications of high cross-cultural similarity (e.g., Adams & Osgood, 1973; Saito, 1996) we (Specker et al., 2018) recently postulated that the association between brightness and positivity is an automatic and universal association that most likely arises early in development. We used the Implicit Association Test (hereafter: IAT, Greenwald, McGhee, & Schwarz, 1998) as well as an explicit rating task to assess the association between brightness and positivity. Our results were consistent with these hypotheses, indicating that this association is both automatic as well as universal.

These notions are also in line with accounts such as Lakens et al. (2013) who proposed that people have a “brightness bias” where the overall brightness of a stimulus gives it an implicit marker of positivity. More recently, building on this idea of brightness as an implicit marker of positivity, Schietecat, Lakens, IJsselsteijn, and De Kort (2018) proposed a dimension-specificity hypothesis. This hypothesis builds on the work of Osgood, Suci, and Tannenbaum (1957) whose research focused on the meaning that people assign to concepts, in our case assigning the meaning of “positivity” to bright stimuli. In their seminal work they distinguished three broad categories of meaning, namely evaluation, activity, and potency. Within this context, the dimension-specificity hypothesis proposes that the attribution of meaning (e.g., “positivity”) depends on the saliency of the context. They

* Faculty of Psychology, Department of Basic Psychological Research and Research Methods, University of Vienna, Liebiggasse 5, 1010 Vienna, AT

† Faculty of Historical and Cultural Studies, Department of Art History, University of Vienna, Vienna, AT

‡ Research Platform Cognitive Science, University of Vienna, Vienna, AT

Corresponding author: Eva Specker (eva.specker@univie.ac.at)

propose that brightness will be associated with calmness when the evaluation dimension is salient (essentially showing a brightness-positive association). In contrast, this hypothesis predicts that brightness will be associated with aggression when the activity dimension is salient (essentially showing a brightness-negative association).

Our previous findings (Specker et al., 2018) would be in line with the dimension-specificity account. In the studies by Schietecat et al. (2018) the assumption is made that brightness is automatically associated with positivity and that thus by manipulating brightness of the stimulus one can make the evaluation dimension salient. Both our and their results seem to support this idea and indicate that brightness is in fact automatically associated with positivity.

Nonetheless, previous studies supporting this idea have been limited. As mentioned, only a small set of studies directly investigated this effect (Hemphill, 1996; Lakens, Fockenberg, Lemmens, Ham, & Midden, 2013; Lakens, Semin, & Foroni, 2012; Meier, Robinson, & Clore, 2004; Meier, Fetterman, & Robinson, 2015), crucially, the majority of these studies investigated achromatic brightness differences (white vs. black). The information these studies provide about the association between brightness and positivity when looking at luminous brightness differences (bright vs. dark) is therefore limited. Especially because in regard to the two studies who did not use achromatic stimuli, one used color patches of different hues that were not controlled for brightness or saturation (Hemphill, 1996) and the other (Lakens et al., 2013) used complex naturalistic pictures.

Both our study (Specker et al., 2018) and the study by Schietecat et al. (2018) addressed this issue. A benefit of our design was the more pure focus on finding the boundary conditions of the effect as well as having better controlled stimuli in comparison to Hemphill (1996) and Schietecat et al. (2018). In the case of Hemphill (1996) participants saw colored cardboard rectangles of 10 different hues. Unfortunately, information on how these cardboard rectangles were produced or under which conditions they were viewed was not reported. In the study by Schietecat et al. (2018) viewing conditions were not held constant. Since participants completed the experiment at home, the monitor settings were not controlled and color presentation may have varied widely as a result. Additionally, in both cases, participants may have completed the task at different times of the day leading daylight conditions to vary which potentially influences the results. In fact, Schietecat et al. (2018) explicitly manipulate this in Experiment 2A and 2B and found that if the surrounding light is brighter this also acts as a cue for positivity, though only with achromatic stimuli.

In sum, our original study (Specker et al., 2018) extended the previous research by improving methodological design and showing consistent effects. In addition, it has the potential to be useful for further research efforts (e.g., investigating the dimension-specificity hypothesis) that assume that an automatic association between brightness and positivity exists. To add more substance to the claims made in our original study we decided to replicate our results. We preregistered the replication on

the Open Science Framework: <https://osf.io/uafyb/>. In the current study we present a direct replication of Study 1 and Study 2 of Specker et al. (2018). Both studies use the IAT; in Study 1 color patches of the same hue are used, in Study 2 black-and-white stimuli are used. In our original study (Specker et al., 2018) we improved the control over the chromatic color stimuli in comparison to previous studies, nonetheless, our control was not perfect and our manipulation of brightness also led to small variations in hue. Therefore, we conducted Study 2 with achromatic stimuli to exclude potential confounds of hue and/or saturation. In the original study only Study 2 included an explicit measure, for our replication we also included this measure in Study 1. If an association between brightness and positivity exists this should apply to both stimulus types as well as both measurement methods. We will first discuss the replication of Study 1 (also Study 1 in our replication) and then the replication of Study 2 (also Study 2 in our replication). Finally, we will present results of an explorative meta-analysis of both the original study and the replication to allow for a comprehensive overview of the effect and investigate potential moderators. Though the term explorative is not usually applicable to meta-analysis we use it here to emphasize that this part of our analysis was not included in our preregistration.

Study 1

Participants

We based our power analysis on the effects reported by Specker et al. (2018), the smallest effect size we reported in that study was a Cohen's d of .90, this was the effect size of the explicit association between brightness and positivity when using black and white stimuli in a Western cultural context. To err on the side of caution we used a slightly smaller estimated effect size of .80 for our power analysis using the G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007). This indicated that we needed 15 participants to achieve a power of .80 with an error rate probability of 0.05. We therefore collected data from 15 participants. All participants were psychology students of the University of Vienna and received course credit for their participation. Eight participants were female and the mean age was 22.93 ($SD = 3.96$). The experiment was carried out in accordance with the Declaration of Helsinki and in a procedure which was approved by the local ethical committee of the University of Vienna. We pre-registered an exclusion criterion of a minimum accuracy of 60%, meaning that anyone below that threshold would be excluded. Since none of our participants met our exclusion criteria we included all participants in our analysis.

Method

Our method, materials and procedure was identical to Study 1 of Specker et al. (2018), thus we used an IAT the only difference being that we included an explicit rating task. The IAT and the explicit rating task were counterbalanced. Meaning that half of the participants first completed the IAT and then gave explicit ratings and the other half first gave explicit ratings and then completed the IAT.

Materials

IAT task. We used the standard IAT design, that includes error feedback for incorrect responses. In this case, the error message remains on the screen until the correct key is pressed. Key assignments were visible throughout the task as recommended by Gawronski, Deutsch, and Banse (2011). We opted to counterbalance the key assignments in order to be able to eliminate block order as potential confound. Meaning that half of the participants first had a congruent blocks (i.e. bright/positive and dark/negative) and then incongruent blocks (i.e. bright/negative and dark/positive) and the other half first had incongruent blocks followed by a congruent blocks.

Explicit Rating Task. In the explicit rating task people were asked to rate each stimulus on positivity. People were asked: “How positive is this color patch?” and answered on a 7 point scale ranging from 1 (negative) to 7 (positive).

Stimuli. We created 18 color patches differing on brightness as stimuli for the IAT task. All color patches were blue. Nine color patches represented the bright category and nine the dark category. The bright category ranged from 10–30% brightness with increasing steps of about 2–3%. The dark category ranged from 70–90% brightness with increasing steps of about 2–3%. All images can be found here: <https://osf.io/w8479/>. As a manipulation check all stimuli were measured by use of a colorimeter (i1Xtreme by X-Rite) this was done by presenting the stimuli full screen on a CRT monitor and then taking the measurements. Measurement was carried out on the monitor that was used by the participant during the test. For this manipulation check we focused on the luminance measures of the LAB color model. This showed that the dark category was lower in luminance (and therefore darker) than the bright category, $t(16) = -15.49$, $p < .001$. All stimuli were presented against a grey background (RGB: 128, 128, 128; LAB: 64.3, -4.4, -23.1). To ensure consistency of color presentation across subjects the IAT task was conducted in a windowless room with constant lighting conditions and on a CRT Monitor.

To represent the positive and negative category, we selected words from the Berlin Affective Word List Reloaded (BAWL-R, Vö et al., 2009). The BAWL-R is a cross-validated resource of affective words in German that have been rated for valence and arousal. Negative words included words such as asocial, sad, and loveless (respectively, “asozial”, “traurig”, and “lieblos” in German) and the positive words included words such as loved, great, and courageous (respectively, “beliebt”, “toll”, “mutig” in German). The words selected to be in the positive and negative category differed in valence, $t(16) = 63.41$, $p < .001$, but not in arousal, $p = .77$. We matched for word length, on average negative words had a length of 7.22 characters and positive words of 6.44 characters, this difference in length was not significant, $p = .244$.

Experimental scripts and related stimuli material is freely available on the Open Science Framework: <https://osf.io/w8479/>.¹

Results

As in Specker et al. (2018) a D-score was calculated as recommended by Greenwald, Nosek, and Banaji (2003) for every person. The D-score ranges from -1.5 to +1.5. Typically, D-scores of $> .65$ are seen as large effects. Since the IAT is a relative measure a positive D-score indicates a larger association between brightness and positivity than brightness and negativity. As a consequence this also means a larger association between darkness and negativity than darkness and positivity. This indicated a large effect with the mean D-score in our sample being $.65$ ($SD = .37$). As in Specker et al. (2018) we then conducted a one-sided one sample t-test. This showed a significant difference from 0 in the expected (positive) direction, $t(14) = 6.58$, $p < .001$, Cohen's $d_z = 1.73$,² CI [.91, 2.51]. For the explicit association, as in Specker et al. (2018), a difference score was created by subtracting the standardized mean rating of the dark category from the standardized mean rating of the bright category. We then conducted a one-sided one sample t-test over the difference score. This did not show a significant difference from 0, $t(14) = -.90$, $p > .05$, Cohen's $d_z = 0.23$ CI [-.74, .28]. Finally, as in Specker et al. (2018) we computed the correlation between implicit and explicit association. The correlation between implicit and explicit association was $r = -.39$, $p > .05$, as illustrated in **Figure 1**.

Discussion

Study 1 replicated the findings of Study 1 of Specker et al. (2018) meaning that both studies found a large effect for the implicit association of brightness and positivity. Study 1 did not show an explicit effect for the association between brightness and positivity that should theoretically have been there, we therefore interpret this as a failure to replicate. The most probable reason for this failure to replicate – besides a simple Type 2 error – is the fact that when varying brightness we did not perfectly control for variation in hue or saturation. Therefore, it is possible that effects found are due to either hue and/or saturation rather than brightness. This was also the main reason of the original study to perform a second study with achromatic stimuli. If similar effects are found with achromatic stimuli, one can be more secure in arguing that this was due to brightness. For an absence of the effect, it might be that because these stimuli were less controlled, the effect disappeared. This becomes more likely when combined with the findings of the original study that the explicit rating effects are smaller than the implicit rating effects. Though it has to be mentioned that the smaller effect size is likely due, at least in part, to the fact that reaction time based measures with a lot of trials have less variance and therefore larger effect sizes than a one-trial explicit rating. Another likely explanation for the absence of an explicit effect, is that in the explicit rating task people confounded “positivity” with “liking”. In other words, if someone saw an image they liked that was dark they may have perceived it as more positive, similarly, when they saw a bright image they disliked they may have perceived it as more negative. In any case, this seems to suggest that stimulus type and measurement type influence the size of the effect. In this case, the odds

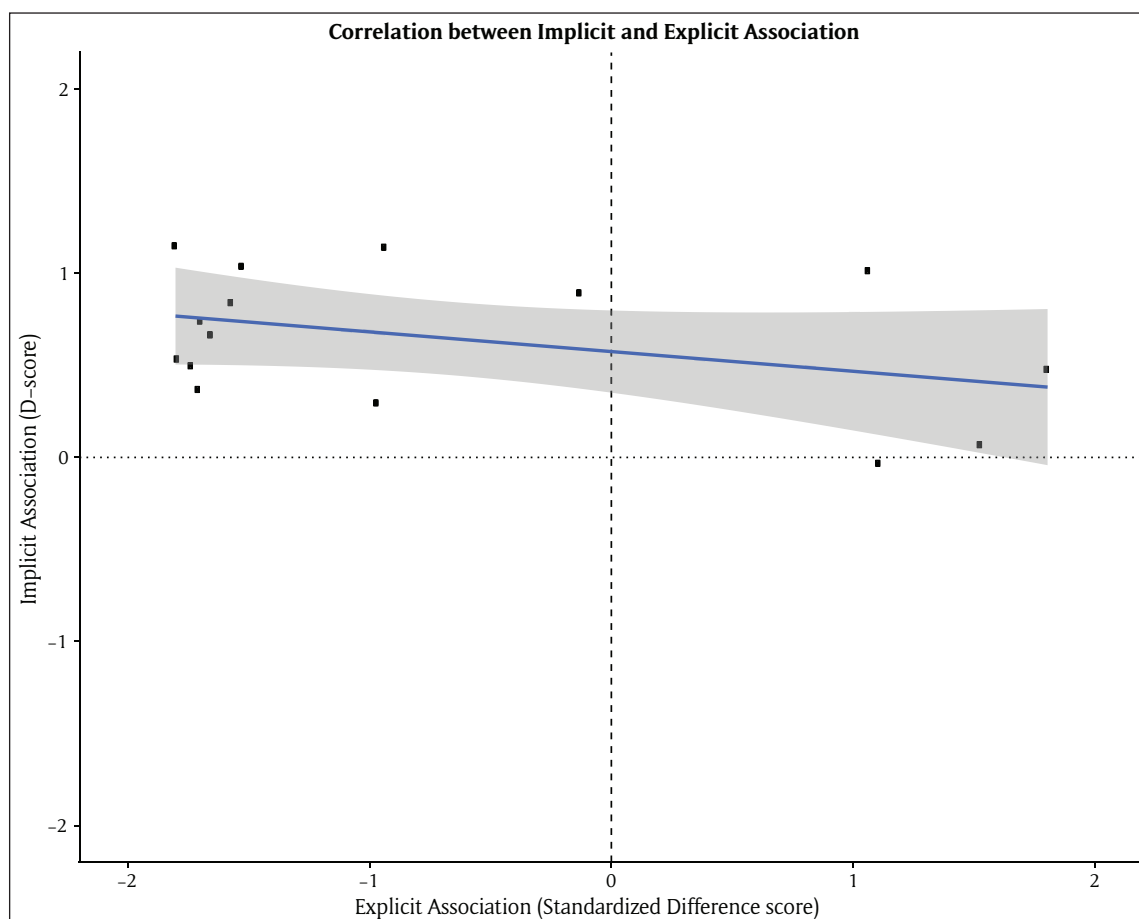


Figure 1: Scatterplot of the correlation between Explicit and Implicit (D score of the IAT) association. To aid interpretation lines are drawn on the 0 point of both axis. Because both measures represent a difference score, a positive score, thus above the 0 point on the horizontal axis and right from the 0 point on the vertical axis, represents an association between brightness and positivity.

were stacked against finding the effect since it combined lesser controlled stimuli as well as explicit rating as a measurement. This is a notion we will investigate in more detail later by use of meta-analysis.

Study 2

Participants

We based our power analysis on the effects reported by Specker et al. (2018), the smallest effect size we reported in that study was a Cohen's d_z of .90, this was the effect size of the explicit association between brightness and positivity when using black and white stimuli in a Western cultural context. To err on the side of caution we used a slightly smaller estimated effect size of .80 for our power analysis using the G*Power software (Faul et al., 2007). This indicated that we needed 15 participants to achieve a power of .80 with an error rate probability of 0.05. We therefore collected data from 15 participants. All participants were psychology students of the University of Vienna and received course credit for their participation. Eleven participants were female and the mean age was 19.67 ($SD = 1.39$). The experiment was carried out in accordance with the Declaration of Helsinki and in a procedure which was approved by the local ethical committee of the University of Vienna. We pre-registered an exclusion criterion of a minimum accuracy of 60%,

meaning that anyone below that threshold would be excluded. Since none of our participants met our exclusion criteria we included all participants in our analysis.

Method

Our method, materials and procedure was identical to Study 2 of Specker et al. (2018) and Study 1 of the current paper with the exception of the color stimuli. Rather than having variations of blue we constructed achromatic stimuli. These employed a ratio from black-to-white. The bright category consisted of 8 color patches which were filled 10 or 20% black (with 90 or 80% white respectively). The dark category consisted of 8 color patches which were filled 10 or 20% white (with 90 or 80% black respectively). All images can be found here: <https://osf.io/w8479/>. Again, the IAT task and the explicit rating task were conducted in a windowless room with constant lighting conditions on a CRT monitor.

Results

We used the same analyses as in Specker et al. (2018) and Study 1. Thus, we calculated a D-score for every person as recommended by Greenwald, Nosek, and Banaji (2003). This indicated a large effect with the mean D-score in our sample being .95 ($SD = .23$). We then conducted a one-sided one sample t-test. This showed a significant difference

from 0 in the expected (positive) direction, $t(14) = 15.77$, $p < .001$, Cohen's $d_z = 4.07$, CI [2.49, 5.64]. For the explicit association, we created a difference score by subtracting the standardized mean rating of the dark category from the standardized mean rating of the bright category. We then conducted a one-sided one sample t-test over the difference score. This showed a significant difference from 0 in the expected (positive) direction, $t(14) = 11.88$, $p < .001$, Cohen's $d_z = 3.06$, CI [1.83, 4.29]. Finally, we computed the correlation between implicit and explicit association. The correlation between implicit and explicit association was $r = .63$, $p < .01$, as illustrated in **Figure 2**. As can be seen from **Figure 2** this correlation was mainly driven by one data point (the one on the far left), to test this hypothesis exploratively we calculated the correlation without this outlier. Without the outlier the correlation was $r = .067$.

Discussion

Study 2 replicated the findings of Study 2 of Specker et al. (2018) meaning that both studies found a large effect for the implicit and explicit association of brightness and positivity. Taking these findings together with the findings of Study 1 we can conclude that we replicated the effect in 3 out of 4 cases. There are two possible explanations as to why we were able to replicate the explicit rating effects in Study 2, one would be that it is due to the higher level of

control over our stimuli, the other option would be that positivity (i.e. the evaluation dimension) is more salient in achromatic stimuli. Since black and white are stronger bipolar opposites than luminous differences in brightness this increases saliency and therefore leads to larger effects. In any case, these results like the results of Study 1 seem to suggest a moderating effect of both stimulus type as well as measurement type on the size of the effect. We will now investigate these notions in more detail as well as to give a more comprehensive estimate of the effect by use of an explorative³ meta-analysis.

Meta-Analysis

We ran a fixed-effects model over the results of both the original study Specker et al. (2018) and our replication. We chose a fixed-effects model rather than a random effects model because the fixed-effects model assumes homogeneity in estimated effect size. Though this assumption often does not hold up when conducting a meta-analysis over a large amount of different studies, the high similarity across our studies made this the best suited model. The results are illustrated in **Figure 3**. The fixed effects model indicated a meta-analytic effect of 1.31 [1.12, 1.51]. However, there was residual heterogeneity not explained by the model, $Q(8) = 86.88$, $p < .001$. This was coherent with previous descriptive comparisons

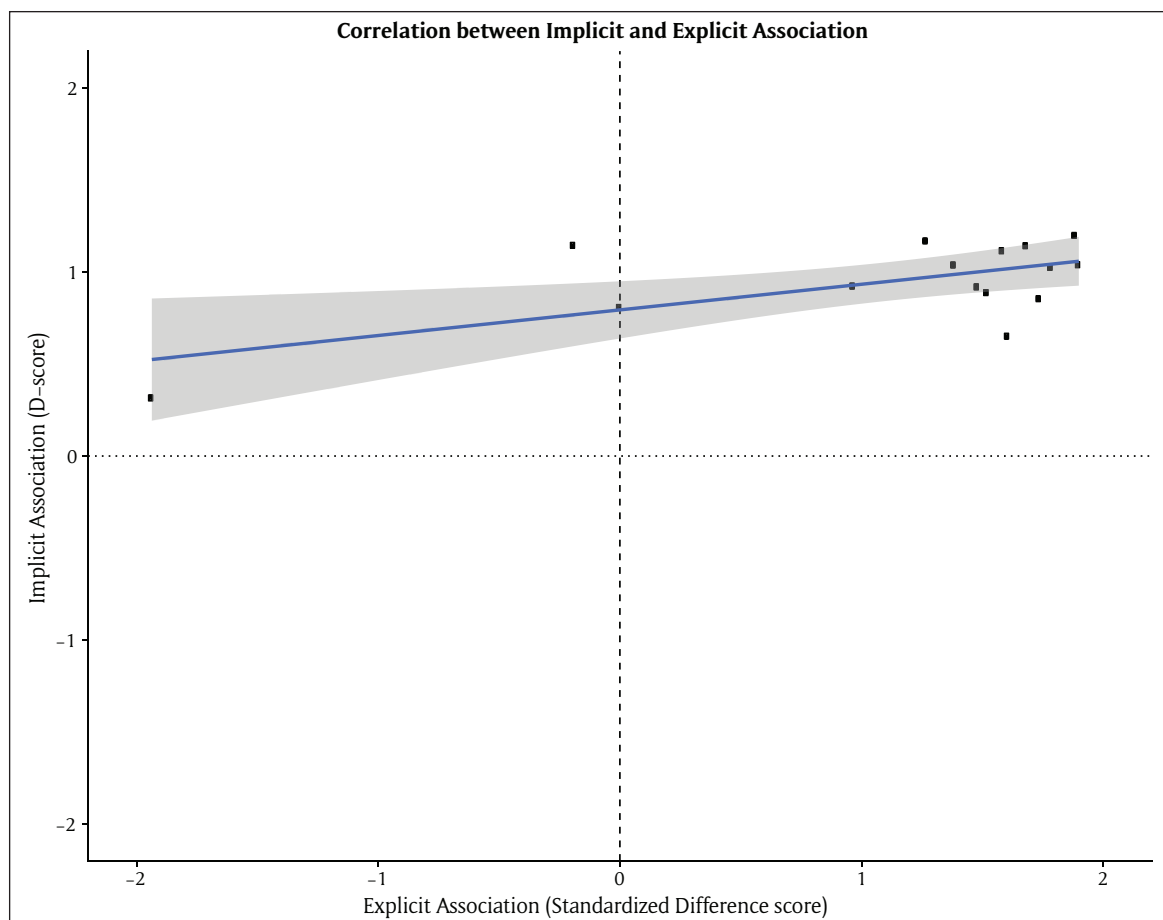


Figure 2: Scatterplot of the correlation between Explicit and Implicit (D score of the IAT) association. To aid interpretation lines are drawn on the 0 point of both axis. Because both measures represent a difference score, a positive score, thus above the 0 point on the horizontal axis and right from the 0 point on the vertical axis, represents an association between brightness and positivity.

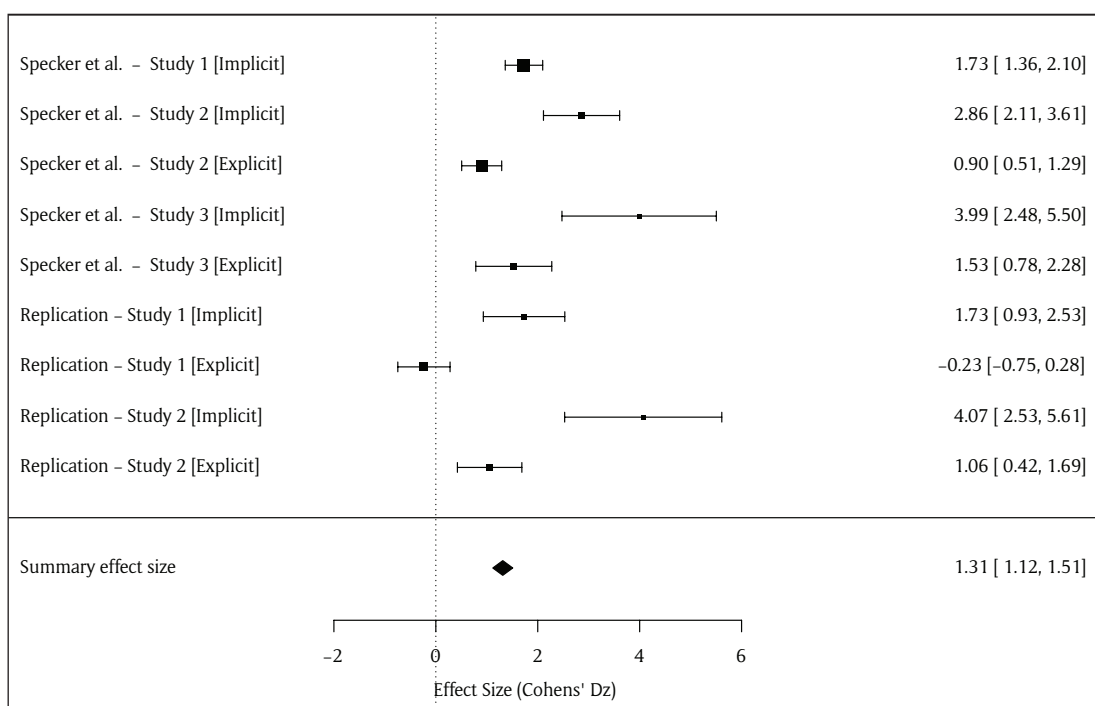


Figure 3: Forest plot of the Fixed-Effects Model. All effect sizes within the plot are Cohen's Dz values. All confidence intervals represent 95% confidence intervals. The size of the boxes represents the sample size (N) of the study.

of our included studies which indicated that the small variations of our study design seemed to influence the effect. We therefore ran a fixed-effects model with moderators to see if this was able to explain the residual heterogeneity. As discussed above, our studies employed two different stimulus types (chromatic vs. achromatic) where the achromatic stimuli represent better controlled and more salient stimuli. In addition, when comparing the results descriptively both this replication and the original study reported larger effects when measuring the effect implicitly in comparison to an explicit measurement. Though it has to be mentioned that the smaller effect size is likely due, at least in part, to the fact that reaction time based measures with a lot of trials have less variance and therefore larger effect sizes than a one-trial explicit measure. We therefore used stimulus type (chromatic vs. achromatic) and measure type (implicit vs. explicit) as moderators. The fixed-effects model with moderators showed a significant effect of the moderators, $Q(2) = 81.43$, $p < .001$. In addition, there was no residual heterogeneity, $Q(6) = 5.44$, $p > .05$, this indicates that all heterogeneity in the effects can be explained by the moderators. The regression coefficient of stimulus type is -1.37 , $p < .001$. This indicates that when stimulus type is chromatic the estimated effect (strength of the association between brightness and positivity) decreases with 1.37. The regression coefficient of measure type is 2.06 , $p < .001$. This indicates that when measure type is implicit the estimated effect increases with 2.06.

General Discussion

We were able to replicate the effects of Specker et al. (2018) in three out of four cases. In addition, the meta-analysis indicated a large effect. In sum, this indicates that there is a strong and replicable association between

brightness and positivity. However, the size of the effect is dependent on measurement type as well as stimulus type, as indicated by the meta-analysis. The effect is stronger when measured implicitly than when measured explicitly. As noted, this is likely due to the fact that reaction time based measures with a lot of trials have less variance and therefore larger effect sizes than a one-trial explicit measure. Another likely explanation would be that people confounded "positivity" with "liking". In other words, if someone saw an image they liked that was dark they may have perceived it as more positive, similarly, when they saw a bright image they disliked they may have perceived it as more negative. In addition, the effect is stronger when using better controlled and more salient stimuli. These findings offer researchers interested in the effect concrete tools when designing a study investigating the effect with regard to effect size estimates for power analysis as well as stimulus and measurement design. Especially researchers who want to use brightness as a cue (e.g., to make the evaluation dimension salient) can use these findings in their research design. To facilitate future research efforts we have made the experimental scripts, analysis scripts, all related stimuli material and our data freely available on the Open Science Framework: <https://osf.io/w8479/>.⁴

Lakens et al. (2012) have suggested that the association between brightness and positivity may be explained by the strong association people have between black and negativity rather than white and positivity. Due to the relative nature of the IAT one cannot exclude this explanation based on our data. However, the inclusion of chromatic stimuli poses a potential argument against this reasoning since it is based on associations with achromatic stimuli. As noted in Specker et al. (2018), the hue we used (blue), is typically regarded as the

most preferred color (e.g., Hemphill, 1996) while being at the same time being stereotypically associated with sadness (Soriano & Valenzuela, 2009). Thus, one can assume blue would be seen as positive (due to being liked) as well as negative (due to its association with sadness). We think it is more likely that people have a general brightness bias as later postulated by Lakens et al. (2013). This more symmetrical relationship is in line with the metaphorical account (Huang et al., 2017; Meier et al., 2004; Meier et al., 2015) where people can represent abstract concepts (such as valence) in terms of concrete concepts (brightness) through metaphoric associations. This notwithstanding, metaphors have to be learned. This is incongruent with the high cross-cultural similarity we found in Specker et al. (2018). In addition, the fact that the effect is larger when using implicit than explicit measures as indicated by the meta-analysis suggest an automatic association that does not rely on explicit metaphors. Admittedly, through learning processes people would be able to internalize explicit metaphorical associations which would lead to implicit associations. Nonetheless, if this were the case one would assume that the original explicit (metaphorical) association would be at least as strong if not stronger than the resulting implicit association. This is not the case, making the metaphorical account less plausible. Thus, our results are in line with accounts such as recently postulated that the association between brightness and positivity is an automatic and universal association that most likely arises early in development.

In sum, the study shows clear support to conclude that there is a strong association between brightness and positivity as well as it providing new insight as to under which conditions the effect is present.

Data Accessibility Statement

As noted in the main body of the text, our experimental scripts, analysis scripts, all related stimuli material and our data are freely available on the Open Science Framework: <https://osf.io/w8479/>. The original pre-registration can be found at: <https://osf.io/uafyb/>.

Notes

- ¹ The interested reader can compare this material with the pre-registration: <https://osf.io/uafyb/>.
- ² In this case and all following instances Cohen's *d* and its confidence interval is calculated by using the method outlined here: <http://daniellakens.blogspot.co.at/2014/06/calculating-confidence-intervals-for.html>.
- ³ As noted, the term "explorative" is not usually applicable to meta-analysis, we use it here to emphasize that this part of our analysis was not included in our preregistration.
- ⁴ The interested reader can compare this material with the pre-registration: <https://osf.io/uafyb/>.

Funding Information

The writing of this paper was supported by a grant to HL by Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF: Projectnumber: CS15-036).

Competing Interests

The authors declare to have no competing interests.

Author Contributions

Contributed to conception and design: ES
 Contributed to acquisition of data: ES
 Contributed to analysis and interpretation of data: ES, HL
 Drafted and/or revised the article: ES, HL
 Approved the submitted version for publication: ES, HL

References

- Adams, F. M., & Osgood, C. E. (1973). A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology*, 4(2), 135–156. DOI: <https://doi.org/10.1177/002202217300400201>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. DOI: <https://doi.org/10.3758/BF03193146>
- Gawronski, B., Deutsch, R., & Banse, R. (2011). Response interference tasks as indirect measures of automatic associations. In: *Cognitive Methods in Social Psychology*, 78–123.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. DOI: <https://doi.org/10.1037/0022-3514.74.6.1464>
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197–216. DOI: <https://doi.org/10.1037/0022-3514.85.2.197>
- Hemphill, M. (1996). A Note on Adults' Color–Emotion Associations. *The Journal of Genetic Psychology*, 157(3), 275–280. DOI: <https://doi.org/10.1080/00221325.1996.9914865>
- Huang, Y., Tse, C.-S., & Xie, J. (2017). The bidirectional congruency effect of brightness-valence metaphoric association in the Stroop-like and priming paradigms. *Acta Psychologica*. DOI: <https://doi.org/10.1016/j.actpsy.2017.10.006>
- Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J., & Midden, C. J. H. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion*, 27(7), 1225–1246. DOI: <https://doi.org/10.1080/02699931.2013.781501>
- Lakens, D., Semin, G. R., & Foroni, F. (2012). But for the bad, there would not be good: Grounding valence in brightness through shared relational structures. *Journal of Experimental Psychology: General*, 141(3), 584–594. DOI: <https://doi.org/10.1037/a0026468>
- Lakoff, G., & Johnson, M. (1980). Conceptual Metaphor in Everyday Language. *The Journal of Philosophy*, 77(8), 453. DOI: <https://doi.org/10.2307/2025464>
- Meier, B. P., Fetterman, A. K., & Robinson, M. D. (2015). Black and White as Valence Cues. *Social Psychology*, 46(3), 174–178. DOI: <https://doi.org/10.1027/1864-9335/a000236>

- Meier, B. P., Robinson, M. D., & Clore, G. L.** (2004). Why Good Guys Wear White: Automatic Inferences About Stimulus Valence Based on Brightness. *Psychological Science*, 15(2), 82–87. DOI: <https://doi.org/10.1111/j.0963-7214.2004.01502002.x>
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H.** (1957). *The Measurement of Meaning*. Urbana, Chicago, and London: University of Illinois Press.
- Saito, M.** (1996). A comparative study of color preferences in Japan, China and Indonesia, with emphasis on the preference for white. *Perceptual and Motor Skills*, 83(1), 115–128. DOI: <https://doi.org/10.2466/pms.1996.83.1.115>
- Schietecat, A. C., Lakens, D., IJsselstein, W. A., & De Kort, Y. A. W.** (2018). Predicting Context-Dependent Cross-Modal Associations with Dimension-Specific Polarity Attributions Part 1 – Brightness and Aggression. *Collabra: Psychology*, 4(1), 14. DOI: <https://doi.org/10.1525/collabra.110>
- Soriano, C., & Valenzuela, J.** (2009). Emotion and colour across languages: implicit associations in Spanish colour terms. *Social Science Information*, 48(3), 421–445. DOI: <https://doi.org/10.1177/0539018409106199>
- Specker, E., Leder, H., Rosenberg, R., Hegelmaier, L. M., Brinkmann, H., Mikuni, J., & Kawabata, H.** (2018). The universal and automatic association between brightness and positivity. *Acta Psychologica*, 186, 47–53. DOI: <https://doi.org/10.1016/j.actpsy.2018.04.007>
- Valdez, P., & Mehrabian, A.** (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, 123(4), 394–409. DOI: <https://doi.org/10.1037/0096-3445.123.4.394>
- Vö, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M.** (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534–538. DOI: <https://doi.org/10.3758/BRM.41.2.534>

Peer review comments

The author(s) of this paper chose the Open Review option, and the peer review comments are available at: <http://doi.org/10.1525/collabra.168.pr>

How to cite this article: Specker, E., & Leder, H. (2018). Looking on the Bright Side: Replicating the Association between Brightness and Positivity. *Collabra: Psychology*, 4(1): 34. DOI: <https://doi.org/10.1525/collabra.168>

Senior Editor/Editor: Rolf Zwaan

Submitted: 15 May 2018

Accepted: 13 September 2018

Published: 17 October 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

2.2. Implementation in the Lines & Colors Project

Psychological

2.2.1

Specker, E., Forster, M., Brinkmann, H., Boddy, J., Immelmann, B., Goller, J., Pelowski, M., Rosenberg, R., & Leder, H. (*under review*). Is Red Warm for Everyone? Assessing the agreement on aesthetic effects of artworks. *Frontiers*.
Submitted: 22.08.2018.

My Contribution:

Frontiers requests you to note all authors contributions. As noted there:

ES, HB, JB, RR, & HL contributed to the initial idea and research questions of the study.

HB, JB, BI, & RR contributed to the art historical research and foundation.

ES, HB, JB, BI contributed to researching and designing the stimulus material.

ES, MF, & JG contributed to the empirical implementation and study design.

ES, MF, & JG contributed to the data analysis.

All authors contributed to the development of the manuscript.

In this case I was the main author of the manuscript but received feedback from all other authors during the process. In addition, sections on the art historical background were written by my collaborators.

Is Red Warm for Everyone? Assessing the agreement on aesthetic effects of artworks

Eva Specker^{1,2*}, Michael Forster¹, Hanna Brinkmann², Jane Boddy², Beatrice Immelmann², Jürgen Goller¹, Matthew Pelowski¹, Raphael Rosenberg^{2,3}, Helmut Leder^{1,3}

¹Faculty of Psychology, Department of Basic Psychological Research and Research Methods, University of Vienna, Vienna, Austria

²Faculty of Historical and Cultural Studies, Department of Art History, University of Vienna, Vienna, Austria

³Research Platform Cognitive Science, University of Vienna, Vienna, Austria

* Correspondence:

Eva Specker

eva.specker@univie.ac.at

Keywords: empirical aesthetics¹, universality², aesthetic effects³, psychology⁴, associations⁵

Abstract

The idea that simple visual elements such as colors and lines are associated with specific effects—for example red being warm—is intuitive to all of us. Such associations have formed a basis for the description of artworks since the 18th century and are still fundamental to discourses on art today. Art historians might describe a painting where red is dominant as warm, aggressive, or lively. However, do beholders actually share their perception of these qualities? We tested whether and to what extent the assumption of universality (sharing of perceived qualities) is justified. For the first time, complex artworks and single elements (lines and colors) extracted from them were used. Participants rated these stimuli on 14 dimensions derived from art literature and empirical aesthetics. We examined on which of the dimensions there is agreement and investigated the influence of art expertise, comparing art historians with lay people. In one study and its replication, we found significantly lower agreement in the effects of artworks and their single elements than expected. For the whole artworks, participants agreed on the effects of warm-cold, heavy-light, and happy-sad, but not on the 11 other scales. Further, we found that the image type (artwork or its constituting elements) was a major factor influencing agreement. Art expertise did not play a significant role and agreement was low on dimensions usually of interest in empirical aesthetics (e.g., like-dislike). Our results pose a challenge to the practice of interpreting artworks based on their aesthetic effects, as these effects may not be as universal as previously thought.

1. Introduction

In Goethe's seminal treatise about colors, he wrote: "The colours on the plus side are yellow, red-yellow (orange), yellow-red (minium, cinnabar). The feelings they excite are quick, lively, aspiring" (Eastlake, 1840, p.306). This statement was part of a chapter in which Goethe examined the aesthetic effects of colors on perceivers, in which he also stated that "red-yellow gives an impression of warmth and gladness" (Eastlake, 1840, p.309). Because of such effects, he considered

color a central aspect of visual art: “colour considered as an element of art, may be made subservient to the highest aesthetical ends” (Eastlake, 1840, p.304). In the art historical literature, attributions such as “warmth” are known as *aesthetic effects*. Aesthetic effects are seen as specific responses in viewers of art that are caused by aesthetic forms, especially lines and colors (Brinkmann et al., 2018). Thus, in art historical discourse an artwork is seen as warm because it has warm colors in it.

The idea that simple visual elements such as colors (but also lines) elicit specific effects in humans is one of the foundations of modern discourses on art since the 18th century and is still crucial for art historians today (Brinkmann et al., 2018). Under this view, if an artwork evokes an impression of warmth, this impression can be explained by the occurrence of red tones in the artwork, and vice versa. Theorists describing the link between colors and forms, and their effects on humans generally conceive these effects as objective (aesthetic) properties of the stimuli. Aesthetic properties (or “aesthetic effects”) such as warmth are often treated as essentially “perceptual” (Robson, 2014), that is, we can “see” that a line is dynamic in the same way as that we can see that it is curved. From a psychological perspective, the terminology of an aesthetic effect, let alone of this effect being perceptual, is strange. In psychology we would rather speak of conceptual associations between separate concepts (i.e. the concept “red” and the concept “warmth”), which lead to an aesthetic experience (Leder, Belke, Oeberst, & Augustin, 2004). We, however, mention this “perceptual” approach since it is first the most common approach of theorists and second it leads theorists to make an implicit assumption of universality¹. This means that they claim universality for their individual perception, suggesting that other perceivers have the same or similar responses to artworks. It is worth noting that art historians and art theorists do not differentiate between “perceptual” and “conceptual” themselves; these are psychological categories that art historians do not use. Perceptual was used by Robson (2014) who is one of the few exceptions who tried capture this in more psychological terms in order to describe how theorists generally use aesthetic effects meaning that they treat aesthetic effects as something that can be “seen”. However, it is not clear, whether and to what extent aesthetic effects are individual or universally shared. In this study, we addressed this issue by testing to what degree ratings on specific aesthetic dimensions of abstract artworks and their single elements (colors and lines) are shared amongst perceivers.

1.1. Theoretical background

Striving to find “general laws”, research in empirical aesthetics sought universality in terms of preference for, or beauty of e.g., the golden section (McManus et al., 2010; McManus & Weatherby, 1997), fractals (Spehar et al., 2003), rectangles (McManus & Wu, 2013), paintings (Leder et al., 2016) and across different aesthetic domains (Vessel & Rubin, 2010).

Research focusing on associations between basic elements and concepts has exclusively focused on single elements, without embedding those elements in complex, realistic stimuli such as artworks. For example, psychological studies have tested the universality of color associations (Specker et al., 2018), color-form associations (Albertazzi et al., 2013; Chen et al., 2015), as well as associations with lines (Lundholm, 1921).

Interestingly, much of what is discussed as an aesthetic effect in art historical literature is reminiscent of what in modern psychology is known as *cross-modal correspondence*. Cross-modal

¹ Though, sometimes also explicit statements can be found, for example Endell (1896) who claimed that every color and every form cause the same feeling in all people.

correspondence refers to the input of two different sensory modalities being congruent, for example perceiving high-pitched vs. low-pitched sounds to resemble bright vs. dark colors (e.g., Hubbard, 1996; Ward, Huckstep, & Tsakanikos, 2006). In the above aesthetic related example, it would be a cross-modal correspondence of a tactile sensation of warmth with the visual input of “warm” colors such as red.

Cross-modal correspondences can already be observed in young (e.g., Mondloch & Maurer, 2004; Walker et al., 2010) and can not only be found in humans but also in animals (e.g., Ludwig, Adachi, & Matsuzawa, 2011).

These types of cross-modal correspondences are particular evident in people with synesthesia, a condition in which experiences in one domain evoke additional experiences in another, as when musical notes or letters of the alphabet evoke colors. An important difference between synesthetes and people without synesthesia is that synesthetes, in this example specifically music-color synesthetes, actually *see* high-pitched notes in brighter colors (Mulvenna & Walsh, 2005) whereas people without synesthesia do not. Nonetheless, people without synesthesia still have the ability to *recognize* cross-modal similarities.

However, this link between crossmodal correspondence and aesthetic theory is not surprising. Historically, many theorists were inspired by linking different senses in a consistent way.

Taken from this, the “perceptual” approach towards aesthetic effects is perhaps not as surprising as it may seem at first hand. Since theorists built on accounts of synesthetic experiences that do include cross-modal correspondence in a perceptual way. In adopting this “perceptual” approach, theorists describing these effects generally make an implicit assumption of universality. This means that they address their individual perception at a universal level, suggesting that other perceivers have the same or similar responses to artworks.

Even though in the history of art theory, the assumption of universality is explicitly or implicitly shared, different theorists have proposed different aesthetic effects for the same color or type of line. Where red for Goethe (Eastlake, 1840) and Edmund Burke (Burke, 1757) was a cheerful color, it was a brutal color to Franz Marc (Cohen, 2013). Such historical discrepancies cast first doubt on the universality of aesthetic effects. Although there have been some attempts for empirical measurements of aesthetic effects (for an overview see Brinkmann et al., 2018), we still lack a systematic test of the assumption of universality.

In this paper, we test the notion of universality empirically, using real artworks and their constituting elements. To trace back shared aesthetic effects, elicited by artworks, to shared aesthetic effects, elicited by their constituting elements, we tested the associations elicited by 1) the artwork as a whole, 2) the combination of all lines of the image, 3) the combination of all colors of the image, 4) the single lines of the image, and 5) the single colors of the image. We manipulated images digitally to separate these elements from the artwork as a whole. This allowed us to investigate the single elements in isolation while retaining stimulus material derived from real artworks. We chose abstract works of art by Wassily Kandinsky, Fritz Winter, and Joan Mirò, since they can be divided in single components. As can be seen in Figure 1, we did not change the placement or size of the single elements. We measured associations of aesthetic effects in 14 different dimensions derived from art literature and empirical aesthetics, based on Brinkmann et al. (2018). This allows for a comparison between the most common dimensions in art literature (e.g., aggressive-peaceful, heavy-light) and the most common dimensions in empirical aesthetics (dislike-like, uninteresting-interesting).

Even though aesthetic effects are treated as essentially “perceptual” (Robson, 2014), it may be the case that some people are more “sensitive” observers than others. Hume (1965), for example, argued that a true judge has “a strong sense, united to delicate sentiment, improved by practice, perfected by comparison, and cleared of all prejudice”(p.17). Similarly, Sibley (1959) argues that aesthetic perception “requires the exercise of taste, perceptiveness, or sensitivity” (p.421). Though it has been argued that (very) few people would meet or come close to meeting Hume’s standards (Hume, 1965; Meskin, 2007; Robson, 2014) or perhaps those of Sibley, we felt that a true test of universality should at least attempt to take into account the notion that some observers are more sensitive than others. We supposed that people who have a higher knowledge of art and interest in art may represent a more sensitive viewer. If these viewers are more sensitive perceivers, and there is a “true” effect, then they should agree more with each other on the aesthetic effects of an artwork than a group of less sensitive viewers. Therefore, we assessed whether people who have a higher knowledge about art and interest in art (we chose expert participants among art history students) will agree more with each other on the aesthetic effects of artworks than people who have little knowledge about and little interest in art.

1.2 Empirical implementation

To measure the proportion of shared evaluation we relied on an index derived from generalizability theory (Brennan, 2001; Shavelson & Webb, 1991). Hönekopp (2006) introduced this method as the beholder index (“bi”). Based on comparisons of variance components, this index allows to quantify the relative proportion of shared to individual variance in evaluations. Following its name, the beholder index reflects how much of the variance is due to private evaluation. For example, if the bi is .75, 75% of the variance is due to private evaluation and by definition 25% is due to shared evaluation. Therefore, to support the notion that aesthetic effects are universally shared, we would need to find low beholder indices where less than half of the variance is due to private evaluation, meaning that more than half of the variance is due to shared evaluation.

In a second step, we investigated the potential role of knowledge about art and interest in art. In order to do this, we collected data of art history students (experts) as well as psychology students (lay people). This allowed us to calculate beholder indices for these two groups separately. If knowledge about art or interest in art leads people to agree more on the aesthetic effects, the art history students should consistently have lower beholder indices than lay people.

2. Method

The bi method estimates variance components that can be interpreted as the observed variance attributed either to the participant or the stimulus (see Leder et al., 2016, page 2 for a comprehensive of the estimation of the variance components). In order to do this, participants need to rate each stimulus twice. The repeated measure allows for estimation not just of how much participants agree with each other on a rating of a stimuli (shared evaluation) but also how much participants agree with themselves on the rating of a stimulus (private evaluation). These variance components can then be combined into the beholder index (“bi”). Hönekopp (2006) provided two different beholder indices: bi_1 and bi_2 . Both indices use the variation explained by the interaction between the participant and the image as an indicator of how much variation is attributable to private evaluation. The crucial difference is that bi_2 additionally uses the variation explained by the participant as an additional indicator for private evaluation. There are two ways to interpret this additional variation: First, one could argue that the differences between participants are not meaningful but simply reflect differences in scale use and not differences in evaluation (bi_1 is indicated). Second, one could argue

that these differences reflect a meaningful part of private evaluation (bi_2 is indicated). We have no strong opinion on which of the two indices is preferable, but rather think that the appropriate value probably lies in between both indices. We chose to use bi_1 in the main body of text and in all the figures. However, to ensure completeness, the tables include bi_1 and bi_2 .

In order to calculate beholder indices Hönekopp (2006) we measured each participant twice. Therefore, participants completed two sessions of one hour each with a minimum of two days in between. We used a blocked design where in each session people first completed Block 1, then an unrelated filler task and then Block 2. Block 1 was always a rating task of aesthetic effects for the single lines and colors with a total of 28 stimuli (see Figure 1, panel d and e). Block 2 was always a rating task of aesthetic effects for the combination of lines and colors (6 stimuli) as well as the three artworks as a whole (see Figure 1, panel a, b, and c). The filler task was included to help with concentration and fatigue. We chose this design in order to exclude the influence of seeing the whole artwork on the rating of the individual elements. We anticipated that if participants saw the whole artwork before (or in between) rating the elements the whole artwork could potentially influence their rating of the single elements to a large extent. This design leaves open the possibility for the reverse effect, namely that the rating of the single elements influences the whole artwork. However, we saw this as less problematic for two reasons. First and foremost, theoretically one would assume that the individual elements contribute to the overall impression of an image constructed of these elements. That is, one assumes that lower-level features inform the impression of the whole image as evidenced by research on lower-level features such as curvature (Bar & Neta, 2006; Bertamini, Palumbo, Gheorghes, & Galatsidas, 2016; Leder, Tinio, & Bar, 2011; Silvia & Barona, 2009) or symmetry (Gartus & Leder, 2013; Leder et al., 2018). Thus it would be impossible to exclude this influence experimentally. Additionally, since the single elements are presented randomly and not blocked per artwork it is unlikely that participants can guess which elements belong “together”, reducing the potential influence. Due to the design completion of Block 1 took about 30 minutes, which was then followed by a 10 minute filler task it is unlikely that participants would remember the single elements after this period. It is very likely that they did take the individual elements into account in their judgement of the whole artwork but as mentioned this would be due to the theoretical notion that lower-level features inform the impression of the whole image and as such this effect would also be present if participants did not rate the images in advance.

2.1 Participants

We excluded seven participants because they dropped out from the first to second session, data loss, or there was low variance in the answers that they gave. We collected data of a total sample of 107 participants, ending up with a final sample of 100 participants (24 men, age range: 19-48 years, $M = 24.44$, $SD = 4.73$) with groups of 50 art history students (9 men, age range: 19-48 years, $M = 25.81$, $SD = 6.26$) and 50 psychology students (24 men, age range: 19-43 years, $M = 25.12$, $SD = 5.56$). All participants received a compensation of 30€ for their participation at the end of the second session. The study was carried out in accordance with the Declaration of Helsinki and the study was approved by the local ethical committee of the University of Vienna.

2.2 Materials

2.2.1 Aesthetic Effects

To measure aesthetic effects, we used 14 rating scales, each scale representing one aesthetic effect. Each rating scale was a 7-point Likert type bipolar scale or semantic differential, meaning that

each aesthetic effect was represented by opposite pairs as poles of the scale, for example: warm – cold. Based on Brinkmann et al. (2018) we used the terms: negative-positive, passive-active, lively-still, happy-sad, aggressive-peaceful, soft-hard, warm-cold, heavy-light, smooth-rough, bodily-spiritual, masculine-feminine, intrusive-cautious, dislike-like, and uninteresting-interesting. In each case the left term was represented by 1 and the right term by 7. Left term and right term are here used to refer to the opposite sides of the dimension such as “warm-cold”, in this case warm would be referred to by 1 and cold by 7. Two terms (liking and interest) were included for their importance to psychologists (i.e. empirical aesthetics) rather than art historians. The question asked was each time “This image appears... warm-cold” (In German: “Dieses Bild wirkt... warm/kalt”). For liking the question was rephrased as “Dieses Bild... Gefällt mir/Gefällt mir nicht” in order to stay grammatically correct².

2.2.2. Stimuli

We used high quality reproductions of the following images: Wassily Kandinsky, *Untitled*, 1934, watercolor, ink, 31.6 × 24.6 cm, Paris, Centre Georges Pompidou; Joan Miró, *Untitled*, 1961, color etching and aquatint, 31.7 × 24 cm, cat raisonnee: 292. Fritz Winter, *Siebdruck 6*, 1950, screen print, 50 × 65 cm. For each image, we digitally extracted the single elements as illustrated in Figure 1, resulting in a total set of 37 stimuli. We selected relatively homogenous stimuli for the purpose of this study given that the agreement between observers has been shown to vary as a function of the heterogeneity of stimuli used (Hönekopp, 2006). All “lines” in the images were black lines, this was done in order to not have our lower-level categories of “lines” and “colors” to be confounded. In addition, all colors were circle-shaped with the exception of one element in the Winter image that had a bow-like shape. To summarize, all color elements were circle shaped and all lines were black lines. All three artworks were abstract artworks that did not consider any element that could not be categorized as a line or color by the description above. As noted and illustrated in Figure 1, we did not change the placement or size of the single elements.

2.2.3. Filler Task

In Session 1, the filler task was the Vienna Art Interest and Art Knowledge questionnaire (VAIAK; Specker et al., 2018). This provided a filler task as well as a manipulation check for the assumed higher art interest and art knowledge of the art history students. In Session 2, the filler task was an unrelated experimental task.

2.3. Procedure

Upon arrival in the lab for the first session, participants received and signed an informed consent form. They were instructed by the experimenter, to follow the instructions on the screen and complete the task (Block 1). After completing Block 1, the experimenter started the filler task for the participant. When the participant finished the filler task, the experimenter started Block 2 for the participant, where again, they were instructed to follow the instruction on the screen and complete the task. Within the blocks the questions were blocked on image, in other words participants saw an image and rated that image for all aesthetic effects before moving to the next image. The question order of the aesthetic effects was randomized, as was the image presentation order. Afterwards the first session was finished. Upon arrival in the lab for the second session, participants went through

² Note that an English and German differ grammatically here.

the same procedure, the only exception being not signing the informed consent again. After the second session was finished, they were thanked for their participation and received their compensation.

3. Results

Art history students were more interested in art ($M = 64.9$) and had more knowledge about art ($M = 15.46$) than lay people ($M_{\text{interest}} = 39.82$, $M_{\text{knowledge}} = 5.24$), both $ps < .001$. This indicated that our selection of art history students and psychology students was a suitable way to select two groups that differed in a meaningful way on these variables.

As outlined above, the beholder index reflects how much of the variance is due to private evaluation, so if the bi is .75 this means that 75% of the variance is due to private evaluation (and by definition that 25% is due to shared evaluation). Therefore, to support the notion that aesthetic effects are universally shared we would need to find low beholder indices where at least less than half of the variance should be due to private evaluation (that is $bi < .50$).

As mentioned by Leder et al. (2016), methods to calculate standard errors and confidence intervals for variance components exist (Burdick & Graybill, 1992), however, because variance components are usually not normally distributed: “summarizing the precision of a variance component estimate by giving an approximate standard error is woefully inadequate” (Bates, 2010; p.19). We will therefore not report any standard errors or confidence intervals. Unfortunately, this makes it impossible to test for significant differences between bi’s, either between the different image types or between the different groups of participants. Therefore, all comparisons here are purely descriptive and should be interpreted as such.

First, we calculated beholder indices of the aesthetic effects over our total sample, separately for each image type as reported in Table 1 and illustrated in Figure 2 (Panel 1). Descriptive means and standard deviations separated by image types and session can be found in the Supplementary Material, Table S1. This showed that across image types there was consistently a higher level of private evaluation for the majority of terms. Interestingly, image type seems to play a major role in determining the agreement amongst raters. For the whole artwork, the aesthetic effects of warm, heavy, and happy were mainly determined by shared evaluation.

< Figure 2>

<Table 1>

This seems to indicate that aesthetic effects are not universally shared but rather are highly determined by private evaluation. In addition, there are large differences between the different image types, essentially making agreement dependent on image type. For the artworks as a whole people agree on the effects of warm-cold, heavy-light, and happy-sad. However, when considering different image types this pattern changes. It is also worth noting, however, that in general people agreed more on the aesthetic effects of the artwork (3 out of 14 effects that are mainly determined by shared evaluation) than on the aesthetic effect of the elements of those artworks (with the exception of the combination of colors). Furthermore, generally speaking the two variables that psychologists are mostly interested in, liking and interest, are highly determined by private evaluation. That they are not strongly agreed upon, is in line with previous findings (Leder et al., 2016; Vessel & Rubin, 2010). In sum, what this seems to indicate is that there is no support for the assumption of universality of aesthetic effects.

In a second step, we calculated beholder indices of the aesthetic effects over our two groups (art history students and lay people) for each image type separately as reported in Table 2 and illustrated in Figure 3. This shows that experts do not consistently agree more than lays (in which case the dark grey bar of experts should be consistently lower than the light grey bar). The only exception is the combination of lines where experts do consistently agree more on the aesthetic effects than lays. As mentioned, there is no way to test for significant differences in the beholder indices. Thus, small differences would not necessarily be significant if subjected to a test.

< Table 2>

< Figure 3>

In sum, what this seems to indicate is that there is no empirical support for the notion that art experts would consistently agree more on the aesthetic effects due to their shared knowledge. Rather, it seems that agreement on aesthetic effects is highly dependent on the specific effect in relation to the image type. As outlined above, the absence of an adequate method to calculate standard errors and confidence intervals for variance components makes it hard to assess the precision of the point-estimates of the bis. We therefore carried out a replication study in order to address this issue.

4. Replication

For the replication study, we only included lay participants as in our original study we found no effect of expertise. Our sample consisted of 50 lay participants (17 men, age range: 18-38 years, $M = 21.98$, $SD = 4.23$). Methods and analysis were identical to the original study. The findings are reported in Table 3 and illustrated in Figure 2 right-side panel (Panel 2). Descriptive means and standard deviations separated by image types and session can be found in the Supplementary Material, Table S2. A comparison between Panel A and Panel B of Figure 1 reveals that we replicated the main findings of our original study. However, even though the general pattern of findings was similar, it is also apparent that the point estimate of the bi can shift. It therefore seems inadvisable to interpret the point estimate of the bi in and of its own. A better-suited way is to interpret the tendency whether the rating is mainly determined by private or mainly by shared evaluation.

<Table 3>

5. Discussion

This study is—to the best of our knowledge—the first attempt to test the universality of aesthetic effects by using original material (high quality reproductions of artworks) and rating scales based on the terms most common in art discourses. The results show that in general we cannot assume universality of aesthetic effects. Across image types most terms were not agreed upon and there was no consistent pattern which terms people agreed on. This seriously challenges the practice of universalizing aesthetic effects. A practice largely used in historical discourses and contemporary discourses on art, in art criticism as well as scholarly art history.

The finding of lacking universality in rating aesthetic effects is reinforced by the finding that we did not see more agreement among experts: The individual differences in knowledge about art or interest in art do not matter for the degree of agreement in the perception of aesthetic effects: If

people have more knowledge or interest, they do not agree more and have no higher level of shared evaluation than people who have less knowledge or interest.

It has to be noted that when taking into account the whole artwork we did find consistent agreement on the aesthetic effects of warm-cold, happy-sad, and heavy-light. Thus, our findings do not suggest an absence of any universality they only imply that the universality is not as universal as previously thought. In addition, it has to be mentioned that universality does not need to imply perfectly shared evaluation. For example, we know that people have quite some shared evaluation of faces, however previous research (Leder et al., 2016) has shown that even for faces the β_1 is around .40. It would therefore be too strong to conclude that there is no agreement, nonetheless we can conclude that the agreement is significantly lower than presupposed.

Analyzing the details of the results shows that whereas the amount of agreement is generally poor, it does vary according to the type of images and to the aesthetic dimension. Two variances are particularly surprising: First, there seems to be more agreement when rating complex abstract artworks as a whole, rather than their single elements. This probably means that the traditional assumption whereupon the effect of the whole results of the sum of more elementary effects of single lines and colors. Second, there is less universality in taste (liking and interest) rather than in rating qualitative aesthetic effects (the 12 other scales we used).

The finding that there is more agreement on the complex image than on the single elements is surprising when considering the vast literature on cross-modal correspondence. A reason for this may be that though the colors and lines in our study varied they did not have a systematic variation. That is, if I wanted to test for a cross-modal correspondence between a visual (e.g. colors, lines) and haptic domain (the haptic domain is implemented in our design by the aesthetic effects of smooth-rough, soft-hard, warm-cold) I would normally focus on one aspect of colors, e.g. brightness. Similarly to the research showing that high-pitched vs. low-pitched sounds are connected to bright vs. dark colors (e.g., Hubbard, 1996; Ward, Huckstep, & Tsakanikos, 2006). However, in our design the different color stimuli did not only vary in brightness but also in hue and saturation. Similarly, the lines differed in orientation and e.g. thickness. Seen this way even the lower-level features in our design had several aspects to them, this variation may have obscured some of the cross-modal correspondences. That said, when we investigate these effects in an aesthetic context we are dealing with stimulus material that will always have relatively high variation. That is, when comparing one image to another it is unlikely that we find them only differing in e.g. line thickness. That is not to say that it is not worthwhile to investigate how line thickness contributes to our aesthetic perception but only that if these cross-modal correspondences only occur at a very low-level and do not translate to complex images such as artworks (because there are simply too many other factors influencing our judgement) the theory that they would be applicable to such complex stimuli is seriously challenged. To test the theory that these effects apply to artworks one has to test this theory with artworks. We have done exactly this in our study. Though our results do not indicate that these effects are not applicable, they do indicate that people do not agree on these effects. That is, the correspondences are highly idiosyncratic. Within the domain synesthesia there are both idiosyncratic correspondences and correspondences that are generally agreed up, the latter type are the type of correspondences that are also evident in non-synesthetes and the focus of cross-modal correspondence research. However, based on our research, it seems that in aesthetics we deal with the former, highly idiosyncratic one.

These interpretations are strengthened by the finding that this general pattern was replicable. However, the results of the replication also highlight the imprecision of the point estimates of the β_1 . It seems that the point estimate in and of its own is not a suitable point of reference, rather, it can be

used to describe the central tendency. Thus, one can say from the bi that participant's ratings are either mainly based on private evaluation or mainly based on shared evaluation. In the case that they are mainly based on shared evaluation, we postulate that this means that people agree on these effects. In the end we can say, even though red may be warm for everyone that does not mean that everyone agrees on it being lively.

6. Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7. Author Contributions

ES, HB, JB, RR, & HL contributed to the initial idea and research questions of the study.
HB, JB, BI, & RR contributed to the art historical research and foundation.
ES, HB, JB, BI contributed to researching and designing the stimulus material.
ES, MF, & JG contributed to the empirical implementation and study design.
ES, MF, & JG contributed to the data analysis.
All authors contributed to the development of the manuscript.

8. Funding

The writing of this paper was supported by a grant to Raphael Rosenberg and Helmut Leder by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF: Project number: CS15-036).

9. Acknowledgements

We would like to thank Lisa Hegelmaier and Maximilian Douda for their help with data collection and the preparation of tables and Karl Pani, Armin Plankensteiner, and René Steyer for editing the stimulus material.

10. References

- Albertazzi, L., Da Pos, O., Canal, L., Micciolo, R., Malfatti, M., & Vescovi, M. (2013). The hue of shapes. *Journal of Experimental Psychology. Human Perception and Performance*, 39(1), 37–47. <https://doi.org/10.1037/a0028816>
- Bar, M., & Neta, M. (2006). Humans Prefer Curved Visual Objects. *Psychological Science*, 17(8), 645–648. <https://doi.org/10.1111/j.1467-9280.2006.01759.x>
- Bates, D. (2010). *lme4: Mixed-Effects Modeling with R*. New York, NY: Springer. Retrieved from <http://lme4.r-forge.r-project.org/book/>.
- Bertamini, M., Palumbo, L., Gheorghes, T. N., & Galatsidas, M. (2016). Do observers like curvature or do they dislike angularity? *British Journal of Psychology*, 107(1), 154–178. <https://doi.org/10.1111/bjop.12132>
- Brennan, R. L. (2001). *Generalizability Theory: Statistics for Social Science and Public Policy*. New York, NY: Springer.
- Brinkmann, H., Boddy, J., Immelmann, B., Specker, E., Pelowski, M., Leder, H., & Rosenberg, R. (2018). Ferocious Colors and Peaceful Lines Describing and Measuring Aesthetic Effects. *Wiener Jahrbuch Für Kunstgeschichte*, 65.
- Burdick, R. K., & Graybill, F. A. (1992). *Confidence Intervals on Variance Components*. New York, NY: Marcel Dekker Inc.
- Burke, E. (1757). *A Philosophical Enquiry into the Origin of Our Ideas of the Sublime and Beautiful*. London: John c. Nimmo. Retrieved from <https://www.gutenberg.org/files/15043/15043-h/15043-h.htm>

- 438 Chen, N., Tanaka, K., & Watanabe, K. (2015). Color-shape associations revealed with Implicit
439 Association Tests. *PLoS ONE*, 10(1), 1–15. <https://doi.org/10.1371/journal.pone.0116954>
- 440 Cohen, W. (2013). *August Macke*. New York: Parkstone International.
- 441 Eastlake, C. L. (1840). *Goethe's Theory of Colours*. London: John Murray. Retrieved from
442 <http://www.gutenberg.org/files/50572/50572-h/50572-h.htm>.
- 443 Endell, A. (1896). *Um die Schönheit. Eine Paraphrase über die Münchener Kunstausstellungen*.
444 München.
- 445 Gartus, A., & Leder, H. (2013). The Small Step toward Asymmetry: Aesthetic Judgment of Broken
446 Symmetries. *I-Perception*, 4(5), 361–364. <https://doi.org/10.1068/i0588sas>
- 447 Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of
448 private and shared taste to judgments of facial attractiveness. *Journal of Experimental*
449 *Psychology: Human Perception and Performance*, 32(2), 199–209.
450 <https://doi.org/10.1037/0096-1523.32.2.199>
- 451 Hubbard, T. L. (1996). Synesthesia-like Mappings of Lightness, Pitch, and Melodic Interval. *The*
452 *American Journal of Psychology*, 109(2), 219. <https://doi.org/10.2307/1423274>
- 453 Hume, D. (1965). Of the Standard of Taste, and Other Essays. Retrieved from
454 <https://philpapers.org/rec/HUMOTS-3>
- 455 Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and
456 aesthetic judgments. *British Journal of Psychology*, 95(4), 489–508.
457 <https://doi.org/10.1348/0007126042369811>
- 458 Leder, H., Goller, J., Rigotti, T., & Forster, M. (2016). Private and Shared Taste in Art and Face
459 Appreciation. *Frontiers in Human Neuroscience*, 10, 155.
460 <https://doi.org/10.3389/fnhum.2016.00155>
- 461 Leder, H., Tinio, P. P. L., & Bar, M. (2011). Emotional Valence Modulates the Preference for
462 Curved Objects. *Perception*, 40(6), 649–655. <https://doi.org/10.1068/p6845>
- 463 Leder, H., Tinio, P. P. L., Brieber, D., Kröner, T., Jacobsen, T., & Rosenberg, R. (2018). Symmetry
464 Is Not a Universal Law of Beauty. *Empirical Studies of the Arts*, 027623741877794.
465 <https://doi.org/10.1177/0276237418777941>
- 466 Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high
467 luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans.
468 *Proceedings of the National Academy of Sciences of the United States of America*, 108(51),
469 20661–5. <https://doi.org/10.1073/pnas.1112605108>
- 470 Lundholm, H. (1921). The Affective Tone of Lines: Experimental Researches. *Psychological Review*,
471 28(1), 43–60. <https://doi.org/10.1037/h0072647>
- 472 McManus, I. C., Cook, R., & Hunt, A. (2010). Beyond the Golden Section and Normative
473 Aesthetics: Why Do Individuals Differ so Much in Their Aesthetic Preferences for Rectangles?
474 *Psychology of Aesthetics, Creativity, and the Arts*, 4(2), 113–126.
475 <https://doi.org/10.1037/a0017316>
- 476 McManus, I. C., & Weatherby, P. (1997). The Golden Section and the Aesthetics of Form and

- 477 Composition: A Cognitive Model. *Empirical Studies of the Arts*, 15(2), 209–232.
478 <https://doi.org/10.2190/WWCR-VWHV-2Y2W-91EE>
- 479 McManus, I. C., & Wu, W. (2013). “The square is ... bulky, heavy, contented, plain, good-natured,
480 stupid ...”: A cross-cultural study of the aesthetics and meanings of rectangles. *Psychology of*
481 *Aesthetics, Creativity, and the Arts*, 7(2), 130–139. <https://doi.org/10.1037/a0030469>
- 482 Meskin, A. (2007). Solving the Puzzle of Aesthetic Testimony. In *Knowing Art* (pp. 109–124).
483 Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-5265-1_8
- 484 Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in
485 young children. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2), 133–136.
486 <https://doi.org/10.3758/CABN.4.2.133>
- 487 Mulvenna, C., & Walsh, V. (2005). Synaesthesia. *Current Biology : CB*, 15(11), R399-400.
488 <https://doi.org/10.1016/j.cub.2005.05.039>
- 489 Robson, J. (2014). Aesthetic Autonomy and Self-Aggrandisement. *Royal Institute of Philosophy*
490 *Supplement*, 75, 3–28. <https://doi.org/10.1017/S1358246114000265>
- 491 Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory: A Primer*. Newbury Park, CA:
492 SAGE.
- 493 Sibley, F. (1959). Aesthetic Concepts. *The Philosophical Review*, 68(4), 421.
494 <https://doi.org/10.2307/2182490>
- 495 Silvia, P. J., & Barona, C. M. (2009). Do People Prefer Curved Objects? Angularity, Expertise, and
496 Aesthetic Preference. *Empirical Studies of the Arts*, 27(1), 25–42.
497 <https://doi.org/10.2190/EM.27.1.b>
- 498 Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H.
499 (2018). The Vienna Art Interest and Art Knowledge (VAIAK) Questionnaire: A Unified and
500 Validated Measure of Art Interest and Art Knowledge. *Psychology of Aesthetics, Creativity, and*
501 *the Arts*. <https://doi.org/10.1037/aca0000205>
- 502 Specker, E., Leder, H., Rosenberg, R., Hegelmaier, L. M., Brinkmann, H., Mikuni, J., & Kawabata,
503 H. (2018). The universal and automatic association between brightness and positivity. *Acta*
504 *Psychologica*, 186, 47–53. <https://doi.org/10.1016/j.actpsy.2018.04.007>
- 505 Spehar, B., Clifford, C. W. G., Newell, B. R., & Taylor, R. P. (2003). Universal aesthetic of fractals.
506 *Computers & Graphics*, 27(5), 813–820. [https://doi.org/10.1016/S0097-8493\(03\)00154-7](https://doi.org/10.1016/S0097-8493(03)00154-7)
- 507 Vessel, E. A., & Rubin, N. (2010). Beauty and the beholder: Highly individual taste for abstract, but
508 not real-world images. *Journal of Vision*, 10(2), 1–14. <https://doi.org/10.1167/10.2.18>
- 509 Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010).
510 Preverbal Infants’ Sensitivity to Synaesthetic Cross-Modality Correspondences. *Psychological*
511 *Science*, 21(1), 21–25. <https://doi.org/10.1177/0956797609354734>
- 512 Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-Colour Synaesthesia: to What Extent Does it
513 Use Cross-Modal Mechanisms Common to us All? *Cortex*, 42(2), 264–280.
514 [https://doi.org/10.1016/S0010-9452\(08\)70352-6](https://doi.org/10.1016/S0010-9452(08)70352-6)

515

516

10.1 Tables

517 Table 1

518 *Beholder Indices Over the Total Sample of the Original Study Separated by Image Type*

	Artworks		Combination of Lines		Combination of Colors		Lines		Colors	
Terms	bi ₁	bi ₂	bi ₁	bi ₂	bi ₁	bi ₂	bi ₁	bi ₂	bi ₁	bi ₂
Warm-cold	.34	.36	.96	.98	.33	.34	.68	.82	.42	.52
Heavy-light	.43	.52	.46	.55	.55	.63	.61	.70	.58	.75
Happy-sad	.41	.47	.88	.93	.20	.28	.60	.67	.36	.53
Masculine- feminine	.50	.50	.79	.79	.32	.39	.64	.72	.60	.70
Smooth-rough	.53	.57	.71	.75	.79	.83	.64	.73	.74	.84
Soft-hard	.54	.63	.89	.90	.67	.69	.57	.65	.69	.81
Lively-still	.63	.71	.83	.87	.31	.45	.36	.43	.53	.74
Intrusive-cautious	.64	.67	.84	.91	.60	.74	.48	.60	.63	.82
Negative-positive	.64	.64	.81	.81	.43	.50	.64	.68	.70	.81
Aggressive- peaceful	.67	.73	.88	.89	.96	.97	.67	.71	.77	.88
Passive-active	.75	.86	.92	.95	.35	.42	.26	.34	.65	.82
Dislike-like	.86	.86	.84	.85	.96	.96	.73	.79	.86	.93
Bodily-spiritual	.92	.94	.96	.98	.74	.85	.92	.94	.65	.84
Uninteresting- interesting	.97	.97	.84	.87	.94	.95	.54	.62	.86	.92

519

520 Table 2

521 *Beholder Indices Over the Total Sample of the Original Study Separated by Image Type and*
 522 *Expertise*

Terms	Expertise	Artworks		Combination of Lines		Combination of Colors		Lines		Colors	
		bi ₁	bi ₂	bi ₁	bi ₂	bi ₁	bi ₂	bi ₁	bi ₂	bi ₁	bi ₂
Warm-cold	Lay	.31	.35	1	1	.32	.33	.69	.81	.41	.51
	Expert	.37	.39	.86	.97	.35	.35	.67	.83	.41	.52
Heavy-light	Lay	.51	.54	.60	.63	.50	.56	.77	.82	.50	.74
	Expert	.38	.51	.28	.47	.62	.71	.44	.57	.59	.72
Happy-sad	Lay	.36	.46	.86	.87	.15	.27	.56	.63	.31	.55
	Expert	.46	.49	.93	.98	.26	.30	.64	.71	.41	.53
Masculine- feminine	Lay	.41	.43	.77	.77	.30	.36	.69	.73	.43	.61
	Expert	.54	.54	.79	.80	.36	.43	.59	.71	.74	.79
Smooth-rough	Lay	.55	.55	.80	.80	.67	.73	.73	.78	.67	.80
	Expert	.54	.61	.61	.72	.92	.94	.50	.65	.81	.88
Soft-hard	Lay	.58	.66	.98	.98	.62	.62	.59	.70	.59	.77
	Expert	.51	.60	.68	.79	.69	.77	.54	.60	.77	.84
Lively-still	Lay	.57	.67	.88	.89	.28	.44	.41	.51	.62	.79
	Expert	.70	.76	.79	.84	.34	.47	.31	.32	.43	.68
Intrusive-cautious	Lay	.70	.70	1	1	.64	.80	.51	.64	.70	.88

Is Red Warm for Everyone? Assessing the agreement on aesthetic effects of artworks.

Negative-positive	Expert	.52	.68	.59	.82	.59	.69	.44	.56	.55	.73
	Lay	.68	.68	.87	.87	.31	.41	.64	.65	.68	.80
Aggressive-peaceful	Expert	.63	.63	.69	.76	.54	.58	.63	.70	.73	.83
	Lay	.70	.71	.90	.90	.86	.90	.70	.73	.74	.87
Passive-active	Expert	.60	.76	.81	.86	1	1	.62	.69	.75	.87
	Lay	.61	.76	.95	.96	.32	.39	.23	.36	.68	.85
Dislike-like	Expert	.94	.97	.70	.77	.41	.46	.27	.29	.62	.80
	Lay	.91	.92	.86	.86	.85	.87	.63	.69	.87	.93
Bodily-spiritual	Expert	.81	.82	.84	.86	1	1	.81	.86	.85	.94
	Lay	.91	.94	.97	.99	.71	.79	.92	.93	.75	.88
Uninteresting-interesting	Expert	.94	.96	.89	.94	.77	.91	.91	.94	.53	.81
	Lay	.98	.98	.88	.88	.90	.91	.38	.48	.82	.91
	Expert	.95	.97	.83	.87	.98	.99	.68	.73	.90	.94

523

524

525

526

527

528

529

530

531

532

533 Table 3

534 *Beholder Indices over the Total Sample of the Replication Study Separated by Image Type*

Terms	Artworks		Combination of Lines		Combination of Colors		Lines		Colors	
	bi ₁	bi ₂	bi ₁	bi ₂	bi ₁	bi ₂	bi ₁	bi ₂	bi ₁	bi ₂
Warm-cold	.41	.52	.98	.99	.27	.27	.60	.77	.48	.48
Heavy-light	.42	.54	.43	.48	.50	.63	.57	.68	.59	.66
Happy-sad	.21	.40	.84	.85	.24	.24	.58	.67	.44	.51
Masculine- feminine	.49	.56	.72	.74	.32	.33	.62	.66	.53	.61
Smooth-rough	.46	.63	.68	.70	.82	.87	.59	.69	.64	.78
Soft-hard	.56	.64	.71	.71	.67	.69	.54	.64	.60	.74
Lively-still	.52	.59	.73	.81	.47	.52	.35	.46	.60	.80
Intrusive-cautious	.57	.63	.57	.60	.62	.75	.43	.50	.60	.71
Negative-positive	.54	.63	.72	.73	.45	.49	.59	.66	.61	.63
Aggressive- peaceful	.59	.61	.64	.64	.87	.95	.66	.73	.65	.71
Passive-active	.80	.83	.85	.88	.41	.51	.30	.39	.59	.77
Dislike-like	.85	.87	.72	.74	.92	.92	.58	.65	.89	.92
Bodily-spiritual	1	1	1	1	.71	.84	1	1	.65	.81
Uninteresting- interesting	.95	.95	.84	.85	.98	.98	.47	.60	.87	.91

535

10.2 Supplementary Tables

536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561

Is Red Warm for Everyone? Assessing the agreement on aesthetic effects of artworks.

562 Table S1

563 *Means and Standard Deviations (in brackets) for Study 1 per Image Type separated by Session*

Terms	<u>Artworks</u>		<u>Combination of Lines</u>		<u>Combination of Colors</u>		<u>Lines</u>		<u>Colors</u>	
	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
Warm-cold	-.65 (1.69)	-.49 (1.54)	1.32 (1.57)	1.15 (1.43)	-1.17 (1.82)	-1.02 (1.81)	1.26 (1.57)	1.02 (1.61)	-1.16 (1.66)	-1.08 (1.68)
Heavy-light	-.34 (1.75)	-.38 (1.68)	-.7 (1.83)	-.76 (1.57)	1.05 (1.89)	.99 (1.77)	-.25 (1.87)	-.35 (1.81)	.89 (1.69)	.78 (1.72)
Happy-sad	-.52 (1.64)	-.31 (1.62)	.21 (1.61)	.43 (1.53)	-.44 (2.02)	-.56 (1.98)	.28 (1.67)	.38 (1.73)	-.56 (1.68)	-.54 (1.78)
Masculine-feminine	-.06 (1.62)	-.16 (1.59)	-.79 (1.6)	-.83 (1.6)	.88 (1.56)	.89 (1.65)	-.67 (1.69)	-.74 (1.66)	.68 (1.55)	.65 (1.62)
Smooth-rough	.25 (1.76)	.33 (1.66)	1.1 (1.6)	1.07 (1.52)	-1.33 (1.59)	-1.32 (1.62)	.48 (1.75)	.49 (1.71)	-1.08 (1.69)	-1.06 (1.65)
Soft-hard	.21 (1.6)	.12 (1.52)	.85 (1.64)	.9 (1.61)	-1.34 (1.57)	-1.31 (1.53)	.54 (1.79)	.48 (1.79)	-1.22 (1.61)	-1.05 (1.59)
Lively-still	-1.18 (1.64)	-1.24 (1.59)	-1.43 (1.51)	-1.19 (1.64)	.72 (1.99)	.72 (2.08)	-.26 (1.99)	-.18 (1.94)	.45 (1.88)	.45 (1.89)
Intrusive-cautious	-1.02 (1.65)	-1.09 (1.61)	-1.21 (1.66)	-1.31 (1.56)	.98 (1.84)	.86 (1.72)	-.34 (1.88)	-.37 (1.84)	.52 (1.81)	.49 (1.82)
Negative-positive	.46 (1.76)	.34 (1.68)	-.5 (1.72)	-.36 (1.64)	.99 (1.78)	.96 (1.82)	-.4 (1.74)	-.43 (1.72)	.77 (1.7)	.73 (1.68)
Aggressive-peaceful	-.18 (1.7)	-.19 (1.78)	-.92 (1.7)	-.81 (1.59)	1.43 (1.56)	1.4 (1.55)	-.11 (1.77)	-.13 (1.77)	.97 (1.74)	.93 (1.71)
Passive-active	1.19 (1.6)	1.26 (1.41)	1.31 (1.66)	1.09 (1.61)	-.54 (1.87)	-.5 (1.87)	.17 (2.03)	.07 (1.97)	-.51 (1.77)	-.37 (1.8)
Dislike-like	.47 (2.08)	.4 (2.01)	-.35 (1.94)	-.4 (1.85)	.54 (2)	.8 (1.95)	-.65 (1.93)	-.59 (1.94)	.63 (1.91)	.5 (1.92)
Bodily-spiritual	-.1 (1.76)	-.24 (1.71)	-.45 (1.99)	-.7 (1.7)	.91 (1.75)	.94 (1.74)	-.66 (1.89)	-.59 (1.84)	.72 (1.79)	.78 (1.69)
Uninteresting-interesting	1.16 (1.71)	.92 (1.75)	.45 (1.8)	.41 (1.79)	.10 (1.86)	.23 (1.97)	-.34 (1.96)	-.32 (2.01)	.1 (1.82)	.17 (1.86)

564 **Note:** All scores were recoded from 1 to 7 into -3 to +3 to make the numeric values more interpretable.
565 Negative values thus can be seen as a tendency towards the "left" term of the scale (e.g. "warm" in warm-cold)
566 and positive values as a tendency to the "right" term of the scale (e.g. "cold" in warm-cold). Values around 0
567 thus express a tendency for neither (i.e. the midpoint of the scale).

568 Table S2

569 *Means and Standard Deviations (in brackets) for the Replication Study per Image Type separated by*
570 *Session*

Is Red Warm for Everyone? Assessing the agreement on aesthetic effects of artworks.

571

Terms	<u>Artworks</u>		<u>Combination of Lines</u>		<u>Combination of Colors</u>		<u>Lines</u>		<u>Colors</u>	
	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
Warm-cold	-.73 (1.59)	-.49 (1.59)	1.35 (1.34)	1.19 (1.31)	-1.06 (1.59)	-1.11 (1.66)	1.12 (1.6)	1.02 (1.49)	-1.12 (1.65)	-1.05 (1.61)
Heavy-light	-.36 (1.69)	-.43 (1.48)	-.8 (1.54)	-.64 (1.61)	1.06 (1.56)	.92 (1.63)	-.21 (1.74)	-.32 (1.69)	.66 (1.69)	.66 (1.57)
Happy-sad	-.27 (1.45)	-.4 (1.44)	.57 (1.37)	.45 (1.41)	-.35 (1.72)	-.66 (1.79)	.39 (1.6)	.38 (1.6)	-.59 (1.64)	-.44 (1.66)
Masculine-feminine	.17 (1.58)	-.02 (1.56)	-.7 (1.54)	-.67 (1.47)	.78 (1.62)	.81 (1.53)	-.46 (1.58)	-.55 (1.55)	.55 (1.57)	.51 (1.6)
Smooth-rough	.06 (1.68)	.27 (1.68)	.86 (1.45)	.74 (1.42)	-1.32 (1.48)	-1.19 (1.53)	.39 (1.65)	.42 (1.61)	-1.11 (1.51)	-1.03 (1.55)
Soft-hard	-.07 (1.5)	.02 (1.54)	.78 (1.58)	.72 (1.59)	-1.33 (1.31)	-1.42 (1.44)	.42 (1.66)	.49 (1.69)	-1.29 (1.42)	-.95 (1.55)
Lively-still	-1.12 (1.59)	-1.05 (1.59)	-1.19 (1.53)	-1.07 (1.52)	.51 (2.02)	.23 (1.91)	-.14 (1.89)	-.15 (1.89)	.61 (1.7)	.46 (1.79)
Intrusive-cautious	-1.04 (1.66)	-1.24 (1.6)	-1.28 (1.53)	-1.33 (1.49)	1.1 (1.54)	.83 (1.63)	-.14 (1.85)	-.19 (1.83)	.55 (1.67)	.42 (1.71)
Negative-positive	.48 (1.66)	.44 (1.61)	-.59 (1.5)	-.6 (1.4)	.9 (1.58)	.96 (1.59)	-.5 (1.65)	-.48 (1.66)	.7 (1.57)	.83 (1.47)
Aggressive-peaceful	-.24 (1.68)	-.33 (1.72)	-.95 (1.53)	-.96 (1.51)	1.51 (1.38)	1.34 (1.41)	0 (1.72)	-.2 (1.64)	.9 (1.58)	.90 (1.53)
Passive-active	.94 (1.66)	1.2 (1.43)	.98 (1.62)	1.32 (1.33)	-.77 (1.65)	-.51 (1.72)	.12 (1.97)	-.02 (1.88)	-.44 (1.54)	-.32 (1.56)
Dislike-like	.36 (2.08)	.29 (2.08)	-.22 (1.8)	-.36 (1.7)	.7 (1.79)	.72 (1.71)	-.8 (1.83)	-.66 (1.79)	.42 (1.65)	.5 (1.57)
Bodily-spiritual	.07 (1.76)	-.19 (1.53)	-.63 (1.77)	-.77 (1.65)	.74 (1.68)	.86 (1.61)	-.53 (1.78)	-.65 (1.65)	.46 (1.72)	.67 (1.61)
Uninteresting-interesting	1.06 (1.66)	.94 (1.76)	.51 (1.65)	.64 (1.61)	-.02 (1.68)	.11 (1.69)	-.42 (2)	-.38 (1.89)	.12 (1.68)	.23 (1.54)

572 **Note:** All scores were recoded from 1 to 7 into -3 to +3 to make the numeric values more interpretable.
573 Negative values thus can be seen as a tendency towards the "left" term of the scale (e.g. "warm" in warm-cold)
574 and positive values as a tendency to the "right" term of the scale (e.g. "cold" in warm-cold). Values around 0
575 thus express a tendency for neither (i.e. the midpoint of the scale).

11. Figure Captions

Figure 1. Separation from lines and colors from the artwork Wassily Kandinsky, Untitled, 1934, watercolor, ink, 31.6×24.6 cm, Paris, Centre Georges Pompidou. (A) Whole artwork, (B) the combination of colors, (C) combination of lines, (D) single colors and, (E) single lines.

Figure 2. Histogram of the beholder index as separated by experts and type: (A) artwork, (B) combination of colors, (C) combination of lines, (D) single colors, and (E) single lines. The left-side panel (Panel 1) shows the data from the original study ($N = 100$) and the right-side panel (Panel 2) shows the data from the replication study ($N = 50$).

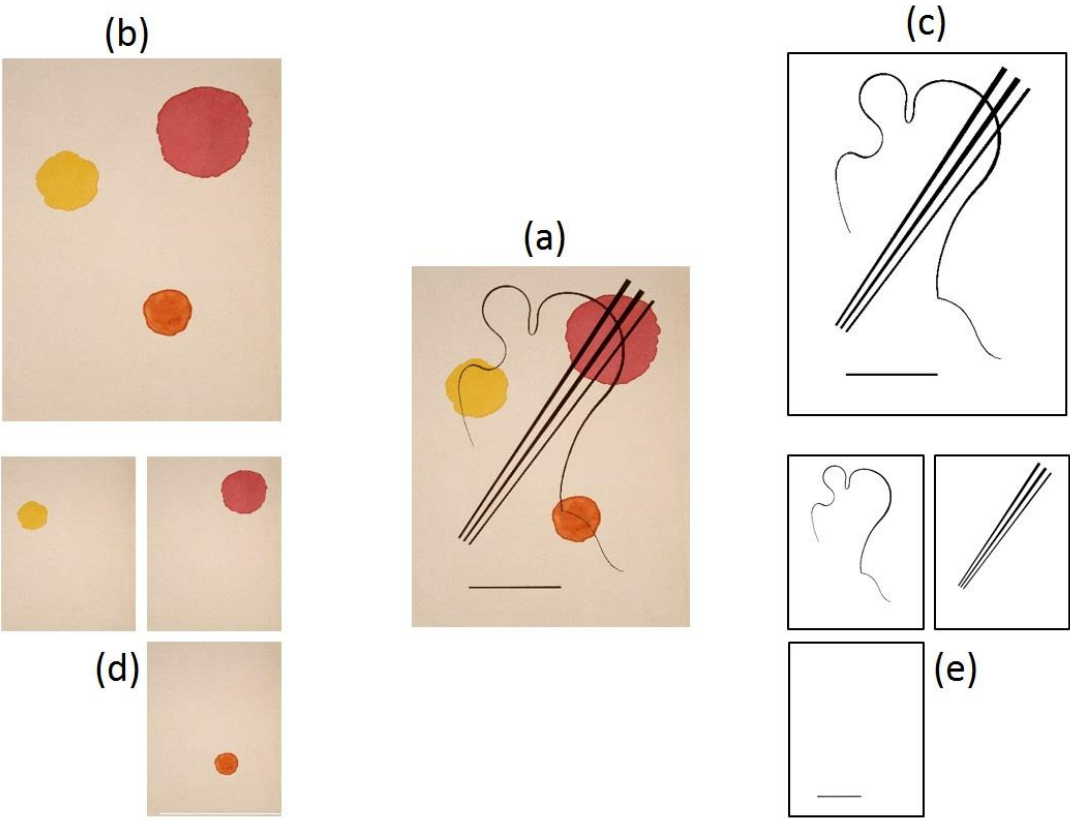
Figure 3. Histogram of the beholder index as separated by experts (light gray) and lays (dark gray) and image type: (A) artwork, (B) combination of colors, (C) combination of lines, (D) single colors, and (E) single lines.

12. Data Availability Statement

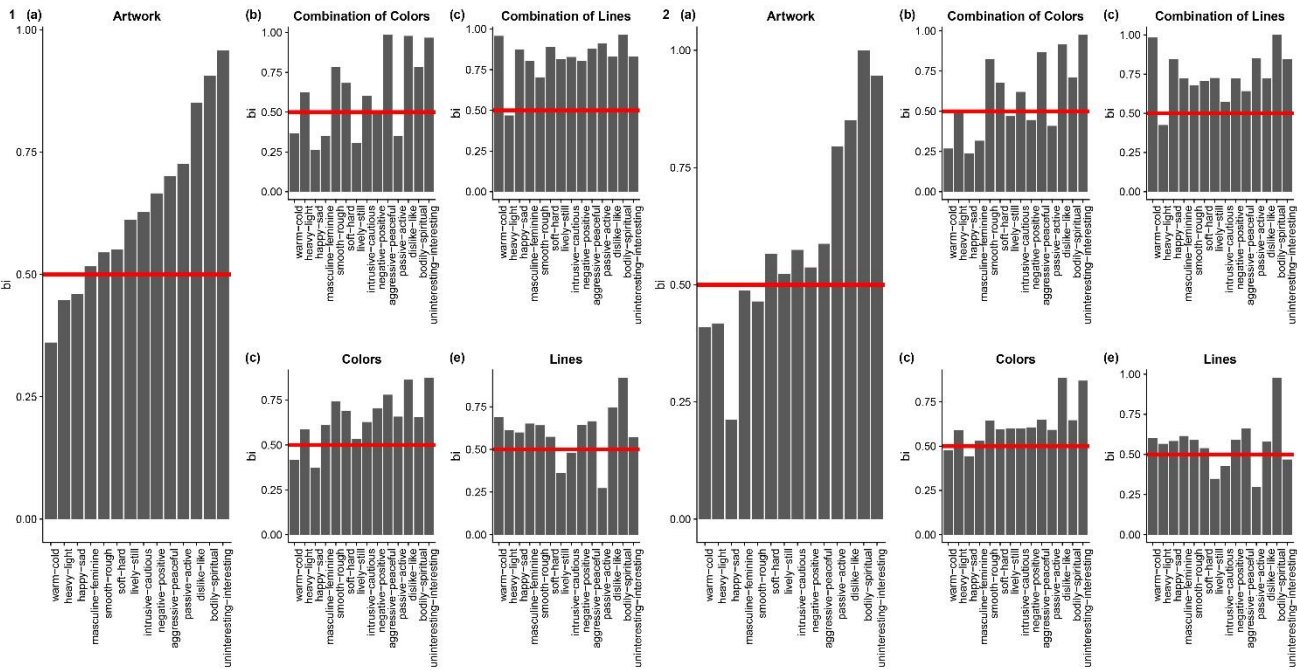
The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher. In addition, all images used as well as experimental scripts will be shared with qualified researchers upon request. Due to copyright issues we are unable to share this material openly.

Figures

Figure 1



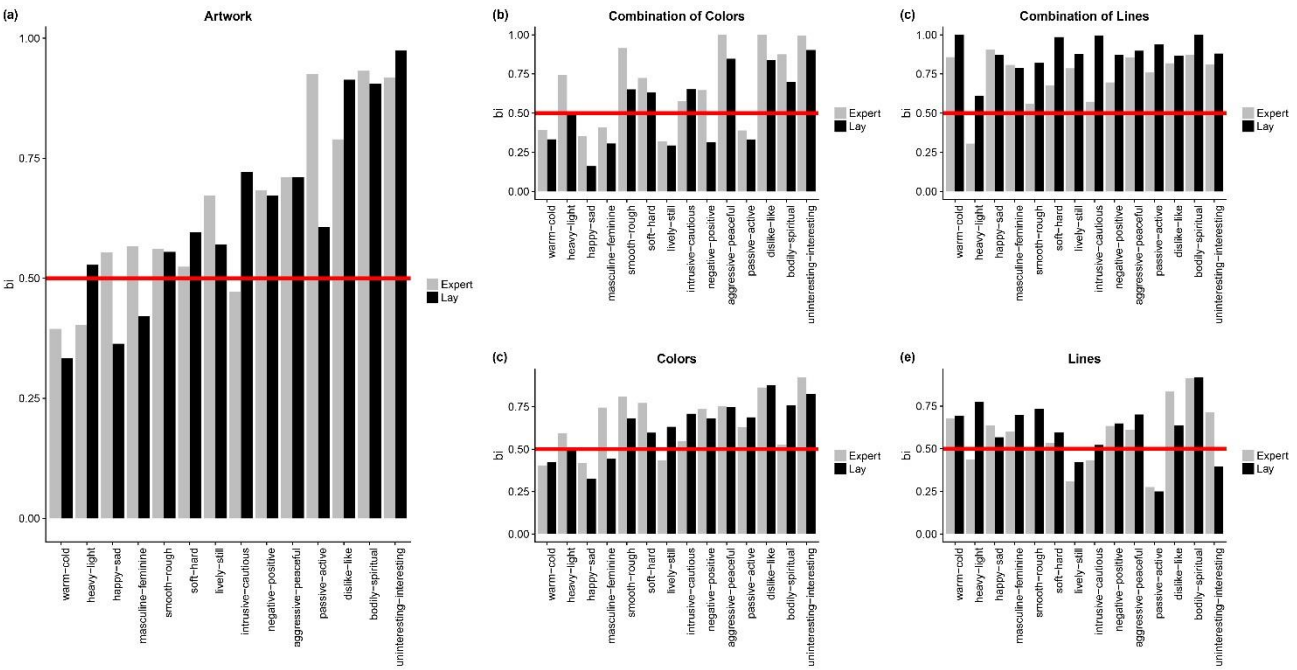
620 Figure 2



621

622

623 Figure 3



624

2.2.2

Specker, E., Fried, E. I., Rosenberg, R., & Leder, H. (*under review*). Sensual Yellows and Moving Lines: Proposing the aesthetic effect network (AEN). *Perspectives on Psychological Science*.

Submitted: 12.02.2019.

My Contribution:

In this case, I came up with the underlying concept and idea of this paper. EF then aided me in study design, refinement of the theory, and interpretation of the results. I collected and analyzed the data. RR and HL offered a wider view of the background in empirical aesthetics and art history/theory, though I was mainly responsible for the literature review. I was the main author of the manuscript but through close collaboration with EF. Once a more refined version of the manuscript was established, it was shared with RR and HL who provided feedback on consecutive versions.

Sensual Yellows and Moving Lines:
A Network Model of Aesthetic Effects

Eva Specker^{1,3}

Eiko I. Fried²

Raphael Rosenberg^{3,4}

Helmut Leder^{1,4}

¹ Department of Basic Psychological Research and Research Methods, Faculty of Psychology,
University of Vienna, Vienna, Austria

² Department of Clinical Psychology, Faculty of Social and Behavioural Sciences, Leiden
University, Leiden, The Netherlands

³ Department of Art History, Faculty of Historical and Cultural Studies, University of Vienna,
Vienna, Austria

⁴ Research Platform Cognitive Science, University of Vienna, Vienna, Austria

Author Note:

The study was supported by a grant to RR and HL by Wiener Wissenschafts-,
Forschungs- und Technologiefond (WWTF: Projectnumber: CS15-036).

Correspondence concerning this article should be addressed to Eva Specker,
Department of Basic Psychological Research and Research Methods, Faculty of Psychology,
University of Vienna, Liebiggasse 5, 1010 Vienna, Austria.

E-mail: eva.specker@univie.ac.at.

Abstract

In this paper we propose the aesthetic effect network (AEN) a novel way to conceptualize aesthetic experience, namely as an associative process where having one association leads to the next association. Through a network of associations, this generates an overall aesthetic experience. In art theory, associations between concepts of this kind are referred to as *aesthetic effects*. Conceptualizing aesthetic experience (at least in part) as an associative process provides an explicit account of a specific cognitive process involved in aesthetic experience. We discuss previous work, outline the aesthetic effect network (AEN), and discuss the results of an empirical study (Study 1, N=255) that explored what can be gained from this approach. In addition, we report on a follow-up study (Study 2, N=133) that provides evidence for the AEN over other alternatives. In summary, we propose a new perspective on the processes underlying aesthetic experience of artworks that extends beyond research on aesthetic effects, that can be applied to aesthetic experience in general, and provides a basis for future studies. It promises to facilitate our understanding of why aesthetic experience is both somehow different and also somehow the same for all of us.

Keywords: aesthetic experience, network, network analysis, empirical aesthetics, aesthetic effects

Introduction

Aesthetic experiences are complex experiences that unfold over time, and assume a strong relation between features in artworks and their effects on the beholder. But how does this actually happen? A key concept is *aesthetic effects* which, since the 18th century, have been used in art literature to explain responses to artworks (Brinkmann et al., 2018; Schnackertz, 2010). The assumption is that works of art produce specific aesthetic effects in viewers that are caused by aesthetic forms and elements; for example, that the color red in a painting causes a sensation of warmth. Franz Marc described colors in a letter to August Macke as follows: “Blue is the male principle, astringent and spiritual. Yellow is the female principle, gentle, gay and sensual. Red is matter, brutal and heavy and always the color to be opposed and overcome by the other two” (Macke, 1964, p. 28). Similar descriptions can be found for lines. Kandinsky, for example, postulated: “The same thing can with complete justification be said about the horizontal and the vertical - low and high. The former is lying, the latter is standing, walking, moving about, finally climbing upward. Supporting - growing. Passive - active. Relatively: feminine - masculine” (Kandinsky, 1947, p. 63).

Art literature implicitly assumes that aesthetic effects are universal, i.e. almost the same for everyone (Brinkmann et al., 2018). However, recent empirical findings suggest that this is not the case (Specker et al., *under review*). As illustrated by the quotes above, aesthetic effects are commonly referred to in relation to specific elements of an artwork. Aesthetic effects are often described as co-occurring, i.e. multiple effects might simultaneously be caused by the same element or artwork, but their connection to each other is rarely made explicitly.

In this paper we propose a novel way to conceptualize these effects by explicitly connecting them to each other. We propose that aesthetic effects can directly influence each other; for example, if a viewer sees an artwork as active, this will lead the viewer to also

perceive it higher in liveliness. We conceptualize these effects as an associative process that proposes direct relationships between aesthetic effects using a network approach.

To rethink the notion of aesthetic effects in this sense, also means to rethink the concept of aesthetic experience in general. Most empirical work focuses on the outcomes of aesthetic experience in terms of judgements made about artworks (e.g., Graf & Landwehr, 2015; Hekkert & Van Wieringen, 1996) or feelings aroused by art (e.g., Gerger, Pelowski, & Leder, 2017; Silvia, 2009; Specker et al., 2017). Even though most theoretical models conceptualize aesthetic experience as a process (e.g., Leder, Belke, Oeberst, & Augustin, 2004; Locher, Overbeeke, & Wensveen, 2010; Pelowski, Markey, Forster, Gerger, & Leder, 2017; Tinio, 2013), these models generally include a large amount of potential relevant variables, often with a lack of clear hypotheses regarding the interrelation of the different sub-processes and variables involved. Nonetheless, all include some type of cognitive processing. Conceptualizing aesthetic experience (at least in part) as an associative process gives an explicit account of a specific cognitive process involved in aesthetic experiences. In addition, this conceptualization offers the possibility to gain a more in-depth understanding of the whole process of art perception. Aesthetic experience seen as an associative process constitutes a measurable interaction between the viewer and the artwork, where the artwork can set into motion an associative process within the viewer from which the viewer can freely associate onwards and end up with a general sense of this artwork. However, this perspective is in stark contrast with previous empirical work done on aesthetic effects in the way that it conceptualizes and explicitly interprets relationships between aesthetic effects.

Previous work typically views relationships between aesthetic effects as due to a common cause (e.g., Takahashi, 1985). That is, if active and lively are correlated this is because they are both caused by the latent factor of “Activity” (see below). In our view, relationships between aesthetic effects, should be seen as direct associations between one

aesthetic effect and another. As noted above, many art theorists argue that aesthetic effects are shared amongst perceivers and thus in a way are universal. But, challenging this idea, it seems that people only agree with each other very little – i.e. on which artwork would be “lively” and which would be “still” (Specker et al., *under review*). However, this does not exclude the possibility that the overall structure of associations may be shared amongst perceivers. That is, rather than that people agree that a specific artwork is lively, they nonetheless might agree that if the artwork is lively it is also active. Though pure speculation, it is intuitive to most of us that if we perceive something as lively that we will also see it as active. It seems counterintuitive to connect lively to passive.

In this paper, we will first describe the theoretical model employed by previous research to highlight the contrast with our proposition of an associative process. Second, we will introduce our theoretical model (aesthetic effect network: AEN) in more detail, and discuss the results of an empirical study that was conducted to explore which information can be gained from taking this approach to understand the experiences people have when looking at art. Finally, we will discuss the results of a follow-up study that provides evidence for the AEN over other alternatives.

Prior Theory

Interestingly, a major body of previous empirical work on aesthetic effects consists of artistic research, in which artists experimented with their art in an attempt to create one or the other effect, rather than scientific research (for a more detailed discussion see Brinkmann et al., 2018). The relatively small amount of scientific research that has been done into associations with art, within the field of empirical aesthetics (e.g., Takahashi, 1995), used the Evaluation, Potency, and Activity model (hereafter called EPA model) of Osgood et al. (1957; for a similar model, focused on aesthetic effects, but only based on historical research, see Wellek, 1929). This model was developed to capture semantic space or meaning. To measure

meaning, Osgood et al. (1957) proposed the semantic differential. A semantic differential (or bipolar scale) uses two opposing adjectives on a five or seven point scale such as good-bad, active-passive, or strong-weak.

According to Osgood et al. (1957), adjective pairs can be interpreted as measures of three latent factors: Evaluation, Potency, and Activity. In this case, the aesthetic effects themselves do not have causal power. What this means in terms of aesthetics is that when a person looks at an artwork, they will perceive a certain level of potency in the artwork, which will consequently lead to a related score in the corresponding indicator. Thus, if the potency is high, a person will rate the painting as strong rather than weak, because strong is an indicator of potency. The main difference between this perspective and the model of associative process that we propose is the conceptualization and interpretation of relationships between aesthetic effects such as active-passive being related to lively-still. In the EPA model, this relation exists because they belong to the same factor – they both measure the level of activity. However, we propose to interpret these as direct relationships between different aesthetic effects: active-passive and lively-still are correlated because perceiving an artwork as more active leads to perceiving it as more lively (or vice versa). This presupposes a level of causality where the aesthetic effects can cause one another. In other words, approaches such as Osgood et al. (1957) represent a relationship between artwork and viewer, where the artwork (or elements of the artwork) inspire aesthetic effects in viewers and the viewer is a passive receiver. The AEN model we propose claims a more active role of the viewer, though the initial stimulation comes from the artwork. This stimulation then goes on by itself by engaging the viewer in an associative process. The initial association caused by the artwork can inspire the next association in the viewer and thus the viewer can freely associate onwards throughout the network; this process will lead to an overall aesthetic experience. This is analogous to semantic networks such as word association networks based on word association

tasks such as a free association task in which people are provided with a cue word (e.g. “love”) and write down the next word that comes to mind (e.g. “heart”), these tasks have been widely described as network structures (see e.g., De Deyne & Storms, 2008; Steyvers & Tenenbaum, 2005) and have been used to predict performance in memory retrieval tasks and explain various priming and interference phenomena (Collins & Loftus, 1975; Nelson, McKinney, Gee, & Janczura, 1998).

Where the Osgood et al. (1957) model only ascribes causality to the artwork and EPA factors, our AEN also ascribes a causal role to the individual aesthetic effects stating that they can cause each other. These conceptual ideas are illustrated in Figure 1. The left panel represents the EPA model, the right represents a network model. In the EPA case, the artwork influences the three EPA factors which, in turn, lead to specific ratings on their indicators. In the aesthetic effect network (AEN) that we propose, the artwork directly leads to aesthetic effects; for example, I see the artwork as lively and warm. These aesthetic effects are modelled as nodes (see below) in the network, these nodes then influence the rest of the network. That is, if I see a lively artwork, this can then inspire the association with activity in me. This shows the associative process where the initial stimulation comes from the artwork and from thereon out the viewer can freely associate onwards throughout the network. Below we propose a statistical network model that can estimate the associations among aesthetic effects.

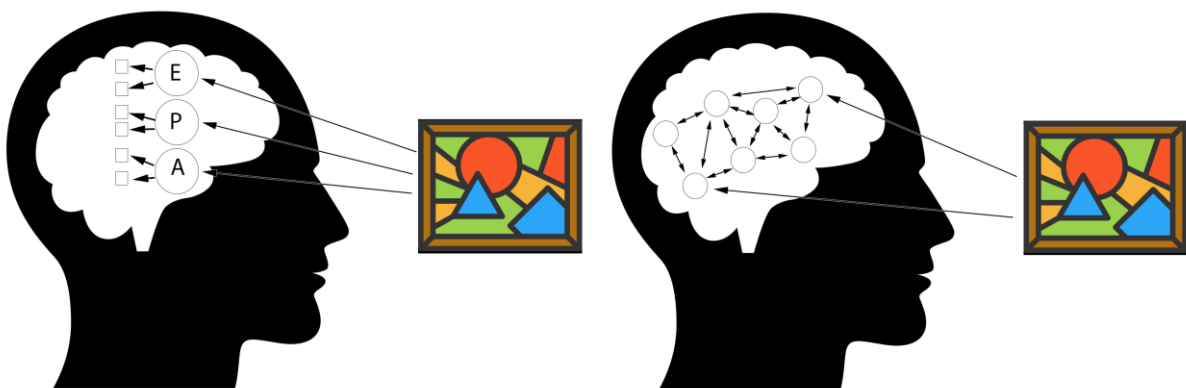


Figure 1. *Conceptual model of how an artwork can influence a person in the EPA model (left) and a network model (right).*

Aesthetic Effect Network (AEN)

Structure

We propose to model this associative process mathematically as a network. Networks consist of *nodes* and *edges*. *Nodes* represent entities and *edges* represent connections between these entities. This general concept is illustrated by the notion of “six degrees of separation”, the idea that any two random people on earth are six or fewer acquaintances apart, originally tested empirically by Travers and Milgram (1969). This is an example of a social network, where nodes represent people, for example Harry and Sally, and edges represent relationships between them. Thus, if Harry and Sally know each other, there will be an edge connecting the node of Harry to the one of Sally.

In the AEN, aesthetic effects (active, lively, etc.) are represented by nodes. Associations between the aesthetic effects are represented as edges. Thus, if active is associated with lively, there will be an edge connecting the nodes. What “associated” in the previous sentence means is not trivial: in a social network, we can simply observe if people are friends (e.g. on Facebook); in aesthetics, we require psychometric models to estimate such relations, which we will explain in more detail below.

We propose that in the AEN, the edges will not be binary (either present or absent) but are more appropriately modelled as continuous, representing the direction and the strength of associations. In addition, active and lively can be positively associated (the more active the more lively) or negatively associated (the more active the less lively). In our case, we conceptualize these edges as partial correlations, because we are interested in the shared unique relations among aesthetic effects active and lively after partialling out all other

aesthetic effects. To go back to the positive edge between active and lively: this means that the variables are dependent conditional on all other items in the network.

On the importance of the Stimulus

Aesthetic effects occur as a reaction to a stimulus, much like attitude networks (Dalege et al., 2016) that represent attitudes towards a specific object or person. We thus propose to call such networks “*stimulus-bound*” since they only exist as a reaction towards a stimulus, in contrast with e.g. social networks that are not stimulus-bound. The stimulus in attitude or aesthetic networks exists in the *external field* (Borsboom, 2017), denoting factors that can influence nodes inside the network from the outside (cf. Figure 1). Thus, the stimulus may cause one (or multiple) nodes in the network but is not itself a part of the network and therefore not represented in the network by a node.

Hybrid models (Fried & Cramer, 2017) provide a useful way to conceptualize stimulus-bound networks. A hybrid model is a mixture of a network model and a common cause model; in the case of stimulus-bound networks, we propose that the stimulus would be seen as the common cause. This means that the stimulus activates certain (not necessarily all) nodes in the network, that is, if Sally perceives a painting, this painting can cause the association of “active” in her. Since the association “active” is a node in the network, the activation can then spread from there throughout the network (see Figure 1). Note that this approach does not answer the question as to how the painting inspires the association of “active”. This perspective holds that a painting can activate the network and then proposes an account of what happens after the initial activation has occurred. This also opens up the possibility for a network approach to individual differences in aesthetic experience. For example, a painting may cause the association “active” in Sally and not in Harry. We will return to this issue in the limitations and future directions section.

Exploring the potential of a network approach empirically

To demonstrate what can be gained from taking this approach to understand the experiences people have when looking at art, we report the results of an empirical study. By doing this we will put forth a preliminary account of the structure of the relationships between different aesthetic effects.

Study 1

Participants

We surveyed 255 participants (92 male, 160 female, 2 identifying as other, 1 unreported, mean age = 20.73, SD = 3.73) in a first year psychology lecture. All participants participated voluntarily. All experiments of the present paper were carried out in accordance with the Declaration of Helsinki and after approval by the local ethical committee of the University of Vienna.

Materials

Aesthetic Effects. To measure aesthetic effects we included 14 items. Each item was a semantic differential, where each aesthetic effect was represented by opposite pairs as poles of the scale, for example: warm – cold. The choice of the semantic differential format was based on a recent historical analysis of the most frequently used terms for aesthetic effects in art literature (Brinkmann et al., 2018). We used the following pairs of terms: negative-positive, passive-active, lively-still, happy-sad, aggressive-peaceful, soft-hard, warm-cold, heavy-light, smooth-rough, bodily-spiritual, masculine-feminine, intrusive-cautious. In addition, we included dislike-like and uninteresting-interesting because they are outcome variables of aesthetic experience that are widely considered as relevant and important by psychologists (e.g. Leder, Gerger, Dressler, & Schabmann, 2012).

Stimuli. As stimuli we used 8 abstract artworks by 5 different artists. Image choice was based on Brinkmann et al. (2018) and (Specker et al., *under review*). All image information and their shorthand titles can be found in Table 1. For ease of reading we chose shorthand titles for each of them (see third row) to be used throughout the text and in all figures and tables. We chose a relatively homogenous set of artworks, because we aimed to average across the artworks later to derive a general network structure. Data on no other artworks was collected, i.e. we did not perform a sub-selection of specific artworks for the purpose of this paper.

Table 1

Artworks used in the study

Artist	Title	Shorthand Title	Year	Collection
Paul Klee	Zeichen in Gelb/ Sign in Yellow	Klee Yellow	1937	Foundation Beyeler, Riehen/Basel
Paul Klee	Blick aus Rot/ Be aware of Red	Klee Red	1937	Zentrum Paul Klee, Bern
Wassily Kandinsky	Regungen/Impulses	Kandinsky Albertina	1928	Albertina Vienna – permanent loan basis, Collection Forberg
Wassily Kandinsky	Untitled	Kandinsky Pompidou	1934	Centre Pompidou, Paris
Richard Mortensen	Øvelsesstykker/ Exercise pieces	Mortensen Pink	1922	Private Collection – sold by Bruun Rasmussen Auctioneers, 6. August 1992, lot 728
Richard Mortensen	Øvelsesstykker/ Exercise pieces	Mortensen Orange	1922	Private Collection – on sale at Bruun Rasmussen Auctioneers, 20. July 1992, lot 729
Joan Mirò	Untitled	Miro	1961	Yvon Taillandier, Pierre Matisse Gallery, cat raisonnee: 292
Fritz Winter	Siebdruck 6/ Silkscreen 6	Winter	1950	Galleri MDA, Sweden, Helsingborg

Procedure

The study took place during a first-year psychology lecture. All images were projected against a black background. Participants were instructed to first look at each image, let it make an impression on them, and then to rate the image on the 14 aesthetic effects within a total viewing time of 75 seconds for each image. After each image, a blank black screen was shown for 5 seconds before moving on to the next image. Ratings could be made either with pen and paper or a digital answering option (tablet, pc, and phone). Pen and paper (N=127) was partly randomized: each participant had the same order for each image, but the question order was different for each image. For digital answering (N=128), the order was random for both images and participants.

Network Estimation

We employed a 3-step analytic strategy similar to Rhemtulla et al. (2016; for a similar, more recent example, see: Fried et al., 2018). In their study, Rhemtulla et al. (2016) used network analysis to compare the interrelation of symptoms in six different datasets of patients with different substance abuse disorders. To do this they first estimated a symptom network for each substance abuse disorder dataset. Then they estimated a global network across all datasets to understand what a general network of substance abuse symptoms would look like. Finally, they estimated a variability network to estimate the variability of the edges across the six individual substance abuse disorder networks. Though we do not deal with mental disorders, our aims in the current study are comparable: we want to compare the interrelation of aesthetic effects for multiple artworks, and estimate what a global network of aesthetic effects looks like. A variability analysis then adds to our understanding of how associations between the aesthetic effects differ depending on which artwork people are looking at. Therefore we followed the approach of Rhemtulla et al. (2016) and performed three analyses. In contrast to Rhemtulla et al. (2016) and based on Fried et al. (2018) we only interpret the

centrality in terms of *expected influence*, because strength takes absolute values and thus does not take negative edges into account, and *betweenness* and *closeness* centrality are often unreliable (e.g., Epskamp, Borsboom, & Fried, 2018). All centrality measures were standardized. All analyses were performed in R using the “*qgraph*” package (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) implementing a Gaussian graphical model (GGM), a network in which edges connecting aesthetic effects represent estimates of partial correlations, including a graphical lasso regularization. In the GGM, edges can be understood as conditional dependence relations among aesthetic effects: If two aesthetic effects are connected in the resulting graph, they are dependent after controlling for all other symptoms. If no edge emerges, aesthetic effects are conditionally independent. Regularization involves estimating a statistical model with an extra penalty for model complexity (Epskamp & Fried, 2018). This procedure leads to networks that are sparse and constrain many of the small coefficients to zero. This means that edges that are likely to be spurious are removed from the model, which leads to networks that are simpler to interpret. Using the graphical lasso to estimate a GGM improves network estimates and leads to a sparse network that describes the data parsimoniously. We used the method of regularization outlined in Epskamp and Fried (2018), using the default setting of the gamma parameter of 0.5. For the global network, we then averaged each edge across the eight separate artwork networks. The GGM with regularization represents the current state-of-the-art for ordinal or continuous data. The final dataset used for the network analyses and the corresponding code can be found on the Open Science Framework: <https://osf.io/zqxbm/>.

Results

Global network

We start with the interpretation of the global network of aesthetic effects (averaged across all 8 individual networks), shown in figure Figure 2. The corresponding centrality plot is shown in Figure 3. The individual networks and their comparison follow afterwards.

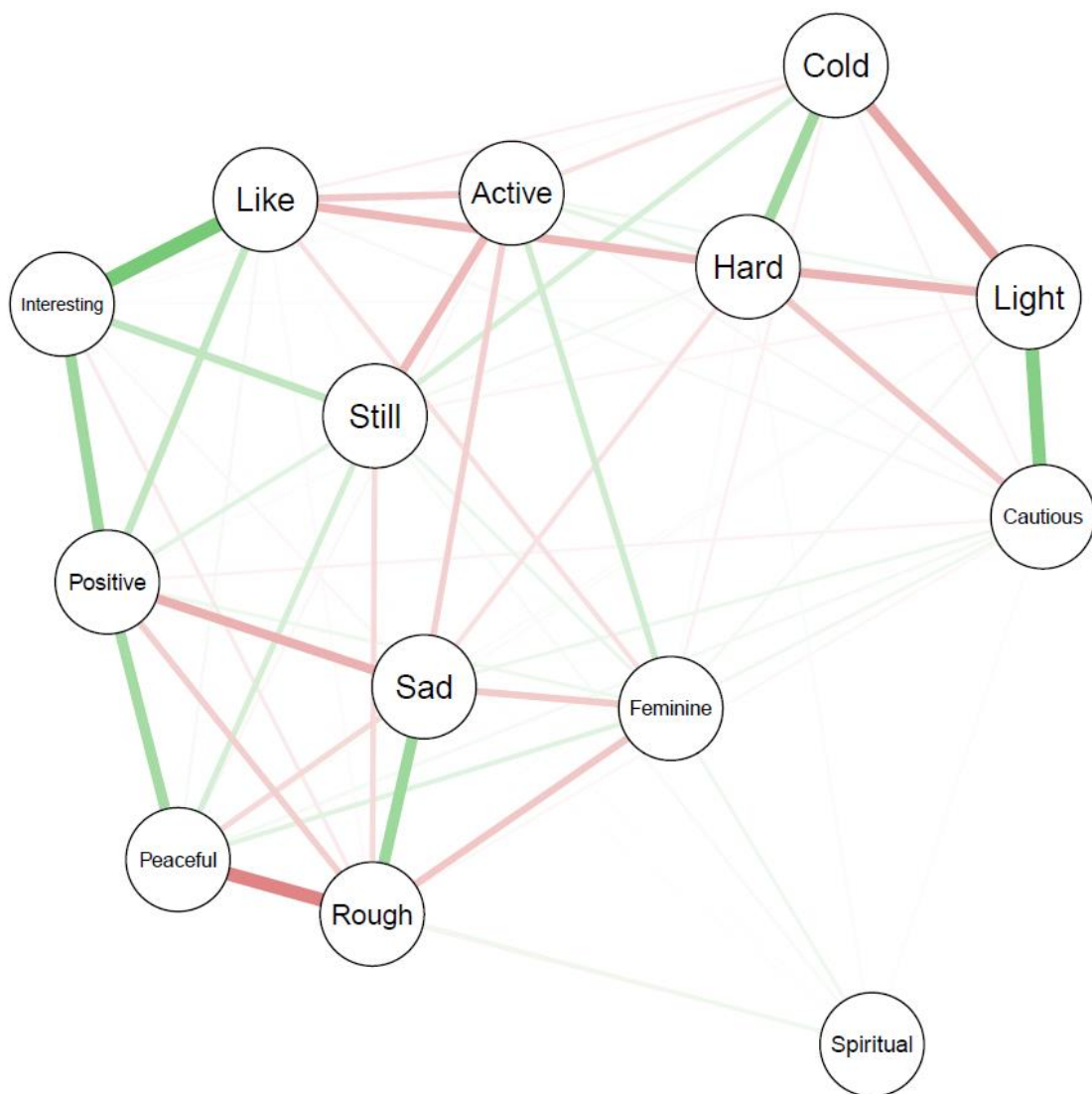


Figure 2. *Global network of aesthetic effects across all artworks. All node names represent the right side of the semantic differential. That is, the item negative-positive is represented by*

the node “positive”. A positive edge from e.g. the node active, with the node positive thus means the more active the more positive, a negative edge means the more active, the more negative. Thickness represents the strength.

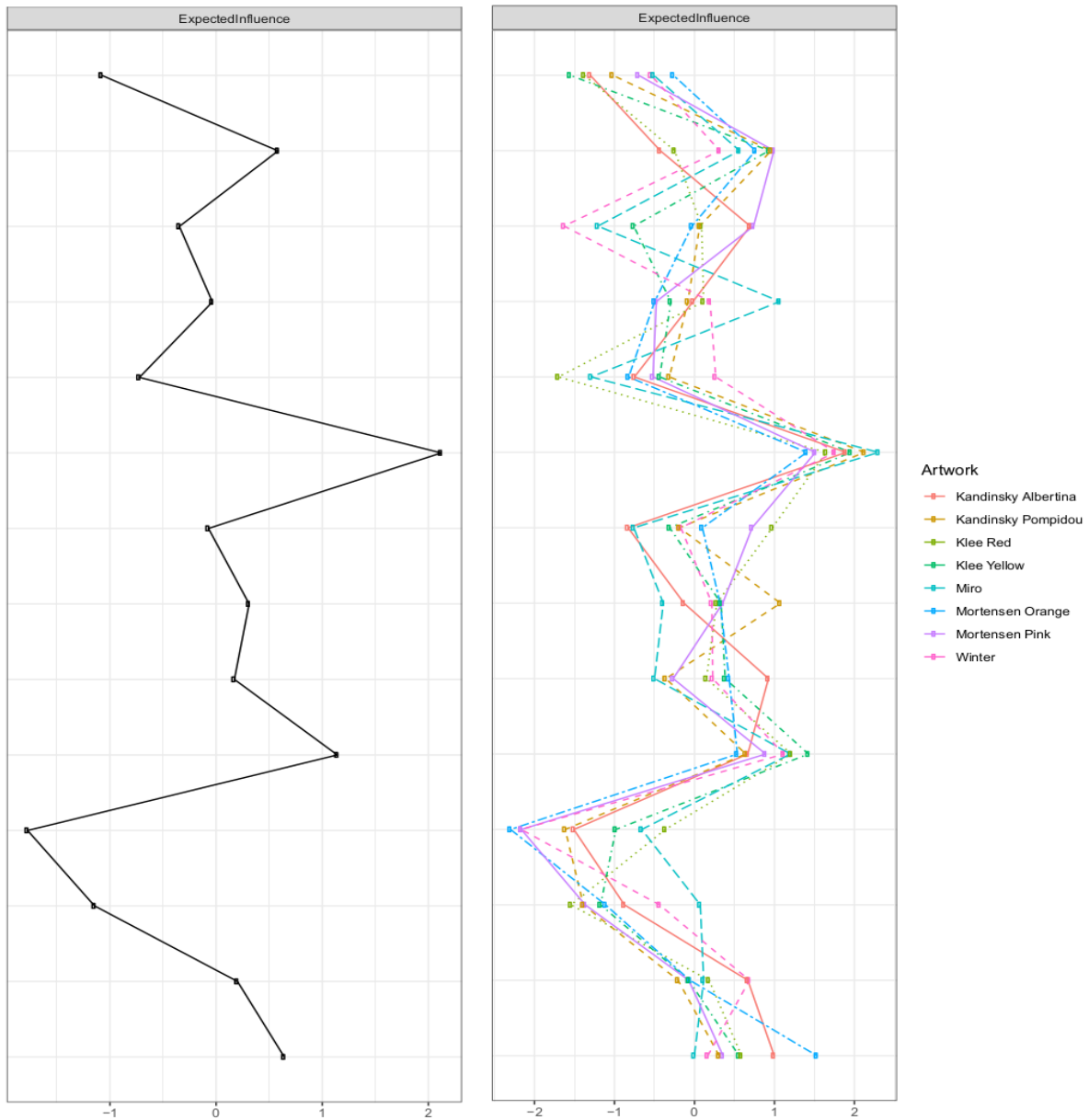


Figure 3. *Standardized centrality measures of expected influence for each aesthetic effect of the global network on the left and for each individual artwork network on the right.*

The general network of aesthetic effects represents a preliminary account of the general structure of the relationships between the different aesthetic effects. As Figure 2 shows, the estimated network was interconnected with strong positive edges between the

nodes “like” and “interesting” (0.27), “light” and “cautious” (0.24), “positive” and “interesting” (0.19), “sad” and “rough” (0.19), “positive” and “peaceful”, and “hard” and “cold” (0.18). The network has strong negative edges between the nodes “peaceful” and “rough” (-0.24), “cold” and “light” (-0.17), “hard” and “light” (-0.15), and “positive” and “sad” (-0.15).

With regard to centrality as measured by *expected influence*, “interesting” is the most central node in the network followed by “positive” and afterwards “cautious” and “still”. This is plotted in Figure 3. Notably, “rough” has a very low level of expected influence and “spiritual” though seemingly quite peripheral in Figure 2 falls somewhere in the middle.

An alternative approach to estimating a global network would be to average the raw data across the artworks (i.e. creating a mean score for each aesthetic effect) and estimate a network from these means. To estimate the robustness of our results, we also estimated the global network this way, the corresponding figures can be found on the Open Science Framework: <https://osf.io/zqxbm/> and in the supplementary materials. The similarity of the two global networks was $r=.83$ (we correlated the adjacency matrices); differences were mainly due to the second method leading to a sparser network due to regularization.¹ In addition, we calculated the correlation between the centrality of the two networks, this showed high similarity ($r=.94$).

Individual artwork networks

Since the network structure may vary dependent on artwork, we also estimated the aesthetic effects network per artwork, shown in Figure 4. The right panel of Figure 3 represents the corresponding centrality plot.

¹ To understand this: imagine that an edge weight is 0 in seven of the individual artwork networks and .4 in one of them. When you take an average over those edge weights, the edge weight will be above zero. In the first method regularization comes first and then an average is taken afterwards, in the second method this order is reversed, leading to a sparser network.

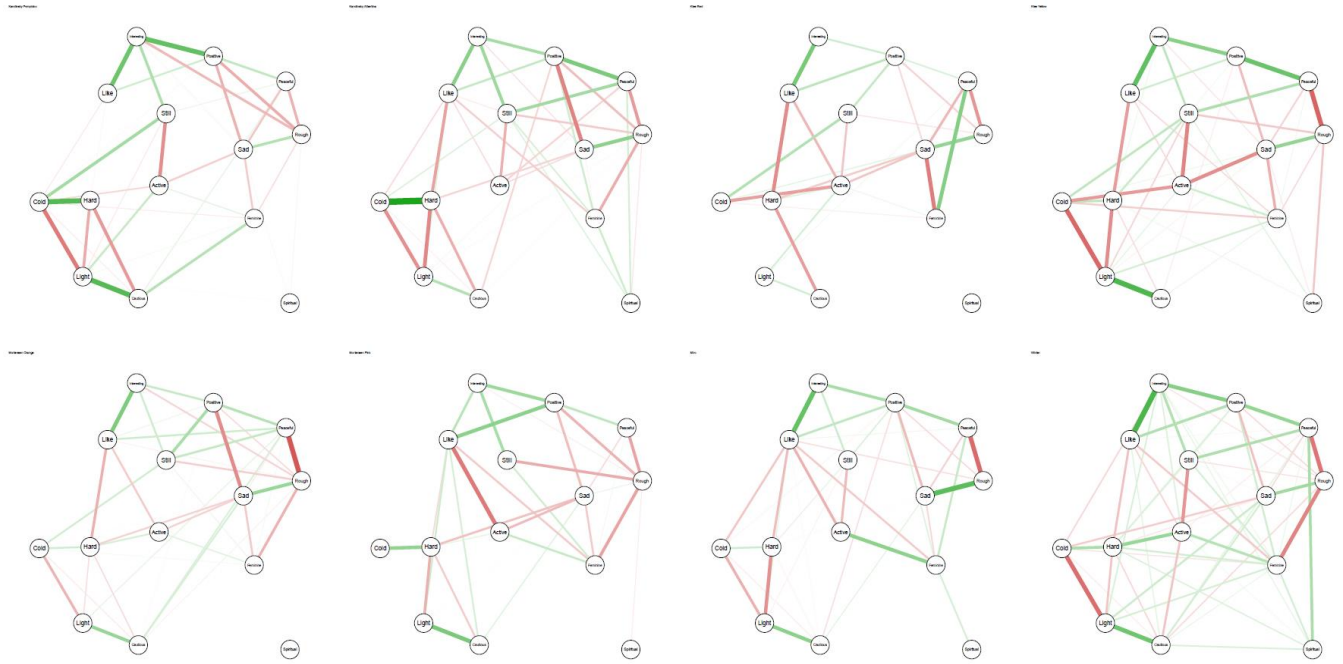


Figure 4. *Aesthetic effects networks for each individual artwork. As in Figure 1, all node names represent the right side of the semantic differential. That is, the item negative-positive is represented by the node “positive”. A positive edge from e.g. the node active, with the node positive thus means the more active the more positive, a negative edge means the more active, the more negative. Thickness represents the strength. All networks use the same graphical standardization, which means that the strength of the edges can be compared across networks.*

Figure 4 shows both considerable similarity and variability across the networks. For example, the positive association between “interesting” and “liking” is present in all networks, and in all networks this a relatively strong association, implying a consistent pattern. In addition, only the Winter network (bottom row, 4th network) involves an association between “active” and “cautious”, and the relation is weak: this implies the edge is largely absent from all networks. In contrast, both artworks by Paul Klee (top row, 3rd and 4th network) have a strong negative association between “cold” and “hard”, whereas the other networks have a strong positive association between these aesthetic effects, indicating greater variability. From the centrality plot, Figure 3, we can see that there is a variability in node

centrality depending on artwork. For example, “light” is relatively high in centrality for Mortensen Pink and Klee Red compared to the other artworks. There is also consistency across artworks, for example, “interesting” and “positive” have high centrality for all artworks.

As a next step to investigate cross-network variability in more detail, we derived an index of network similarity by correlating the adjacency matrices and centrality across artwork networks. Table 2 presents the corresponding correlation matrices. The lowest correlations ($r=.51$, $r=.57$, $r=.58$) of the adjacency matrix correlations would typically be considered moderate correlations, all others would typically be considered strong correlations. The centrality correlations show a similar pattern the lowest five correlations range from .42 to .56 and all others would again typically be seen as strong correlations.

Table 2

Correlation matrix of artwork networks

Artwork	Kandinsky Pompidou	Kandinsky Albertina	Klee Red	Klee Yellow	Mortensen Orange	Mortensen Pink	Miro	Winter
Kandinsky Pompidou		.71	.69	.87	.81	.89	.60	.72
Kandinsky Albertina	.80		.70	.74	.76	.67	.53	.61
Klee Red	.58	.51		.76	.63	.69	.56	.46
Klee Yellow	.83	.78	.69		.76	.75	.70	.78
Mortensen Orange	.71	.76	.70	.79		.87	.46	.69
Mortensen Pink	.68	.74	.57	.68	.72		.42	.62
Miro	.72	.75	.68	.82	.79	.70		.70
Winter	.75	.80	.55	.80	.80	.71	.79	

Note: Lower triangle are adjacency matrix correlations and upper triangle are centrality correlations

Variability network

As a final step to assess differences across networks we assessed the variability of edge weights. Figure 5 represent a network in which edges represent the *variability* in connection strength across the artworks rather than average connection strength. The network is scaled to a standard deviation of .15. In this network the greatest variability is found between the associations of “cold” and “hard” (0.14). Another relatively variable association

is between “cold” with “light” (0.12). Looking at Figure 4, we can see that these differences between the association with “cold” and “hard” may well be driven by the two artworks by Paul Klee. As noted, they include a negative association between these aesthetic effects while the other artworks include a positive association. Together these results indicate that there is some variation in the associations between the aesthetic effects dependent on artwork, however, aesthetic effects networks seem to be relatively similar in their structure.

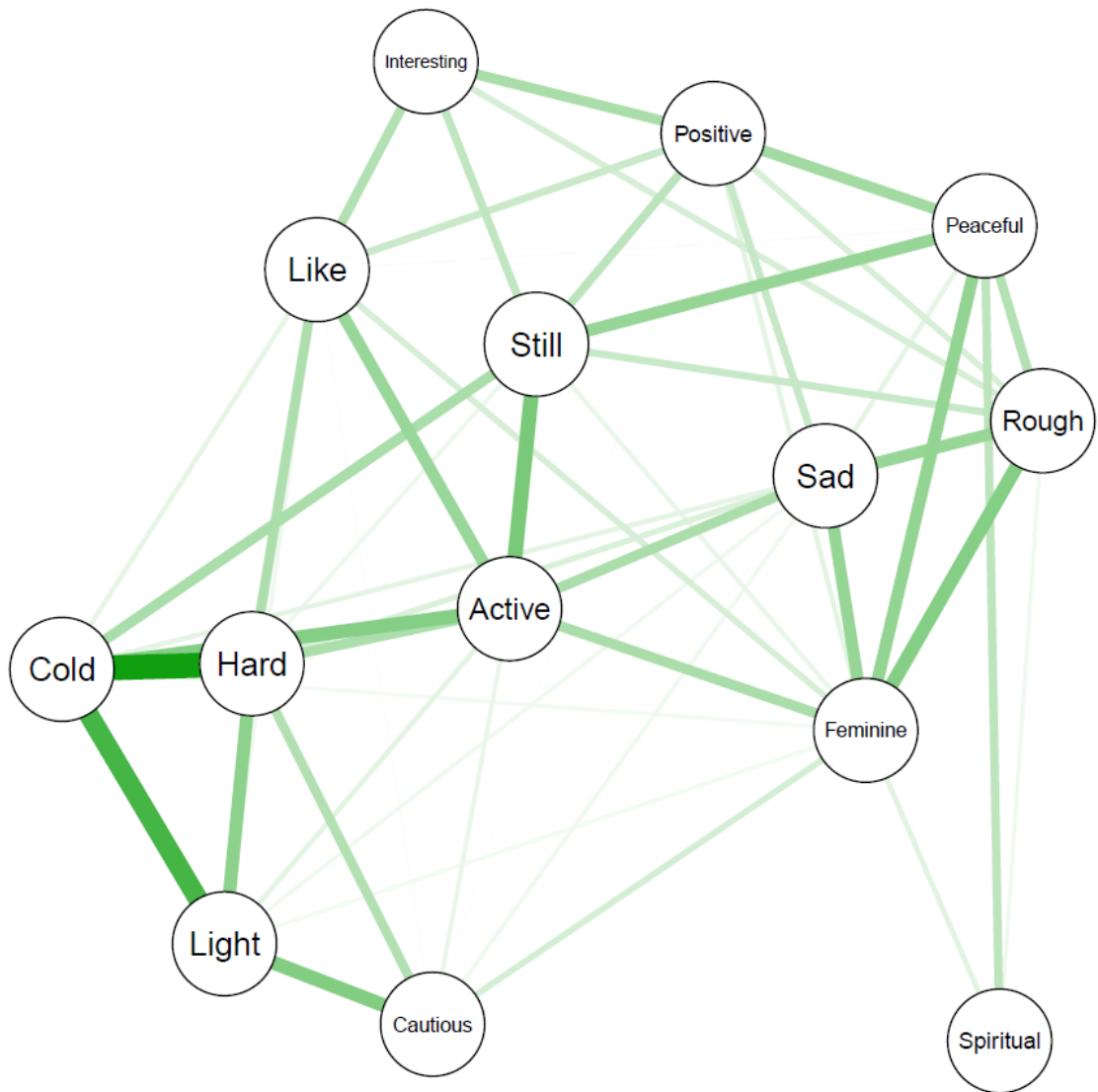


Figure 5. *Variability network. Each edge represents the standard deviation across the edge*

weights of each individual artwork network. Thicker edges imply a higher variability of each given edge.

Discussion

This data allowed us to formulate a preliminary account of the general structure of the relationships between the different aesthetic effects within the AEN. From this general structure, hypotheses can be derived on how the different aesthetic effects relate to each other. For example, participants in our study who rate an artwork as happy are likely to also see that artwork as positive, feminine, and smooth since these nodes are strongly connected. Though some of these associations may feel like “common sense”, recent empirical work has shown that people have a relatively low level of agreement on the aesthetic effects they attribute to artworks (Specker et al., *under review*). Speculatively, it could be that people share an association structure: people agree that artworks that are, for example, happy are also positive, feminine, and smooth but people disagree as to which artworks they think are happy.

Study 2

A limitation to Study 1 is that theoretical common cause models (like the EPA model) and network models can give rise to the same correlations among items (Fried & Cramer, 2017; Kruis & Maris, 2016; Van Der Maas et al., 2006). It is not possible to figure out which true model underlies the data based on estimating models from merely observing participants, but it is possible to distinguish between the models by experimental manipulation.

Therefore, we conducted a follow-up study to test the predictions from AEN model. The AEN not give an account of how the stimulus activates the network, it only postulates that the stimulus can activate the network. However, one can draw from the existing literature to generate hypotheses on how an artwork may activate the network. Research shows that brightness (of colors) is generally associated with positivity (Lakens, Fockenberg, Lemmens,

Ham, & Midden, 2013; Lakens, Semin, & Foroni, 2012; Meier, Robinson, & Clore, 2004; Meier, Fetterman, & Robinson, 2015; Specker et al., 2018; Specker & Leder, 2018). We therefore decided to manipulate the brightness of artworks as a means to activate the node of positivity. This manipulation is particularly suitable since there are indications that the association between brightness and positivity applies to the general brightness level of complex images (Lakens et al., 2013) as well as that it is universal (Specker et al., 2018), thus the manipulation should be applicable to artworks as well as relatively insensitive to individual differences. From our AEN model, this manipulation should lead the node of positivity to be activated, and from this node the activation can then spread throughout the network. Since positivity is the second most central aesthetic effect in terms of expected influence (Figure 3), this makes it a particularly suitable aesthetic effect to target via manipulation. Due to the high level of centrality, we would posit that activation should spread throughout the network.

In contrast, this should not happen if the data are generated by a common cause model, due to the causal direction the model proposes: Evaluation in the EPA model is the shared causal origin for both positive and peaceful. It is important to note here that in the conception of Osgood et al. (1957) the factors are causes and not summaries of the items. They note that there is no definite set of items to test for the EPA model, the specific items chosen depend on the specific subject matter, however, that “Standardization, and hence comparability, lies in the allocation of concepts to a common semantic space defined by a common set of general factors, despite variability in the particular concepts and scales employed”. (p.76) Thus, as Evaluation is the cause this means that if we can change the rating of positivity (by manipulating brightness), this should not lead to a change in Evaluation and thus not lead to a change in peaceful.

This is perhaps most easily understood when drawing a parallel with illnesses such as measles. If a patient has measles, this will lead to symptoms that correspond to this illness, such as fever and Koplik's spots. In this situation, symptoms are passive indicators of the underlying disease, similar to the EPA model. Treating the symptoms will alleviate suffering for the patient, but it will not change the fact that the patient has measles: a cure has to target the disease itself, not its indicators. The same applies to our case, and we predict that changing the rating of positivity (by manipulating brightness) should not lead to a change in Evaluation (the underlying cause) and thus should not influence how the artwork is rated on e.g., peacefulness or smoothness.

This is a general hypothesis that assumes that if positivity changes this can have an effect on the rest of the network as a whole. This would be strong evidence in favour for the AEN model. This is because common cause models exclude direct influences of positivity on *any* of the other nodes. Thus, any effect of positivity on any of the other nodes can provide strong evidence for a network approach over a common cause approach. Nonetheless, based on the general network structure identified in Study 1 (see also Figure 2) we also formulated specific hypotheses as to which nodes were likely to be (most) affected by our manipulation based on the strength of their connection to the node of positivity. Notably, this should lead to the artwork being seen as peaceful, smooth, and happy. These are the direct effects, but we also assume that there can be indirect effects that are caused not by direct connections with positivity but are a result of a longer causal chain. Thus, we predicted that still, peaceful, liking, and interest to go up (lighter images are more still) and that sad, rough, and cautious would go down (lighter images are less sad). We pre-registered our study on the Open Science Framework: <https://osf.io/7upyh/register/5771ca429ad5a1020de2872e>.

To ensure that our manipulation would be effective, we first conducted a pilot-study to test if the brightness–positivity association found with colors would translate to more complex stimuli specifically artworks.

Pilot-study

We adapted the methods of Specker et al. (2018) and Specker and Leder (2018), and used an Implicit Association Test (IAT) and an explicit rating task that were counterbalanced (half of the participants first completed the IAT and then gave explicit ratings, and the other half first gave explicit ratings and then completed the IAT). Participants saw dark and bright versions of the artworks used in Study 1. These versions were created by use of GIMP software making them either 30% darker or 30% brighter. We had 27 participants²—mean age = 22.93 (SD = 6.94), 18 female—all students of the University of Vienna who received course credit for their participation. All participants participated voluntarily. Results indicated a strong implicit association between brightness and positivity, $t(26) = 13.08$, $p < .001$, $d = 2.52[1.73, 3.29]$, as well as a strong explicit association between brightness and positivity: $t(26) = 4.994$, $p < .001$, $d = .96[.49, 1.41]$. Indicating that the effects should be generalizable to paintings and that our manipulation should be effective. The dataset and analysis syntax can be found on the Open Science Framework: <https://osf.io/3n8zy/>.³

Method

Our method for Study 2 was similar to Study 1. Participants saw eight artworks and rated each artwork on 14 scales, however, in this study there were three versions (control, dark, and bright) of each artwork. Dark and bright versions were the same as in the pilot-study. Control versions represent “original” versions of the artworks meaning the same

² Our initial sample consisted of 28 participants; one participant was excluded because she did not speak German.

³ Due to copyright the images are not available on the OSF though they can be seen by running the OSF script from Study 2 and will, of course, be shared upon request.

version as in Study 1. Thus “control version” refers to the fact that these images were not manipulated. Since these images are all reproductions of artworks and thus not strictly “originals” we use control version throughout the text to refer to these images. From the association between brightness and positivity it should be the case that: bright version > control version > dark version for ratings of positivity which would constitute a manipulation check. If this holds, one can then analyze whether the manipulation had the expected effect.

To make the variation in brightness salient we wanted to ensure that each participant saw control versions, dark versions and bright versions of artworks. At the same time, to cancel out effects of repeated exposure, we wanted each participant to see each artwork in only one version. We created three groups which differed in which version of the artwork they saw (see Table 3). These groups were not used as a basis for analysis. The experimental script, final dataset and analysis scripts are made available on the Open Science Framework:

<https://osf.io/3n8zy/>.

Table 3

Presentation of Images

Shorthand Title	<u>Version</u>		
	Group 1	Group 2	Group 3
Klee Yellow	Dark	Control	Bright
Klee Red	Bright	Dark	Control
Kandinsky Albertina	Dark	Control	Bright
Kandinsky Pompidou	Bright	Dark	Control
Mortensen Pink	Bright	Dark	Control
Mortensen Orange	Dark	Control	Bright
Miro	Control	Bright	Dark
Winter	Control	Bright	Dark

Participants

We tested 133 participants (22 male, 111 female, none identifying as other; mean age = 20.41, SD = 2.60). All participants participated voluntarily and were students of the University of Vienna who received course credit for their participation.

Procedure

As in Study 1 participants saw 8 artworks which they rated on 14 scales. Contrary to Study 1, participants participated in the lab, the study was self-timed, and presentation order of the scales was randomized. Upon arrival in the lab participants signed informed consent and then completed the rating task.

Results and Discussion

To analyze our results we first did a manipulation check to see if our manipulation was effective. To test this we ran a one-way ANOVA (bright, dark, control) as an within-subjects independent variable and positivity as a dependent variable. If our manipulation was effective it should be that the brighter an image, the more positive it is rated (so: bright version > control version > dark version). This is exactly what we found, $F(2,264) = 3.77$, $p = .024$, $\eta G^2 = .018$. A plot of this can be found on the Open Science Framework: <https://osf.io/3n8zy/>. It has to be noted that the effect was smaller than anticipated based on the pilot-study ($d = .96$). This is most likely due to the fact that in the pilot-study participants saw two versions of the same images thus leading to a more direct comparison based on the brightness feature. Whereas in this study participants only saw one version of each image.

We interpreted this as our manipulation being significant and then ran a MANOVA with version (bright, dark, control) as an within-subjects independent variable and all aesthetics effects (within the exception of positive) as dependent variables to test whether the manipulation changed how the aesthetic effects were rated. If the MANOVA is significant

this would provide strong evidence for a network approach over a common cause approach which is exactly what we found: $F(2, 264) = 3.55, p < .001$.

Furthermore, based on the general network structure identified in Study 1 (see also Figure 2) we formulated specific hypotheses as to which nodes were likely to be (most) affected by our manipulation based on the strength of their connection to the node of positivity. To test our specific hypotheses we ran a one-way ANOVAs (bright, dark, control) as an within-subjects independent variable for each dependent variable separately. All ANOVA results and corresponding means and standard deviations are reported in Table 4. All corresponding figures can be found on the Open Science Framework: <https://osf.io/3n8zy/>.

Table 4

Follow-up ANOVA and descriptive statistics per aesthetic effect

Aesthetic Effect	DFn	DFd	F	p	$G\eta^2$	Mean (SD)		
						Bright	Control	Dark
Aggressive – Peaceful	2	264	1.35	.26	.007	4.42 (1.07)	4.31 (1.09)	4.21 (1.09)
Intrusive-Cautious	2	264	.86	.42	.005	3.87 (1.18)	3.73 (1.06)	3.69 (1.06)
Happy-Sad	2	264	21.12	<.001	.090	3.49 (.9)	3.68 (.87)	4.17 (.96)
Dislike-Like	2	264	.54	.59	.003	3.91 (1.26)	4.02 (1.21)	4.06 (1.31)
Bodily-Spiritual	2	264	1.16	.32	.006	4.3 (1.35)	4.07 (1.24)	4.16 (1.1)
Lively-Still	2	264	2.68	.07	.013	3.69 (.99)	3.5 (.98)	3.77 (1.03)
Masculine-Feminine	2	264	8.08	<.001	.044	4.35 (.98)	4.19 (1.11)	3.83 (.98)
Passive – Active	2	264	.76	.47	.004	4.44 (.94)	4.45 (1.01)	4.32 (1.14)
Smooth-Rough	2	264	2.79	.06	.014	3.81 (1.14)	4.09 (1.12)	4.09 (1.04)
Heavy-Light	2	264	13.93	<.001	.067	4.47 (1.05)	4.15 (1.01)	3.81 (1)
Uninteresting-Interesting	2	264	.52	.60	.002	4.34 (1.16)	4.48 (1.13)	4.43 (1.26)
Warm-Cold	2	264	.01	.99	<.001	3.85 (1.19)	3.85 (1.19)	3.87 (1.17)
Soft-Hard*	2	264	.95	.50	<.001	3.8 (1.05)	3.93 (.94)	3.94 (1.11)

*In this case, the assumption of sphericity was violated so the value in the F column represents a Greenhouse-Geisser epsilon and all other values represent corrected values.

As predicted we found that darker images were significantly perceived as sadder, however, we did not find significant effects for liking, interest, still, and peaceful.

We found an unexpected significant effect on heavy-light, it would be possible that this effect was an indirect effect of positivity it is more likely that our manipulation influenced both the positivity node as well as the "weight" node. The second explanation is more likely due to existing literature on cross-modal correspondences showing an association between lightness and weight (Alexander & Shansky, 1976; De Camp, 1917; Walker, Francis, & Walker, 2010; Wright, 1962). Since there was no direct relationship between heavy-light and negative-positive it seems unlikely that this effect on cautious was due to a longer chain of causality.

However, with changes in heavy-light one would assume that warm-cold and soft-hard would also change due to their strong connections: the lighter the warmer and the lighter the softer. This seems not to be the case, however, we know from the variability analysis of Study 1 that the association between warm-cold with heavy-light is one of the more variable connections in our network. In addition, the association between warm-cold and soft-hard is the other more variable connection in our network. Thus, it is potentially this variability that may be partially responsible.

The unexpected result for liking and interest most likely has a different cause. One possible explanation centers around the directionality of edges in the network. All other nodes in the network represent associations which makes it more likely that these relationships are bi-directional. It is relatively unlikely that someone who associates dark colors with negativity does not associate bright colors with positivity (though this claim is sometimes made, see e.g. Lakens et al., 2012). What could tentatively explain these results is that the edges between liking and interest with positivity represent uni-directional relationships. People rate artworks as more positive if they like them and find them interesting, but people do not like artworks

more and find things more interesting if they rate them as more positive. The underlying mechanism would be that people would misattribute their own felt valence to the valence of the stimulus (e.g. seeing a sad song as more happy because I feel happy listening to it). Nonetheless, this would be a strong theoretical claim that would need further and more detailed experimental investigation.

Another possible explanation centers around the fact that many things factor into making a judgment about if you like a stimulus or not. The same can be said for whether you think a stimulus is positive or not. And, crucially, some of these may overlap. It may be, for example, that a balanced composition makes people like an artwork more and makes them think that it is more positive. This would result in a correlation between positivity and liking that would be insensitive to a manipulation of brightness (since the balance is the same in both the bright and the dark images). In our case, when just computing a correlation between liking and interest with positivity you find that they are correlated: $r = .46$ for liking and $.32$ for interest.

In sum, even though the current results bring up new questions for the AEN one thing is clear from these results, namely: they are clearly not in accordance with a common cause perspective. From a common cause perspective it should not be the case that if ratings of positivity goes up (as achieved by our manipulation) that ratings of the other aesthetic effects change. Our results are evidence in favor of the network model over a common cause perspective. Nevertheless, since we only found a significant effect on happy-sad, we did not find strong evidence in favour of the network structure formulated by us in Study 1.

General Discussion

We presented a new perspective that changes the way we, at least formally, conceptualize aesthetic effects and to a larger degree aesthetic experience. In addition, the

results of Study 2 show empirical support in favor of the AEN model over other potential alternatives. But did not show strong evidence for the specific network structure outlined in Study 1 (see Figure 2).

This new perspective introduces a more process-oriented view. As noted, most theoretical models actually do conceptualize aesthetic experience as a process (e.g., Leder, Belke, Oeberst, & Augustin, 2004; Locher, Overbeeke, & Wensveen, 2010; Pelowski, Markey, Forster, Gerger, & Leder, 2017; Tinio, 2013). Nevertheless, experimental work tends to focus on the outcomes of aesthetic experience (e.g., Gerger et al., 2017; Graf & Landwehr, 2015; Silvia, 2009; Specker et al., 2017). Empirical studies like the approaches of Osgood et al. (1957) and Wellek (1929) focus on specific factors such as features of the artwork that lead to outcomes, basically stating $A \rightarrow B$. This leaves open the question as to the underlying process: how does A lead to B? The current approach shifts the focus to the process.

Refocusing on understanding the processes underlying aesthetic experience can lead to valuable insights into how people experience art. This new approach extends beyond research on aesthetic effects and can be applied to aesthetic experience in general. When taking this wider perspective, future work could model other variables of interest such as relevant individual differences (e.g. art interest or art knowledge) or emotions. Within a network perspective, these variables can influence each other representing a process-based approach to aesthetic experience in which aesthetic experience represents an active interaction between a viewer and an artwork.

Limitations and Future directions

As noted in the introduction, the way a stimulus activates or influences the network may vary across individuals, as illustrated by the example that a painting may cause the association “active” in Sally and not in Harry. This is illustrated in Figure 6.

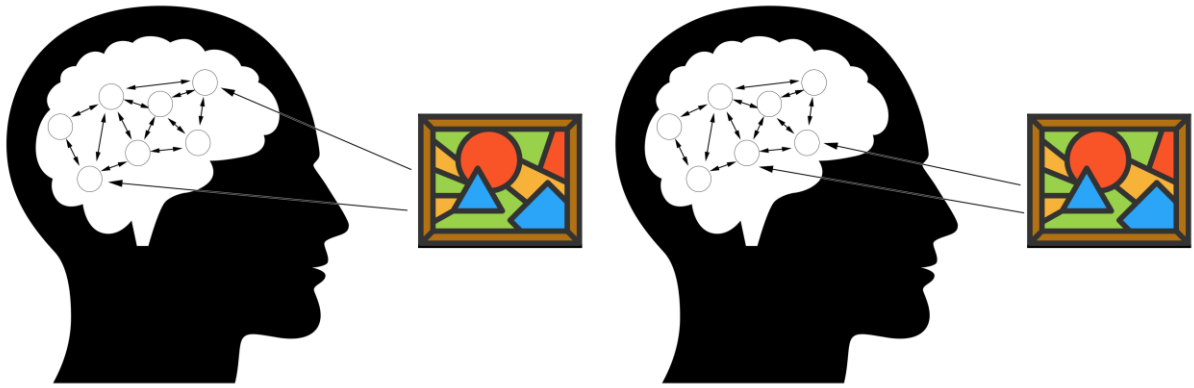


Figure 6. *Conceptual model of how a painting influences two people differently. The same painting influences the person on the left differently from the person on the right by activating different nodes in their networks.*

This is not necessarily incompatible with the notion of a general network structure. If we call the left head in Figure 6 “Harry” and the right “Sally”, we can see that in this example Harry and Sally have the same network structure; what makes them different is how the artwork influences their network. The same artwork activates different nodes in Harry than in Sally. Note that in this example the artwork activates two nodes in both Harry and Sally, though in a real-world case it might be that the amount of nodes an artwork activates varies across people. To investigate how people differ, mixture network modelling would be particularly beneficial, unfortunately, this is not available at the current date (see Fried & Cramer, 2017; for a comprehensive outline of some of the methodological challenges in the field).

Mixture network modeling holds great potential for understanding individual differences in aesthetic experience since these differences could be explicitly connected to differences in network structure. If we want to understand how people experience art, then questions pertaining to why different people have different experiences when looking at the same artworks as well as to investigate why people have similar experiences when looking at

different artworks are central. In fact, current trends within empirical aesthetics have seen an interest in using mixed linear models specifically to account for differences in the person by stimulus interaction (Brieber, Forster, & Leder, 2018; Gartus & Leder, 2014; Laming et al., 2016; Leder, Goller, Rigotti, & Forster, 2016; Silvia, 2013).

Due to, amongst other issues, the large sample sizes required for mixture *network* modeling it is unlikely that this technique will become available in the foreseeable future. That said, what is possible at the moment would be to look at differences on a group level. To give some examples: based on the vast literature showing differences in art experience based on art expertise (Chamberlain & Wagemans, 2015; Chirumbolo, Brizi, Mastandrea, Mannetti, & Martens, 2014; Leder et al., 2014; Winston & Cupchick, 1992), one could for instance investigate whether art experts (as a group) have a more strongly connected network of aesthetic effects than lay people (as a group). Based on findings related to art styles (Augustin & Leder, 2006; Leder et al., 2012), one could also compare a network for abstract art with one for figurative art. Though these approaches are not as rich as mixture network modeling would be, they could still provide new and interesting information that cannot be gained from methods used so far. A richer and potentially very fruitful approach would be to use latent class analysis to see if one could derive latent classes within the data, this would still require large sample sizes but is more feasible than mixture network modeling at the moment. These classes can consist of either artworks or people. Essentially, what this would entail is that within a given subgroup, people (or artworks) would share a network structure, whereas across groups the network structure differs. This is already a highly relevant point in non-stimulus bound cross-sectional data where people commonly average over people. In our stimulus-bound case this becomes even more relevant because we average not only over people but also average over stimuli. The possibility to test for subgroups either of people or of stimuli (and ideally both) would be extremely valuable.

This would also address limitations of our study in terms of generalizability. The generalizability of our findings is limited by our homogenous set of stimuli and our homogenous set of participants. The homogeneity of our stimuli makes it unclear to what extent our results are generalizable to the whole range of abstract art, as well as to the even wider range of artworks in general. The homogeneity of our participants makes it unclear to what extent our results are generalizable to populations that do not consist of so-called “WEIRD” (Western Educated Industrialized Rich Democratic; Henrich, Heine, & Norenzayan, 2010) people.

Our use of a relatively homogenous set of stimuli was beneficial to our aim of estimating a general structure of aesthetic effects that would generalize across artworks. We found evidence for this generalizability by finding high correlations of edge weights between the artwork networks (Table 2). To put this into perspective, the highest correlation reported by Rhemtulla et al. (2016) was .64, with most other correlations being in the .20-.50 range. We found additional evidence for generalizability by finding low variance across artworks in our variability network (Figure 5). However, this network similarity may, in part, be driven by our homogenous set of stimuli and participants. The homogeneity of our stimuli likely limits the variance between different artwork networks since the different artworks are relatively similar. The homogeneity of our sample further limits the variance between different artwork networks since all individual artwork networks were derived from the same sample. Nonetheless, if the homogeneity of our sample limited the variance in our data, this would mean that people are relatively consistent in their associations between aesthetic effects (network structure) across artworks.

To conclude, we have proposed a new way to conceptualize aesthetic effects by using a network approach. In addition, we have shown evidence for this approach. Conceptualizing aesthetic effects as an associative process provides an explicit account of a specific cognitive

process involved in aesthetic experience. In addition, it shifts theoretical and empirical focus back to a process-based understanding of aesthetic experience. Refocusing on understanding the processes underlying aesthetic experience can lead to valuable insights into how people experience art.

References

- Alexander, K. R., & Shansky, M. S. (1976). Influence of hue, value, and chroma on the perceived heaviness of colors. *Perception & Psychophysics*, 19(1), 72–74.
<https://doi.org/10.3758/BF03199388>
- Augustin, D., & Leder, H. (2006). Art expertise: a study of concepts and conceptual spaces. *Psychology Science*, 48(2), 135–156.
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13.
<https://doi.org/10.1002/wps.20375>
- Brieber, D., Forster, M., & Leder, H. (2018). On the Mutual Relation Between Art Experience and Viewing Time. *Psychology of Aesthetics, Creativity, and the Arts*.
<https://doi.org/10.1037/aca0000210>
- Brinkmann, H., Boddy, J., Immelmann, B., Specker, E., Pelowski, M., Leder, H., & Rosenberg, R. (2018). Ferocious Colors and Peaceful Lines Describing and Measuring Aesthetic Effects. *Wiener Jahrbuch Für Kunstgeschichte*, 65.
- Chamberlain, R., & Wagemans, J. (2015). Visual arts training is linked to flexible attention to local and global levels of visual stimuli. *Acta Psychologica*, 161, 185–197.
<https://doi.org/10.1016/j.actpsy.2015.08.012>
- Chirumbolo, A., Brizi, A., Mastandrea, S., Mannetti, L., & Martens, A. (2014). ‘Beauty Is No Quality in Things Themselves’: Epistemic Motivation Affects Implicit Preferences for Art. *PLoS ONE*, 9(10), e110323. <https://doi.org/10.1371/journal.pone.0110323>
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428. <https://doi.org/10.1037/0033-295X.82.6.407>
- Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas,

- H. L. J. (2016). Toward a Formalized Account of Attitudes: The Causal Attitude Network (CAN) Model. *Psychological Review*, 123(1), 2–22.
<https://doi.org/10.1037/a0039802>
- De Camp, J. E. (1917). The influence of color on apparent weight. A preliminary study. *Journal of Experimental Psychology*, 2(5), 347. <https://doi.org/10.1037/h0075903>
- De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, 40(1), 213–231. <https://doi.org/10.3758/BRM.40.1.213>
- Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50(1), 195–212.
<https://doi.org/10.3758/s13428-017-0862-1>
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, 48(4), 1–18. Retrieved from <http://www.jstatsoft.org/>
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*. <https://doi.org/10.1037/met0000167>
- Fried, E. I., & Cramer, A. O. J. (2017). Moving Forward: Challenges and Directions for Psychopathological Network Theory and Methodology. *Perspectives on Psychological Science*, 12(6), 999–1020. <https://doi.org/10.1177/1745691617705892>
- Fried, E. I., Eidhof, M. B., Palic, S., Costantini, G., Huisman-van Dijk, H. M., Bockting, C. L. H., ... Karstoft, K.-I. (2018). Replicability and Generalizability of Posttraumatic Stress Disorder (PTSD) Networks: A Cross-Cultural Multisite Study of PTSD Symptoms in Four Trauma Patient Samples. *Clinical Psychological Science*, 6(3), 335–351.
<https://doi.org/10.1177/2167702617745092>

- Gartus, A., & Leder, H. (2014). The white cube of the museum versus the gray cube of the street: The role of context in aesthetic evaluations. *Psychology of Aesthetics, Creativity, and the Arts*, 8(3), 311–320. <https://doi.org/10.1037/a0036847>
- Gerger, G., Pelowski, M., & Leder, H. (2017). Empathy, Einfühlung, and aesthetic experience: the effect of emotion contagion on appreciation of representational and abstract art using fEMG and SCR. *Cognitive Processing*. <https://doi.org/10.1007/s10339-017-0800-2>
- Graf, L. K. M., & Landwehr, J. R. (2015). A Dual-Process Perspective on Fluency-Based Aesthetics. *Personality and Social Psychology Review*, 19(4), 395–410. <https://doi.org/10.1177/1088868315574978>
- Hekkert, P., & Van Wieringen, P. C. W. (1996). Beauty in the Eye of Expert and Nonexpert Beholders: A Study in the Appraisal of Art. *The American Journal of Psychology*, 109(3), 389. <https://doi.org/10.2307/1423013>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Kandinsky, W. (1947). *Point and Line to Plane*. New York, NY: The Solomon R. Guggenheim Foundation.
- Kruis, J., & Maris, G. (2016). Three representations of the Ising model. *Scientific Reports*, 6(1), 34175. <https://doi.org/10.1038/srep34175>
- Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J., & Midden, C. J. H. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion*, 27(7), 1225–1246. <https://doi.org/10.1080/02699931.2013.781501>

- Lakens, D., Semin, G. R., & Foroni, F. (2012). But for the bad, there would not be good: Grounding valence in brightness through shared relational structures. *Journal of Experimental Psychology: General*, 141(3), 584–594. <https://doi.org/10.1037/a0026468>
- Lauring, J. O., Pelowski, M., Forster, M., Gondan, M., Ptito, M., & Kupers, R. (2016). Well, if they like it... Effects of social groups' ratings and price information on the appreciation of art. *Psychology of Aesthetics, Creativity, and the Arts*, 10(3), 344–359. <https://doi.org/10.1037/aca0000063>
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95(4), 489–508. <https://doi.org/10.1348/0007126042369811>
- Leder, H., Gerger, G., Brieber, D., & Schwarz, N. (2014). What makes an art expert? Emotion and evaluation in art appreciation. *Cognition and Emotion*, 28(6), 1137–1147. <https://doi.org/10.1080/02699931.2013.870132>
- Leder, H., Gerger, G., Dressler, S. G., & Schabmann, A. (2012). How art is appreciated. *Psychology of Aesthetics, Creativity, and the Arts*, 6(1), 2–10. <https://doi.org/10.1037/a0026396>
- Leder, H., Goller, J., Rigotti, T., & Forster, M. (2016). Private and Shared Taste in Art and Face Appreciation. *Frontiers in Human Neuroscience*, 10, 155. <https://doi.org/10.3389/fnhum.2016.00155>
- Locher, P., Overbeeke, K., & Wensveen, S. (2010). Aesthetic Interaction: A Framework. *Design Issues*, 26(2), 70–79. https://doi.org/10.1162/DESI_a_00017
- Macke, F. (1964). *August Macke, Franz Marc; Briefwechsel*. Köln: M. DuMont Schauberg.
- Meier, B. P., Fetterman, A. K., & Robinson, M. D. (2015). Black and White as Valence Cues.

Social Psychology, 46(3), 174–178. <https://doi.org/10.1027/1864-9335/a000236>

- Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why Good Guys Wear White: Automatic Inferences About Stimulus Valence Based on Brightness. *Psychological Science*, 15(2), 82–87. <https://doi.org/10.1111/j.0963-7214.2004.01502002.x>
- Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, 105(2), 299–324. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9577240>
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Urbana, Chicago, and London: University of Illinois Press.
- 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. <https://doi.org/10.1016/J.PLREV.2017.02.003>
- Rhemtulla, M., Fried, E. I., Aggen, S. H., Tuerlinckx, F., Kendler, K. S., & Borsboom, D. (2016). Network analysis of substance abuse and dependence symptoms. *Drug and Alcohol Dependence*, 161, 230–7. <https://doi.org/10.1016/j.drugalcdep.2016.02.005>
- Schnackertz, H. J. (2010). Wirkung/Rezeption. In K. Barck, M. Fontius, F. Wolfzettel, & B. Steinwachs (Eds.), *Ästhetische Grundbegriffe*. Stuttgart/Weimar: Metzler.
- Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, 3(1), 48–51. <https://doi.org/10.1037/a0014632>
- Silvia, P. J. (2013). Interested Experts, Confused Novices: Art Expertise and the Knowledge

Emotions. *Empirical Studies of the Arts*, 31(1), 107–115.

<https://doi.org/10.2190/EM.31.1.f>

Specker, E., & Leder, H. (2018). Looking on the Bright Side: Replicating the Association between Brightness and Positivity. *Collabra: Psychology*, 4(1), 34.

<https://doi.org/10.1525/collabra.168>

Specker, E., Leder, H., Rosenberg, R., Hegelmaier, L. M., Brinkmann, H., Mikuni, J., & Kawabata, H. (2018). The universal and automatic association between brightness and positivity. *Acta Psychologica*, 186, 47–53. <https://doi.org/10.1016/j.actpsy.2018.04.007>

Specker, E., Tinio, P. P. L., & van Elk, M. (2017). Do you see what I see? An investigation of the aesthetic experience in the laboratory and museum. *Psychology of Aesthetics, Creativity, and the Arts*, 11(3), 265–275. <https://doi.org/10.1037/aca0000107>

Steyvers, M., & Tenenbaum, J. B. (2005). *The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth*.

https://doi.org/10.1207/s15516709cog2901_3

Takahashi, S. (1995). Aesthetic properties of pictorial perception. *Psychological Review*, 102(4), 671–683. <https://doi.org/10.1037//0033-295X.102.4.671>

Tinio, P. P. L. (2013). From artistic creation to aesthetic reception: The mirror model of art. *Psychology of Aesthetics, Creativity, and the Arts*, 7(3), 265–275.

<https://doi.org/10.1037/a0030872>

Travers, J., & Milgram, S. (1969). An Experimental Study of the Small World Problem. *Sociometry*, 32(4), 425–443. Retrieved from <http://www.jstor.org/stable/2786545>

Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive

manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861.

<https://doi.org/10.1037/0033-295X.113.4.842>

Walker, P., Francis, B. J., & Walker, L. (2010). The Brightness-Weight Illusion. *Experimental Psychology*, 57(6), 462–469. <https://doi.org/10.1027/1618-3169/a000057>

Wellek, A. (1929). Das Doppelempfinden in der Geistesgeschichte. *Zeitschrift Für Ästhetik Und Allgemeine Kunstwissenschaft*, 23(1), 14–42.

Winston, A. S., & Cupchick, G. C. (1992). The Evaluation of High Art and Popular Art By Naive and Experienced Viewers. *Visual Arts Research*, 18(1), 1–14.

Wright, B. (1962). The Influence of Hue, Lightness, and Saturation on Apparent Warmth and Weight. *The American Journal of Psychology*, 75(2), 232.

<https://doi.org/10.2307/1419606>

Art Historical

2.2.3

Brinkmann, H., Boddy, J., Immelmann, B., **Specker, E.**, Pelowski, M., Leder, H., & Rosenberg, R. (2018). Ferocious Colors and Peaceful Lines Describing and Measuring Aesthetic Effects. *Wiener Jahrbuch Für Kunstgeschichte*, 65, 2-26.

Summary:

This paper aims to give an overview of the literature on aesthetic effects. Specifically it shows that the interest in aesthetic effects set the stage for a sustained exchange between art literature and psychological research. It shows how interest in the effect of colors and lines emerge in the 18th century in aesthetic theory by way of e.g., Burke and Hogarth as well as their artistic counterparts such as William Turner and Wassily Kandinsky. It then shows how description of aesthetic effects uses an underlying principle of polarities that can be traced back to Goethe and summarizes the most frequently used categories for description (i.e. temperature, sound, feelings, gender). From this it introduces early empirical work on aesthetic effects, focusing mainly on Kandinsky's famous questionnaire that focused on color-shape correspondences. But also includes early psychological studies such as the work by Lundholm (1921) that studied the "emotional quality" that people ascribed to different types of lines. Notably, it can be derived from this discussion that though initial studies in the period between roughly 1910-1935 were conducted empirical interest in the topic faded with the exception of one study by Takahashi (1995) which found some continuation in the work by Stamatopoulou (2007; 2008). From this the paper reviews the work by Osgood and colleagues (notably Osgood et al. 1957) that though it was focused on the measurement of meaning or semantic space could be applied to art and has interesting parallels with aesthetic effects. The semantic differential that they developed is a type of bipolar scale which is in line with the noted polarity in description of aesthetic effects. Furthermore, Osgood's work mainly deals with associations, as noted, from a psychological perspective, the terminology of an aesthetic effect can be translated into what we call conceptual associations between separate concepts (i.e. the concept "red" and the concept "warmth"). It is therefore not surprising that his work found itself reflected in art historical discussions, for example by Ernst Gombrich but also in a series of interdisciplinary aesthetic congresses paired by the journal "Exakte Ästhetik" both managed by psychologist William Simmat. The final section of the manuscript constitutes an empirical study to see how a bipolar scale (semantic differential) relates to unipolar scales or a selection method in order to find the optimal way to measure these effects in contemporary viewers.

My Contribution:

In this case the concept and idea underlying the manuscript was developed by HB, JB, BI, and RR. My contribution lies in section 5 of the manuscript "An experimental study and outlook on future research". Here I was responsible for the design, data collection, data analysis, and interpretation of the results of the experimental study. As well as writing this section of the article. In addition, I provided feedback on consecutive versions of the article by my co-authors.

WIENER JAHRBUCH FÜR KUNSTGESCHICHTE

Herausgegeben vom
Bundesdenkmalamt Wien
und vom Institut für Kunstgeschichte
der Universität Wien

BAND LXV

2018

BÖHLAU VERLAG WIEN · KÖLN · WEIMAR

Das Wiener Jahrbuch für Kunstgeschichte setzt folgende Zeitschriften fort: Jahrbuch der kaiserl. Königl. Central-Commission zur Erforschung und Erhaltung der Baudenkmale (Jg. I/1856–IV/1860); Jahrbuch der k. k. Zentral-Kommission für Erforschung und Erhaltung der Kunst- und historischen Denkmale (NF I/1903–NF IV/1906); Kunstgeschichtliches Jahrbuch der k. k. Zentral-Kommission für Erforschung und Erhaltung der Kunst- und historischen Denkmale bzw. Jahrbuch des Kunsthistorischen Institutes der k. k. Zentral-Kommission für Denkmalpflege bzw. Jahrbuch des Kunsthistorischen Institutes (Bd. I/1907–Bd. XIV/1920); Jahrbuch für Kunstgeschichte (Bd. I[XV]/1921 f.). Es erscheint unter dem Titel Wiener Jahrbuch für Kunstgeschichte seit dem Band II (XVI)/1923.

Gedruckt mit Unterstützung durch:

Historisch-Kulturwissenschaftliche Fakultät der Universität Wien



universität
wien

REDAKTION :

Für das Bundesdenkmalamt :
N.N.

Für das Institut für Kunstgeschichte der Universität Wien :
MICHAEL VIKTOR SCHWARZ

ISSN 0083–9981

ISBN 978-3-205-20022-2

Druck: Holzhausen, Wolkersdorf

Printed in the EU

INHALTSVERZEICHNIS

HANNA BRINKMANN, JANE BODDY, BEATRICE IMMELMANN, EVA SPECKER, MATTHEW PELOWSKI, HELMUT LEDER, RAPHAEL ROSENBERG Ferocious Colors and Peaceful Lines Describing and Measuring Aesthetic Effects	7
SZILÁRD PAPP Eine frühe Kopie von Giotto's Navicella in Siebenbürgen	27
THOMAS RENKL Dürers „Meerwunder“ als politische Mythenrezeption des deutschen Frühhumanismus	43
WERNER TELESKO <i>Meta-Medien</i> . Zum plurimedialen Charakter von Medaille und Druckgrafik in der Frühen Neuzeit	59
MASSIMILIANO MARAFON PECORARO Un nuovo vulcano sorto dal mare Testimonianze scientifiche, cartografiche e artistiche dell'Isola Ferdinandea	95
WALTRAUD MAIERHOFER Die Titelpuffer von Moritz von Schwind, Vinzenz Gruner und weiteren Wiener Künstlern zu zwei Goethe-Werkausgaben	105
Memorabilia	
IAN VERSTEGEN Otto Pacht and "National" Constants in Late-Gothic Painting	133

FEROCIOUS COLORS AND PEACEFUL LINES DESCRIBING AND MEASURING AESTHETIC EFFECTS

HANNA BRINKMANN, JANE BODDY, BEATRICE IMMELMANN, EVA SPECKER, MATTHEW
PELOWSKI, HELMUT LEDER, RAPHAEL ROSENBERG

In art history, the description of artworks is often based on the description of the effects of lines and colors. This practice has a long tradition. In the middle of the 16th century, Venetian and Florentine art literature theorized *disegno* (drawing) and *colore* (color) as central elements of painting. In the 18th century, the notion of “effect” became a key concept in art literature, and was often used with regard to lines and colors.¹ The word “effect” evoked a relationship between observable properties of artworks and the responses of beholders. Specifically, it suggested a cause and effect relationship between aesthetic forms and perceptions. In this paper, we point to the fact that aesthetic effects were often described with bipolar adjective terms, such as ferocious and peaceful, in both art literature and scientific research (i. e. psychology). The interest in aesthetic effects set the stage for a sustained exchange between art literature and psychological

research that has not yet received the attention that it deserves. The historical analysis brought forward in this paper is the starting point for our own empirical studies, and was developed to critically examine historical assumptions.²

In the first section of this paper, we trace how interest in the effect of colors and lines emerged in 18th century aesthetic theory. In the second section, we show that since Goethe’s *Theory of Colors* the description of such effects was framed by conceptual polarities and we present four exemplary categories frequently used to describe the appearances of lines and colors. In section three, we describe how aesthetic and artistic descriptions were integrated in psychological investigations at the beginning of the 20th century. Our examples center on Wassily Kandinsky and subsequent scientific studies using his writings. Without being fully aware of the history of describing aesthetic effects, psychologists al-

1 H. J. SCHNACKERTZ, Wirkung/Rezeption, in: K. BARCK/M. FONTIUS/F. WOLFZETTEL/B. STEINWACHS (eds.), *Ästhetische Grundbegriffe*. Vol. VI, Stuttgart/Weimar 2010, pp. 670–693. In French: “effet”, in German: “Wirkung”. See C. WEISSERT, Effet/Effekt, in: K. THOMAS (ed.), *Emotionen im Kunstdiskurs Frankreichs im 19. Jahrhundert*. Ein Handbuch der Schlüsselbegriffe, in print. As opposed to emotional response, aesthetic effects do not necessarily entail the actual affective responses that art can elicit in beholders and that is often identified as “aesthetic emotion” by psychologists: G. HOSOYA/I. SCHINDLER/U. BEERMANN/V. WAGNER/W. MENNINGHAUS/M. EID/K. SCHERER, Mapping the Conceptual Domain of Aesthetic Emotion Terms: A Pile-Sort Study, in: *Psychology of Aesthetics, Creativity and the Arts*, August 2017, DOI: 10.1037/aca0000123 (15.12.2017). It is also important to notice that in psychology, responses to art have been far more often studied within the categories of “preference” and “liking”, rather than in regard to aesthetic effects as discussed in this paper. It remains unclear, whether and how aesthetic effects and preference relate to each other.

2 This paper is part of the cross-disciplinary research project “*Universal aesthetics of lines and colors? Effects of culture, expertise, and habituation*” (<http://linesandcolors.univie.ac.at/en/>), financed by the WWTF (CS15–036).

so applied polar categories. A case in point is the “semantic differential”, an experimental tool designed in the 1950s to measure association in a controlled manner. In section four, we give a

brief history of this tool. In the final section, we assess the use of the semantic differential against other tools and discuss our own empirical studies in this field.

I. LINES AND COLORS AS BASIC VISUAL FEATURES

Art literature of the 18th century promoted the idea that specific visual features are structurally foundational for the perception of artworks. In particular lines and colors were analyzed for the aesthetic effects they have on beholders.³ Joseph Addison was one of the first authors who developed an aesthetic theory that centered on the effects caused by aesthetic objects (art or nature) on beholders. In the essay *On the Pleasures of the Imagination* (1712) he distinguished primary and secondary pleasures.⁴ Primary pleasures are produced from what is presented to us in the act of seeing. Addison considered the impressions great, uncommon, and beautiful as “primary pleasures”. He distinguished these from “secondary pleasures”, resulting from a mental comparison, in which the spectator compares the object under current observation with the memory of past experiences.⁵ For Addison, observable features are structurally foundational for primary pleasures.⁶ Beauty, for example, “consists either in the gaiety or variety of colours, in the symmetry and proportion of parts, in the ar-

rangement and disposition of bodies, or, in a just mixture and concurrence of all together. Among these kinds of beauty the eye takes most delight in colours.”⁷ Addison related the specific observable features to the physiological process of perception, and especially the movement of the eyes.⁸ “[A] spacious horizon is an image of liberty, where the eye has room to range abroad, to expatiate at large on the immensity of its views, and to lose itself.”⁹ Despite such a description of the aesthetic effect of the horizon, Addison did not explicitly mention lines as foundational elements for primary pleasures. It was however William Hogarth who first introduced lines to the discussion of aesthetic effects.¹⁰ In *The Analysis of Beauty* (1753), Hogarth argued that the pleasure of any object depends mainly on the type of its (out)lines. He defined the S-shaped serpentine line as the “line of beauty”, which “leads the eye in a pleasing manner along the continuity of its variety.”¹¹

Just a few years later, in 1757, Edmund Burke popularized the distinction between “beauty” and the “sublime” as two different poles.¹² He

3 For a detailed description see especially: R. ROSENBERG, *Entdeckung der Abstraktion*, Turner, Hugo, Moreau, München 2007, pp. 17–53.

4 J. ADDISON, *On the Pleasures of the Imagination*, in: *Spectator* No. 411, June 21, 1712, cited in: *The Works of Joseph Addison. Complete in Three Volumes. Embracing the whole of the “Spectator,” & c.*, New York 1837, Vol. II, p. 138.

5 *Ibid.*, *Spectator* No. 412, June 23, 1712, ed. 1837, p. 138.

6 See J. BODDY, *Disentangling the Aesthetic Effects of Formal Elements of Art*, in: V. A. JOURNEAU/C. V. KAYSER, *Notions esthétiques: La perception sensible organisée*, Paris 2015, pp. 51–61.

7 ADDISON, *On the Pleasures* (cit. n. 4), p. 140.

8 *Ibid.*, pp. 138–140.

9 *Ibid.*, p. 139.

10 ROSENBERG, *Entdeckung der Abstraktion* (cit. n. 3), pp. 21–23.

11 W. HOGARTH, *The Analysis of Beauty*, London 1753, pp. 38–39. See also: P. GUYER, *A History of Modern Aesthetics. Vol. I: The Eighteenth Century*, Cambridge 2014, p. 142.

12 E. BURKE, *A Philosophical Enquiry into the Origin of Our Ideas of the Sublime and Beautiful*, London 2005 [1757], Vol. II, Section XV, p. 195.

emphasized the stimulation of emotions as part of aesthetic experiences and correlated specific colors to specific emotions. He considered a strong red to be cheerful, whereas brown and a deep purple were sad colors.¹³ “Black”, he maintained, “will always have something melancholy in it, because the sensory will always find the change to it from other colours too violent.”¹⁴ Archibald Alison followed Burke with his thesis *On the Nature and Principles of Taste* (1790) in the distinction of “beauty” and “sublime” as the main aesthetic effects and offered the concept of association as an explanation for the perception of these effects.¹⁵ According to Alison these can not only be found in the perception of colors, but also in lines:

*“The most obvious definition of FORM, is that of Matter, bounded or circumscribed by Lines. As no straight line, however, can include Matter, it follows, that the only Lines which can constitute Form, must be either 1st, Angular Lines, or 2ndly, Curved or winding Lines. Every Form whatever must be composed either by one or other of these Lines, or by the Union of them. [...] These different Forms seem to me to be connected in our minds with very different Associations, or to be expressive to us of very different Qualities.”*¹⁶

The qualities that Alison suggested are for example the thickness and the shape of lines: “1. Even Lines are expressive to us of Softness and Smoothness, 2. Uneven Lines are either Angular or Winding. Angular Lines are expressive of Harshness, Roughness. Winding Lines of Pliancy, Delicacy, Ease.”¹⁷

The above-mentioned authors were not the first ones to discuss the aesthetic effects of colors. As early as 1584 Giovanni Paolo Lomazzo gave a systematic description of the characteristics of single colors in his *Trattato dell'arte della pittura*.¹⁸ Yet the major step in this discourse came from Goethe, who published *Zur Farbenlehre*, the very first rigorous and comprehensive theory of color in 1810.¹⁹ In this treatise, he considered color in its phenomenal, experiential and moral qualities, thereby distinguishing the effect of colors on the eye (physical) and on the mind (psychological).²⁰ In this context, he explicitly used the wording “aesthetic effect”.²¹

Goethe's *Theory of Colors* had a broad aftermath. Artists like Philipp Otto Runge, William Turner, and Wassily Kandinsky, just to name the most well-known examples, used it in artistic practice.²² The text became a reference point for the scientific investigation of aesthetics. Due in large part to the substantial skepticism of Goethe's criticism of Newton, the impact within the realm of the sciences was more limited, but

13 Ibid., Vol. II, Section XV, p. 159.

14 Ibid., Vol. II, Section XVIII, p. 232.

15 A. ALISON, *Essays on the Nature and Principles of Taste*, Edinburgh 1790, p. 239.

16 Ibid., pp. 232–233.

17 Ibid., pp. 238–239.

18 G. P. LOMAZZO, *Trattato dell' arte de la Pittura*. [...] Diviso in sette libri. Ne' quali si contiene tutta la Theorica & la prattica d'essa pittura, Milano 1584, cf. Rosenberg, *Entdeckung der Abstraktion* (cit. n. 3), p. 43.

19 T. LERSCH, *Farbenlehre*, in: *Reallexikon zur deutschen Kunstgeschichte*, München 1974, Vol. VII, cols. 157–274.

20 J. W. VON GOETHE, *Zur Farbenlehre*, Tübingen 1810. Translated from German into English by Charles Eastlake in 1840: C. L. EASTLAKE, *Goethe's Theory of Colours*. Translated from the German: with Notes by Charles Lock Eastlake, R.A., F.R.S., London 1840. Part II (p. 59) refers to “physical” and part VI (p. 306) to “psychological”.

21 EASTLAKE, *Goethe's Theory of Colours* (cit. n. 20), § 848. In German: “ästhetische Wirkung”.

22 J. GAGE, *Color and Culture. Practice and Meaning from Antiquity to Abstraction*, Boston/Toronto/London 1993, pp. 201–204.

it did reach up until Thomas Young and Hermann von Helmholtz.²³ In 1829, the young Gustav Theodor Fechner, who would be later one of the founders of psychology, wrote a short and critical essay on Goethe's color theory. Similar to Goethe's discrimination on the effect of colors on the eye and mind, Fechner distinguished a direct impression ("directem Eindruck") and an associative impression ("associativem Eindruck").²⁴

The direct impression resulted from immediate perception and was induced by the forms themselves, comparable to Goethe's effect on the eye. The associative impression was conditioned by the purpose and meaning that the viewer attributes to a given object and, according to Fechner, was induced by related recollections of any sort. It is thus similar to Goethe's effect on the mind.

2. THE DESCRIPTION OF AESTHETIC EFFECTS AS POLARITIES

Goethe's book on colors was a milestone in several regards. The crucial factor for us is that although he did not invent the concept of aesthetic effects, he took this discourse far beyond the discussion on colors and brought it to a new level through his high authority, his explicit and systematic discussion, and the underlining of the principle of polarities. Polarities were a central category in Goethe's worldview, a kind of world-formula. Goethe assumed that nature generally functions through antagonistic forces, opposed entities that attract and repulse each other.²⁵ He might have taken this idea from natural sciences. He may have drawn from prevailing theories concerning magnetism, electricity and chemical attraction, which were already in his time

described in polarities between a positive and a negative pole.²⁶ In Goethe's color theory, blue and yellow represented the plus and the minus pole. Yellow represented the plus side, and was attributed to warmth, to light and to strength. Whereas blue was associated with the minus side, signified to darkness, cold and weakness. Hence, colors and their aesthetic effects varied on a continuous scale between two poles, described for instance as warm or cold.

A general review of art literature since the 19th century provides numerous examples for words often used to encode aesthetic effects using such polar terms.²⁷ Such words usually do not describe the form of lines, for example a "straight" line, or the darkness or clarity of color.

23 See: J. NEUBAUER, *Quellen und Wellen der Wissenschaft. Goethe und Thomas Young*, in: K. MANGER (ed.), *Goethe und die Weltkultur*, Heidelberg 2003, S. 17–32. H. Von Helmholtz, *Ueber Goethe's naturwissenschaftliche Arbeiten*, in: *Vorträge und Reden von Hermann von Helmholtz*, 1903 [1853], Vol. I, pp. 23–45, 46–47.

24 G. T. FECHNER, *Kurze Darstellung der Goethe'schen Farbentheorie*, in: J. B. BIOT (ed.), *Lehrbuch der Experimental-Physik*, Leipzig 1829, pp. 486–488. Cf. EASTLAKE, *Goethe's theory of Colours* (cit. n. 20), §915 and G. T. FECHNER, *Vorschule der Aesthetik*, II Vol., Leipzig 1876, pp. I, 100 and II, 212.

25 R. C. ZIMMERMANN, *Goethes Polaritätsdenken im geistigen Kontext des 18. Jahrhunderts*, in: F. MARTINI/W. MÜLLER-SEIDEL/B. ZELLER (eds.), *Jahrbuch der deutschen Schillergesellschaft*, Stuttgart 1974, pp. 304–347. Goethe writes for instance „[Aus der] Anziehungskraft und Zurückstossungskraft des Wesens der Materie [...] ging mir die Urpolarität aller Wesen hervor, welche die unendliche Mannigfaltigkeit der Erscheinungen durchdringt und belebt“, in: *Goethes Werke*, X, *Autobiographische Schriften II*, München 2002, p. 314.

26 P. PROBST, *Polarität*, in: J. RITTER/K. GRÜNDER (eds.), *Historisches Wörterbuch der Philosophie*, Vol. VII, P–Q, Basel 1989, pp. 1026–1027.

27 We do not review other related subjects which are repeatedly discussed when it comes to the aesthetic effects of colors and lines, such as color combining, contrast and also the correspondences of particular lines with particular colors. Runge for example described the effect of the combination of the three colors blue, yellow and red with their "opposite" colors (orange, violet and green) as "harmonic", whereas he considered the effect of a combination with only the three pure colors as "disharmonic", see P. O. RUNGE, *Farbenkugel*, Köln 1999, pp. 28–39.

Rather, they describe the appearance in reference to sensory crossovers and are often emotionally connoted. In many instances, the crossover is an act of metaphor and should not be confused with actual synesthesia, which is elicited automatically, is extremely idiosyncratic and

highly stable in individuals over time.²⁸ For the sake of space, we shall limit ourselves to four examples of categories involving sensory crossovers (temperature and sound) and expressive qualities (feelings and masculinity/femininity).

2 a. TEMPERATURE

A pervasive category for the description of color effects is temperature – colors were often arranged on a scale between cold and warm. Indeed, Joshua Reynolds suggested this scale in his *Discourse VIII* (1778).²⁹ Although still relatively unknown, Mary Gartside (1805) was possibly the first art theorist to use temperature as a category for a systematic classification of colors. She had connections to the Royal Academy in London and was hence probably aware of Reynolds' writings.³⁰ For his part, Goethe, who was also familiar with Reynolds, suggested that the perception of warm colors was akin to the sensation of warmth: "We find from experience, again, that yellow excites a warm and agreeable impression."

Conversely, blue by Goethe's estimations "gives us an impression of cold, and thus, again, reminds us of shade."³¹ In a similar fashion, William Turner divided his "cycloid of light" in "opposed compounds", which are warm and cold. He suggested a structure of color that progressed from warm to cold: red, orange, yellow, green, blue and violet.³²

The ordering of colors along the axis warm-cold remained relevant in the twentieth century, often with references to Goethe. For instance, Adolf Hölzel's lecture at the 1. *Deutscher Farbentag* (1st German Day of Color) in 1919 was a direct plea for Goethe's color theory.³³ Hölzel responded critically to the color theory of Wilhelm

28 Cf. A. GOTTDANG, *Vorbild Musik. Die Geschichte einer Idee in der Malerei im deutschsprachigen Raum 1780–1915*, München/ Berlin 2004, p. 109 and R. ROSENBERG, *Die Kartographie der Aura aus dem Geist der Wirkungsästhetik. Synästhesie und das Verhältnis von Kunst und Esoterik um 1900*, in: M. NEUGEBAUER-Wölk/R. GEFFARTH/M. MEUMANN (eds.), *Aufklärung und Esoterik: Wege in die Moderne*, Berlin 2013, pp. 583–604, esp. p. 586 f.

29 J. REYNOLDS, *The Works of Sir Joshua Reynolds. Containing His Discourses, Idlers, A Journey to Flanders and Holland, and his Commentary on Du Fresnoy's Art of Painting. In Two Volumes. To Which Is Prefixed an Account of the Life and Writing of the Author*, By Edmond Malone, London 1797, Vol. I, p. 182.

30 M. GARTSIDE, *An Essay on Light and Shade, on Colours, and on Composition in General*, London 1805, p. 13, cf. ROSENBERG, *Entdeckung der Abstraktion* (cit. n. 3), p. 48, 325 and A. LOSKE, *Mary Gartside: A Female Colour Theorist in Georgian England*, in: *St Andrews Journal of Art History and Museum Studies* 14, 2010, pp. 17–30, esp. p. 18.

31 Eastlake, *Goethe's Theory of Colours* (cit. n. 20), §768 and §282.

32 Cited from: J. GAGE, *Colour in Turner*, London 1969, p. 210. Turner possessed the English translation of Goethe's *Theory of Colours* from 1840 and dedicated two paintings to it, (*Ibid.*, p. 18). Turner also knew Reynolds: he studied at the time Reynolds was the president of the Royal Academy. Given that Turner's lecture in 1827 included remarks on warm and cold colors, and that the first English translation of Goethe's *Theory of Colours* was published in 1840, Turner could have derived the idea from Reynolds rather than Goethe (GAGE, *Colour in Turner* (cit. n. 32), pp. 209–210).

33 A. HÖLZEL, *Rede von Adolf Hölzel auf dem ersten deutschen Farbentag. Erster deutscher Farbentag auf der 9. Jahresversammlung des deutschen Werkbundes in Stuttgart*, Berlin 1919, pp. 10–26, esp. p. 11. "Art is always a matter of feeling. This talk, therefore, should mention a few theoretical things related to sensation. As to the development of visual harmony in artworks, our circle, especially I, stands on Goethean fundamentals and seeks to continue on this fundament. To us, Goethe is not outdated. Rather, he is eternal, at least as long as there are human eyes". Orig-

Ostwald who had given a talk about the failure of Goethe at the same event.³⁴ Josef Albers underlined the influence of culture on the categorization of warm and cold colors: “as to warm and cool, it is accepted in Western tradition that normally blue appears cool and that the adjacent group, yellow-orange-red looks warm.”³⁵ In a letter to Rudolf Arnheim, Albers acknowledged that his early readings of Goethe’s color theory had influenced him.³⁶ Arnheim, in turn, dedicated a whole chapter in his book *Art and Visual Perception* (1954) to the tradition of describing the effects of color with the help of temperature and referred to many artists, such as Albers, Itten, and Kandinsky.³⁷

Temperature was also attributed to lines, but often in combination with colors. In 1827 Humbert de Superville ascribed red to an ascending

line, white to a horizontal line and black to a falling line. This opinion was followed by Paul Signac, who in 1921 recommended that artists combine rising lines with warm colors and falling lines with cold colors.³⁸ Similarly, Kandinsky claimed in 1926 that the horizontal line is flat and cold, the vertical line is warm, whereas the diagonal line unites both poles.³⁹ August Endell in 1898 took an even more comprehensive approach by composing a table with 64 emotion words that showed that lines can consistently induce feelings in us.⁴⁰ He ordered the words on the dimensions tempo (“Tempo”) and tension (“Spannung” translated in fig. 2 as effort) (fig. 1 and 2).⁴¹ Lines with medium tension and high speed are “cold” (translated as “frigid”), whereas lines with a bit less tension and almost no speed are “warm”.

2 b. SOUND

A typical cross-sensorial effect attributed to color is sound. Louis Bertrand Castel, the inventor of an ocular harpsichord (1725), suggested

precise relations between sounds and colors.⁴² Goethe drew from Castel’s idea that dark and light colors corresponded to low and high tones

inal: „Kunst ist und bleibt Empfindungssache. Hier soll darum einiges der Empfindung nahestehendes Theoretische erwähnt sein. Unser Kreis, speziell ich, stehen hinsichtlich einer bildharmonischen Entwicklung im Kunstwerke, auf Goetheschem Fundamente und suchen darauf weiterzubauen. Goethe ist für uns nicht veraltet, sondern ewig, wenigstens solange es menschliche Augen gibt“ (translation by the authors).

- 34 W. OSTWALD, *Farbkunde. Ein Hilfsbuch für Chemiker, Physiker, Naturforscher, Ärzte, Physiologen, Psychologen, Koloristen, Farbtechniker, Drucker, Keramiker, Färber, Weber, Maler, Kunstgewerber, Musterzeichner, Plakatkünstler, Modisten*, Leipzig 1923.
- 35 J. ALBERS, *Interaction of Color*, New Complete Edition, New Haven/London 2009 [1963], p. 55.
- 36 Josef Albers to Rudolf Arnheim, March 14, 1973, cited in R. ARNHEIM, *A Critical Account of some of Joseph [sic] Albers’ Concepts of Color*, in: *Leonardo* 15:2, 1982, pp. 174–76. For more detail on Albers and Goethe see: B. DANILOWITZ, *A Short History of Josef Albers’s Interaction of Color*, Amherst 2015, pp. 13–28.
- 37 R. ARNHEIM, *Art and Visual Perception. A psychology of the Creative Eye*, Berkeley/Los Angeles/London, 1974 [1954], pp. 369, 371.
- 38 D. P. G. HUMBERT DE SUPERVILLE, *Essai sur les signes inconditionnels dans l’art*, Leiden 1827, p. 23; P. SIGNAC, *D’Eugène Delacroix au Néo-Impressionnisme*, Paris 1921, p. 63.
- 39 W. KANDINSKY, *Point and Line to Plane*, New York 1947 [1926], pp. 58–59.
- 40 A. ENDELL, *Formenschönheit und dekorative Kunst*, „II. Die gerade Linie“, in: *Dekorative Kunst* I, 2, 1898, pp. 119–123, esp. p. 121.
- 41 Translation in: C. HARRISON/P. WOOD (eds.), *Art in Theory 1900–1990. An Anthology of Changing Ideas*, Oxford 1992, p. 65.
- 42 L. B. CASTEL, *L’optique des couleurs, fondée sur les simples Observations, & tournée sur-tout à la pratique de la Peinture, de la Teinture & des autres Arts Coloristes*, Paris 1740, p. 299–300.

FORMENSCHÖNHEIT UND DEKORATIVE KUNST								
boshaft	höhnisch	hochmütig	pathetisch	kalt	unerbittlich	grausam	fürchterlich	
frivol	übermütig	herausfordernd	stolz	streng	gewaltsam	wild	grässlich	
kokett	chic	prächtig	kühn	rücksichtslos	wuchtig	kolossal	scheusslich	
geziert	graziös	elegant	energisch	kraftvoll	derb	ungeschlacht	roh	
süßlich	zierlich	geschmeidig	feurig	stark	gedrungen	gewaltig	entsetzlich	
fade	zart	hingebend	gross	vornehm	machtvoll	ungeheuer	furchtbar	
nüchtern	einfach	innig	warm	ernst	tief	erhaben	grausig	
stumpf	kraftlos	müde	bekümmert	traurig	schwerfällig	düster	finster	

Quick ↑ T E M P O ↓ Slow	Malicious	Scornful	Haughty	Pathetic	Frigid	Pitiless	Cruel	Terrible
	Facetious	Frivolous	Provocative	Arrogant	Harsh	Violent	Savage	HIDEOUS
	Coquettish	Chic	Ebulliant	Daring	Reckless	Majestic	Awesome	Dreadful
	Affected	Gracious	Elegant	Energetic	Vigorous	Firm	Brutal	Bestial
	Sweet	Dainty	Flexible	Fiery	Strong	Rugged	Powerful	Frightful
	Insipid	Delicate	Devoted	Generous	Distinguished	Mighty	Monstrous	Awful
	Vacuous	Straight-forward	Sincere	Warm	Solemn	Profound	Sublime	Gruesome
	Dim	Weak	Tired	Troubled	Sad	Melancholic	Sombre	Desperate
	Light ← EFFORT → Heavy							

[It is impossible, of course, to translate the German words precisely: this Table is intended to give the essence of Endell's idea.]

Fig. 1 and 2: August Endell, *Table of words capturing nuances of feeling (Gefühlsnuancen)* 1898 and English translation in HARRISON / WOOD 1992

and used multiple scales describing the cross-sensorial effect of color and taste with the example of yellow-red as sour and blue as alkaline.⁴³ This was also reflected in the empirical investigation of aesthetics. Fechner, for instance, asserted in the *Vorschule der Aesthetik* (1876), that blue could be compared to the sound of a flute and red to a very loud tone like a trumpet fanfare.⁴⁴ In 1912 Kandinsky was more specific, differentiating the hue of blue: "In music, light blue is like

a flute, dark blue like a cello, and when still darker, it becomes a wonderful double bass".⁴⁵

In similar fashion, from around 1910 on Gertrud Grunow assigned (twelve) colors to tones and to movements, stating that "[t]he tone f is the equivalent of the color green. Blue can be assigned to a bowing and yellow to a stretching movement".⁴⁶ Other artists connected to the Bauhaus, such as Paul Klee and Johannes Itten, also used musical terms to describe the effects of lines and colors.⁴⁷ While Itten focused on col-

43 J. W. VON GOETHE, *Maximen und Reflexionen*, Zürich 2001, p. 284. Goethe and Castel see: J. JEWANSKI, *Ist C = Rot? Eine Kultur- und Wissenschaftsgeschichte zum Problem der wechselseitigen Beziehung zwischen Ton und Farbe; von Aristoteles bis Goethe*, Sinzig 1999, p. 569.

44 FECHNER, *Vorschule der Aesthetik* (cit. n. 24), p. 216.

45 W. KANDINSKY, *On the Spiritual in Art*, New York 1946, p. 65.

46 G. GRUNOW, *Der Gleichgewichtskreis. Ein Bauhaus Dokument*, Weimar 2001, pp. 42–48 (translation by authors). See also: R. RADRIZZANI (ed.), *Die Grunow-Lehre. Die bewegende Kraft von Klang und Farbe*, Wilhelmshaven 2004, p. 30.

47 H. MAKABE, *What Makes Synesthesia so Fascinating? An Overview of the History of its Study and Practice*, in: *Art and Music – Search for New Synesthesia*, Tokyo 2012, pp. 177–194.

or accords (which involved at least two different colors),⁴⁸ Klee is one of the few authors who linked sounds to lines. In a lecture from 1922, he attempted to translate some musical notes into a drawn line.⁴⁹ In 1927, the psychologist Albert Wellek published research on the history of words used to describe color-hearing. One of his conclusions was that the most usual adjec-

tive pairs used to describe acoustic analogies can be divided between a plus and a minus pole. On the plus side are high pitches and warm colors on the minus side low pitches and cold colors. In Wellek's view, high and low pitches also corresponded to happy and sad emotions. He saw the plus side as the side of life and the minus side the one of death.⁵⁰

2 C. FEELINGS

The arrangement of plus and minus was transferred to happy and sad, and, more generally, to positive and negative feelings. In Goethe's words, "[t]he colors on the *plus* side are yellow, red-yellow (orange), yellow-red (minium, cinnabar). The feelings they excite are quick, lively, aspiring. The colors on the *minus* side are blue, red-blue, and blue-red. They produce a restless, susceptible, anxious impression."⁵¹ Similarly, Fechner assigned colors to feelings: red, orange and yellow are active and exciting, blue and violet are receptive.⁵² In 1910 August Macke wrote in a letter to Franz Marc: "After our request, our friend Job played the three corresponding sounds on the piano for me, well, she claimed to play the same important role in music. So: Blue, Yellow, Red – parallel phenomena: sad, happy, brutal (in tones as well as in colors). All the lines (or melody) determine the sequence of the colours (or sounds).

Ascending and descending melodies that are in full integration".⁵³

A dominant theory was that the direction of lines is correlated with feelings. In 1827 Humbert de Superville stated that ascending lines are perceived as happy, descending as sad, horizontal as neutral.⁵⁴ Signac, who was probably aware of Humbert de Superville's theory through its reception by Charles Blanc, stated that a horizontal line elicits a calm effect, a rising line happiness, and a falling line sadness,⁵⁵ while Georges Seurat proposed that a happy line is above the horizontal, a calm line is horizontal and a descending line sad.⁵⁶ A similar idea is expressed in Endell's table of feelings (fig. 1 and 2): the 36 inner terms describe positive feelings ("Lustgefühl"), whereas the 28 adjective terms outside the square are, according to Endell, negative ones ("Unlust").

48 J. ITTEN, *Kunst der Farbe*, Ravensburg 1961, pp. 21–32.

49 P. KLEE, *Beiträge zur bildnerischen Formlehre*. Faksimilierte Ausgabe des Originalmanuskripts von Paul Klees erstem Vortragszyklus am staatlichen Bauhaus Weimar 1921/22, Basel 1979, p. 56. Cf. R. BONNEFOIT, *Die Linientheorie von Paul Klee*, Petersberg 2009.

50 A. WELLEK, *Doppelempfinden und Programmusik*. *Beiträge zur Psychologie, Kritik und Geschichte der Sinnessprehung und Sinnensymbolik*, Wien/Prag 1927, Vol. II, p. 24.

51 EASTLAKE, *Goethe's Theory of Colours* (cit. n. 20), §764.

52 FECHNER, *Vorschule der Aesthetik* (cit. n. 24), p. 215.

53 English translation from W. COHEN (ed.), *August Macke*, New York 2013, p. 72.

54 HUMPERT DE SUPERVILLE, *Essai sur les signes* (cit. n. 38), pp. 6–7.

55 SIGNAC, *D'Eugène Delacroix* (cit. n. 38), p. 63.

56 For an English translation see: J. C. TAYLOR (ed.), *Nineteenth-Century Theories of Art*, Berkeley/Los Angeles/London 1987, p. 542.

2 d. MASCULINITY/FEMININITY

Characteristics of lines and colors were also often described as masculine or feminine. Runge understood the “warm side” of the color spectrum (yellow-orange) as male and the “cold side” (blue-violet) as female.⁵⁷ He associated “active” with “male”, but also with “warm”, whereas “passive” was associated with “feminine” and “cold”.⁵⁸ The same classification could be made in the opposite direction. In 1910, Marc expressed in a letter to Macke that “blue is the male principle, astrigent and spiritual. Yellow is the female principle, gentle, gay and sensual. Red is matter, brutal and heavy and always the color to be opposed and overcome by the other two”.⁵⁹ Adolf Loos (1908) and Piet Mondrian (1914) asserted that the horizontal line is female and the vertical line is male.⁶⁰ And this again derived from the association with active and passive, also asserted by Kandinsky in 1926:

*“The same thing can with complete justification be said about the horizontal and the vertical – low and high. The former is lying, the latter is standing, walking, moving about, finally climbing upward. Supporting – growing. Passive – active. Relatively: feminine – masculine.”*⁶¹

Finally, Klee distinguished between a straight line and a free line. The straight line appears aggressive, active and therefore masculine.⁶²

The various examples quoted above show that the aesthetic effects of lines and colors were often described in art literature by involving cross-modal perceptions: qualities from different sense modalities. Such perceptions are not meant to be real in a physical sense. Whereas a color may vary in luminance, none of the authors quoted so far assumed that it can physically transform into taste or sound. However, it is a common feature in art literature since the 18th century that the perception of these qualities is not purely subjective, but rather collectively shared. For instance the authors quoted so far, as well as many others, assumed that something about the impression of color is really quite similar to the impression of temperature. This idea reappears in subsequent theories and leads to two basic questions: 1) How can one assess such effects? and 2) What is the cause of the effects themselves?

Since authors in art literature considered aesthetic qualities of lines and colors as objectively given, they tended not to discuss their cause, or did so only in an implicit or transcendental way. Explanations came rather from psychologists. Fechner already explained that color associations may derive from references to objects. For Fechner, the object associations that humans are most likely to have with the color yellow are the lemon and the canary.⁶³ Kandinsky delivers a post-hoc confirmation of Fechner statement.⁶⁴ In 1912 he asserted that yellow can be described

⁵⁷ P. O. RUNGE, *Hinterlassene Schriften*, Göttingen 1965 [1810], p. 164.

⁵⁸ This is in accordance with Goethe, who also attributed “activity” to red. EASTLAKE, *Goethe’s Theory of Colours* (cit. n. 20), § 764.

⁵⁹ A. MACKE/F. MARC, *Briefwechsel*, Köln 1964, p. 28.

⁶⁰ A. LOOS, *Ornament und Verbrechen* (1908), in: Loos, *Sämtliche Schriften*, Vol. I, Wien/München 1962, pp. 276–288, p. 277; P. MONDRIAN, *Two Mondrian Sketchbooks: 1912–1914*, Amsterdam 1969, p. 16.

⁶¹ KANDINSKY, *Point and Line to Plane* (cit. n. 39), p. 63.

⁶² P. KLEE, *Notebooks*, Vol. I: *The Thinking Eye*, London 1961, p. 333.

⁶³ FECHNER, *Vorschule der Aesthetik* (cit. n. 24), p. 101.

⁶⁴ KANDINSKY, *On the Spiritual in Art* (cit. n. 45), p. 63. Kandinsky makes no explicit reference to Fechner. Still, we know that he was familiar with psychological literature of his time, for example with Wilhelm Wundt, see: A. WEISSENHACH, *Wassily Kandinsky. Unterricht am Bauhaus 1923–1933*, Vol. I, pp. 194, 508.

as being “shrill” and “sour”, which he explained with correspondences to the sound of a canary bird and the taste of lemon.⁶⁵ He also asserted a direct form-color correspondence, stating that “sharp colors have a stronger sound in sharp forms (e.g., yellow in a triangle).”⁶⁶

Over time, the question of how to uncover and eventually measure aesthetic effects was more and more discussed. Goethe already conducted introspective experiments with various instruments to achieve this description of the effects of colors. For instance, he described how

objects seen through a yellow glass became warm and vivid and that the same objects seen through a blue glass appeared cold, sad and dead to him.⁶⁷ Experimental methodologies were however mainly developed in the 20th century and involved exchanges between artists and psychologists to develop experimental methodologies to study the perceptual experience of aesthetic forms. To illustrate this, in the next section we concentrate on Kandinsky, who seems to be the foremost example for such cooperations.

3. THE EXPERIMENTAL INVESTIGATION OF AESTHETIC EFFECTS

Having been an academic researcher before starting an artist career at the age of 30 in 1896, Kandinsky was even as an artist, eager to conduct empirical experiments and inspired a number of psychological investigations. In 1920, he tried to set up a department for psychophysical aesthetics at the Institute of Artistic Culture (INChUK) in Moscow. This was part of a larger project, in which psychologists, philosophers, and other thinkers from various domains worked together on a program to establish a science of art.⁶⁸ In this context, Kandinsky mobilized the color theory that he had developed in *Über das Geistige in der Kunst* (1912) for a questionnaire on painting. It aimed to investigate how feelings and forms correlated and included questions about similarities between forms and colors. Kandinsky claimed that three elementary plane-shapes – the circle, the triangle, and the square – correspond to the colors blue, yellow and red respectively. For Kandinsky, these colors and shapes connected on a fundamen-

tal level for following reasons: Yellow manifested a centrifugal movement, coming toward the viewer. This movement expressed madness and impurity. The link to the triangle was clear, Kandinsky insisted, because both yellow and the triangle expressed sharpness and acuity. In similar fashion, blue manifested effects that were equivalent to the circle; passive, round and serene. Finally, red possessed expressive properties that can be described in similar terms to the square, which made their fundamental connection evident. With the questionnaire, Kandinsky reached out to a wider audience and asked his colleagues to write down their impressions as accurately as possible:

*“For example, how do you see a triangle? Would you say that it moves and, if so, in which direction? Would you say that it is smarter than a square? Is not the sensation elicited by a triangle more similar to that elicited by a lemon? What is the singing of a canary more like, a triangle or a square?”*⁶⁹

65 KANDINSKY, *On the Spiritual in Art* (cit. n. 45), p. 40.

66 *Ibid.*, p. 46.

67 J. W. VON GOETHE, *Betrachtungen über die Farben*, geschrieben vor Mainz im Juni 1793, zit. nach: *Beiträge zur Optik und Anfänge der Farbenlehre. 1790–1808*, Weimar 1951, pp. 125–126.

68 It drew on the work of Wundt and Fechner. See: M. VÖHRINGER, *Avantgarde und Psychotechnik. Wissenschaft, Kunst und Technik der Wahrnehmungsexperimente in der frühen Sowjetunion*, Göttingen 2007, p. 77; Weissenbach, Wassily Kandinsky (cit. n. 64).

69 W. KANDINSKY, *Questionnaire* (1920). Translation from Russian: J. E. BOWLT, in: *Experiment* 2002, pp. 157–160, here p. 158.

With such questions, Kandinsky asked people to connect seemingly unrelated objects in their minds and to reflect on the feelings that this elicited. His objective was to establish “how frequent the same answers are” and to find “conformities in impression and observation.”⁷⁰ This was not only meant to refer to the attribution of correspondences, but also to the rationale that people offered for these attributions. A year later, in 1921, Kandinsky continued this approach at the RAKhN (Russian Academy of Sciences) and incorporated it in his pedagogical work at the Bauhaus in 1923. He taught his color-shape-correspondences to first semester students and his assignment became widely used at the Bauhaus.⁷¹ Together with the wall painting workshop, Kandinsky distributed a questionnaire to verify the relationship between the three basic shapes and the three primary colors, but, unfortunately, he never published any results. 27 completed questionnaires still exist today.⁷²

This obviously raises the question: Is the matching of color-shape-correspondences, such as the system Kandinsky postulated, something which people should agree on? The scien-

tific experimentation with Kandinsky's theory continues today. In the last fifteen years, several psychologists attempted to test the Bauhaus questionnaire with contemporary spectators.⁷³ Those studies showed that most test persons proposed different combinations from the ones that Kandinsky had assumed and that those combinations are rather driven by object associations: for example, the sun is yellow, hence round; triangular street signs are red.

Even before Kandinsky set up psychological studies at art academies, psychologists started to integrate his paintings into experimental studies to test the perception of similarity relations between visual forms and the perception of expressive properties. The psychologist Helge Lundholm examined Kandinsky's *Composition VI* as a topic for psychological investigation as early as 1917.⁷⁴ In 1921, Lundholm published an experiment that studied “the emotional quality” that people ascribed to different types of lines.⁷⁵ He wanted to answer the following questions: “Is the affective character of the line a quality which is bound to the line itself, or is it suggested by the literary subject of the masterpiece? Further-

70 Ibid. p. 160.

71 WEISSENBACH, Wassily Kandinsky (cit. n. 64), p. 205; M. DROSTE (ed.), Bauhaus, Köln 2006, pp. 90, 146.

72 The Bauhaus-Archiv in Berlin holds one completed questionnaire (inv. nr. 991). It was mistakenly attributed to Alfred Arndt. It remains unclear who actually filled out the questionnaire. We thank Nina Schöning of the Bauhaus-Archiv for this information. For a reproduction, see: GAGE, Color and Culture (cit. n. 22), p. 261.

The Centre Pompidou in Paris holds 26 more completed questionnaires. Two in the Cabinet d'art graphique (inv. nr. 04-515919/AM81-65-1249(1) and inv.nr. 04-515920/AM81-65-1249(2)) and 24 in the Fonds Kandinsky (inv.nr. CP, FK 9200). For a reproduction of the first questionnaire, see M. Tupitsyn, Gegen Kandinsky/Against Kandinsky, Ostfildern 2006, p. 45. For those in the Fonds Kandinsky see: A. WEISSENBACH, Wassily Kandinsky (cit. n. 64), Vol. II, pp. 131-142.

73 T. JACOBSEN/C. WOLSDORFF, Does History Affect Aesthetic Preference? Kandinsky's Teaching of Colour-Form Correspondence, Empirical Aesthetics, and the Bauhaus, in: The Design Journal 10/3, 2015, pp. 16-27; T. JACOBSEN, Kandinsky's Questionnaire Revisited: Fundamental Correspondence of Basic Colors and Forms?, in: Perceptual and Motor Skills 95, 2002, pp. 903-913; A. V. KHARKHURIN, Is Triangle Really Yellow? An Empirical Investigation of Kandinsky's Correspondence Theory, in: Empirical Studies of the Arts 30, 2012, pp. 167-182; A. D. J. MAKIN/S. M. WUERGER, The IAT Shows no Evidence for Kandinsky's Color-Shape Associations, in: Frontiers in Psychology 4, 2013, pp. 1-7; N. CHEN/K. TANAKA/K. WATANABE, Color-Shape Associations Revealed with Implicit Association Tests, in: PLoS ONE 10, 2015, pp. 1-15.

74 B. LÄRKNER, “Humbler” or “genius”? When Kandinsky came to Sweden, in: A. ROSDAHL (ed.), Nya perspektiv på Kandinsky. New perspectives on Kandinsky, Konsthall & Sydsvenska Dagbladet, Malmö 1990, pp. 23-35, here p. 31.

75 H. LUNDHOLM, The Affective Tone of Lines: Experimental Researches, in: Psychological Review 1921, pp. 43-60.

more, is this quality a phenomenon that appears equally to different observers?"⁷⁶ In his experiment, he asked eight participants to "draw lines, each of which was to express the affective tone of an adjective given verbally."⁷⁷ He used 48 adjectives, divided into 13 groups of synonyms, which represented feelings such as sadness, tranquility, cheerfulness and anger. He classified the lines that the participants produced in groups of ones with angles and ones with curves and specified the interval of the angles and curves (big, medium, and small: fig. 3). He found that lines that matched with the emotions sad, quiet, lazy, etc., had large curves, whereas emotions such as merry were expressed with lines with short waves (e.g., type c in fig. 3).⁷⁸ Lundholm found consensus among observers, which, he thought, was "probably dependent upon the suggestion of movement in the lines."⁷⁹



Fig. 3: Helge Lundholm, *Examples of types of lines used in the experiment published in 1921*

Building on Lundholm's experiment, Poffenberger and Barrows (1924) used a set of similar adjectives to examine the relation between emo-

tions and lines; however, they gave their participants a different task.⁸⁰ Lundholm had asked them to produce drawings that resemble particular feeling states. Poffenberger and Barrows looked at the interaction in the opposite direction and asked participants to match words to a set of lines. Similar studies were pursued in the 1930s. Kate Hevner (1935) for instance asked participants to match a list of adjectives to linear patterns and colors (red and blue). Her results showed that "curves are found to be serene, graceful, and tender-sentimental. Angles are robust, vigorous, and somewhat more dignified."⁸¹

Gertrud Grunow is another example of an artist with affiliations to psychology's methods and ideas. In 1911, she developed a theory about color-tone-movement-correspondences and later cooperated with the psychologist Heinz Werner to test her assumptions. In her essay *Von der Wirkung der Farbe auf das sehende Auge* (The effect of color on the viewing eye), Grunow claimed that the eyes behave differently when viewing different colors. Looking at red elicits a "mode of beholding", while looking at yellow leads to a "mode of observing". She wrote that she had studied "several hundred human beings".⁸² Werner referred to this experiment and further collaborative studies with Grunow in his book *Einführung in die Entwicklungspsychologie*.⁸³ Werner cited biographical statements and theoretical proposals from artists, which, for him, were a valuable source for investigating a mode

⁷⁶ Ibid., p. 43.

⁷⁷ Ibid.

⁷⁸ Ibid., p. 49.

⁷⁹ Ibid., p. 60.

⁸⁰ A. POFFENBERGER/B. BARROWS, The Feeling Value of Lines, in: *Journal of Applied Psychology* 8, 1924, pp. 187–205.

⁸¹ K. HEVNER, Experimental studies of the affective value of colors and lines, in: *Journal of Applied Psychology* 1914, pp. 385–398, here p. 398. Another example is Krauss who asked participants to match lines to gold, iron and silver (R. KRAUSS, Über graphischen Ausdruck. Eine experimentelle Untersuchung über das Erzeugen und Ausdeuten von gegenstandsfreien Linien. Beihefte zur Zeitschrift für angewandte Psychologie, Leipzig 1930, cited by H. Werner, *Einführung in die Entwicklungspsychologie*, München 1959, pp. 46–47).

⁸² G. GRUNOW, Von der Wirkung der Farbe auf das sehende Auge, in: *Kunst und Jugend* 15, 1935, pp. 149–150, esp. p. 149.

⁸³ WERNER, *Einführung* (cit. n. 81), pp. 66–67.

of physiognomic perception. Despite the subjective nature of artists' perceptions, they could be mobilized for psychological inquiry.

Picking up this tradition, in 1995, Shigeko Takahashi, conducted an experiment wherein one group of participants (from an art academy) was asked to produce abstract line drawings with a specific expressive content. Then a second group (not from an art academy) was asked to assign adjectival terms to these line drawings.⁸⁴ The terms were anger, joy, tranquility, depression,

human energy, femininity and illness.⁸⁵ Takahashi found consensus among the viewers, and suggested "that many people may share a synesthetic tendency by which drawings communicate on a level of shared awareness rather than on a level of verbal expression and visual symbols".⁸⁶ Takahashi's study was the starting point for work by Despina Stamatopoulou, whose research investigates the influence of cultural differences on the perception of aesthetic effects.⁸⁷

4. CONTROLLED ASSOCIATIONS

Moving back to the mid-twentieth century, in psychology, a common experimental tool that has similar properties to the tradition discussed above is the semantic differential. It was developed by Charles Osgood in the 1950s as a way of better capturing meanings.⁸⁸ Broadly, it was meant to help uncover the "feeling-tones" of concepts as a part of their total meaning.⁸⁹ For this purpose, Osgood asked participants to rate spontaneously and as quickly as possible a stimulus for associations with a large number of polar adjectival pairs, scaled on either numerical or verbal anchors—for example, extremely cold, quite cold, neither cold nor hot, quite hot, extremely hot. The participants received the following instructions:

*"The purpose of this study is to measure the meanings of certain words of various people by having them judge each word against a series of descriptive scales. In taking this test, please judge the words on the basis of what they mean to you. Each numbered item presents a concept and a scale. You are to rate the concept on the 7-point scale [indicated for each side: very closely associated, quite closely associated, only slightly related, or both sides equally associated]."*⁹⁰

Out of the ratings, any given concept may be mapped in a semantic space, which captures the global assessment of the concept along one or more axes. Osgood contextualized this type of research in the study of metaphor. He held that the "relation of this phenomenon to ordi-

84 S. TAKAHASHI, Aesthetic Properties of Pictorial Perception, in: *Psychological Review* 102, 1995, pp. 671–683.

85 How these terms were chosen is not described in the paper.

86 TAKAHASHI, Aesthetic Properties (cit. n. 84), p. 675. For interest in this topic from the perspective of design, see for example T. Rieke, haptic VISUALS. Oberfläche und Struktur – Farbe und ihre Beziehung zur Tastwahrnehmung, Frammersbach 2008.

87 D. STAMATOPOULOU, Embodied Imagination: Emotional Expression Perception of Dynamic Line-Drawings Created by Artists, in: *Imagination, Cognition and Personality* 26, 2007, pp. 213–248.; D. STAMATOPOULOU, Perception of Emotional Expression in Line Drawings, in: *Hellenic Journal of Psychology* 5, 2008, pp. 117–146; D. STAMATOPOULOU/G. C. CUPCHIK/T. AMEMIYA/M. HILSCHER/T. MIYAHARA, A Background Layer in Aesthetic Experience: Cross-cultural Affective Symbolism, in: *Japanese Psychological Research* 58, 2016, pp. 233–247.

88 C. E. OSGOOD/J. SUCI/P. H. TANNENBAUM, The Measurement of Meaning, Urbana 1957.

89 C. E. OSGOOD, On the Whys and Wherefores of E, P, and A, in: *Journal of Personality and Social Psychology*, 12/3, 1969, pp. 194–199, esp. p. 195.

90 C. E. OSGOOD/J. SUCI, Factor Analysis of Meaning, *Journal of Experimental Psychology* 50/5, 1955, pp. 325–338, here p. 328.

nary metaphor is evident: a happy man is said to feel 'high', a sad man 'low'; the pianist travels 'up' and 'down' the scale from treble to bass; souls travel 'up' to the good place and 'down' to the bad place; hope is 'white' and despair is 'black'.⁹¹ Although he did not develop the semantic differential for the empirical study of art, he acknowledged that the "notion of using polar adjectives to define the termini of semantic dimensions grew out of research on synesthesia,"⁹² especially on color-music synesthesia. In fact, the semantic differential was repeatedly used for research on art and Osgood, for his part, defined art as "a kind of human communication [that] is a perfectly legitimate area of scientific study."⁹³

As a result of the experiments, Osgood found a reduction of adjective scales into three independent groups or factors (calculated with statistical advancements dealing with multidimensional scaling for factor reduction).⁹⁴ These are: "Evaluation", represented by scales such as good-bad, pleasant-unpleasant, and positive-negative; "Potency", with terms such as strong-weak, heavy-light, and hard-soft; and, "Activity", which covers fast-slow, active-passive, excitable-calm.⁹⁵ William Tucker, a student of Osgood, found that art appreciation required a different set of words than Osgood had used in his first studies. He derived a set of 40 adjectives from a pre-study with museum visitors and persons who saw slides of artworks. In his subsequent experiments, he asked artists and non-artists to

Table 7
FACTOR LOADINGS FOR NON-ARTISTS ON SEVEN REPRESENTATIONAL PAINTINGS
(TUCKER STUDY)

Scale	Factor I (Activity)	Factor II (Evaluation)	Factor III (Potency)
hot-cold	.64	-.08	.00
pleasant-unpleasant	-.02	.59	-.60
lush-austere	.64	-.16	-.23
vibrant-still	.91	-.08	.29
repetitive-varied	-.81	-.48	.29
happy-sad	.34	.38	-.71
chaotic-ordered	.55	-.84	.00
smooth-rough	-.57	.53	.00
superficial-profound	.18	-.72	-.58
passive-active	-1.00	.00	.00
blatant-muted	.80	-.28	.11
meaningless-meaningful	-.33	-.79	.28
simple-complex	-.66	.55	-.48
relaxed-tense	-.57	.39	-.54
obvious-subtle	-.23	.80	.01
serious-humorous	-.22	-.05	.97
violent-gentle	.41	-.37	.69
sweet-bitter	-.32	.23	-.67
static-dynamic	-.78	.19	-.53
clear-hazy	-.04	.88	.38
unique-commonplace	.50	.22	.72
emotional-rational	.67	.09	.40
ugly-beautiful	.12	-.51	.42
dull-sharp	-.53	-.34	-.74
sincere-insincere	.18	.80	.34
rich-thin	.56	.35	.46
bad-good	-.33	-.77	-.27
intimate-remote	.09	.45	-.46
masculine-feminine	.31	.13	.76
vague-precise	-.04	-.84	-.43
ferocious-peaceful	.39	-.46	.83
soft-hard	-.39	.09	-.84
usual-unusual	-.52	-.16	-.70
controlled-accidental	.00	.80	.34
wet-dry	-.37	-.89	.35
strong-weak	.37	.46	.81
stale-fresh	-.45	-.54	-.51
formal-informal	-.38	-.40	.24
calming-exciting	-.54	.26	-.55
full-empty	.60	.31	.52

Fig. 4: The list of terms used in the semantic differential by TUCKER 1955

rate abstract and representational contemporary artworks on this adjective list (Fig. 4).⁹⁶ Tucker's list overlaps with the terms discussed in section two of this paper.⁹⁷

At the outset of its development, the semantic differential was reflected in art historical dis-

91 Ibid., p. 21.

92 OSGOOD/SUCL/TANNENBAUM, *The Measurement of Meaning* (cit. n. 88), pp. 20, 290.

93 Ibid., p. 291.

94 C. E. OSGOOD, *Semantic Differential Technique in the Comparative Study of Cultures*, in: *American Anthropologist*, New Series, 66/3, Part 2: Transcultural Studies in Cognition, June 1964, pp. 171–200.

95 Ibid., p. 173.

96 W. T. TUCKER, *Experiments in Aesthetic Communications*, unpub. doct. Diss. University of Illinois, USA, 1955.

97 Tucker's terms were: hot-cold, pleasant-unpleasant, lush-austere, vibrant-still, repetitive-varied, happy-sad, chaotic-ordered, smooth-rough, superficial-profound, passive-active, blatant-muted, meaningless-meaningful, simple-complex, relaxed-tense, obvious-subtle, serious-humorous, violent-gentle, sweet-bitter, static-dynamic, clear-hazy, unique-commonplace, emotional-rational, ugly-beautiful, dull-sharp, sincere-insincere, rich-thin, bad-good, intimate-remote, masculine-feminine, vague-precise, ferocious-peaceful, soft-hard, usual-unusual, controlled-accidental, wet-dry, strong-weak, stale-fresh, formal-informal, calming-exciting, full-empty.

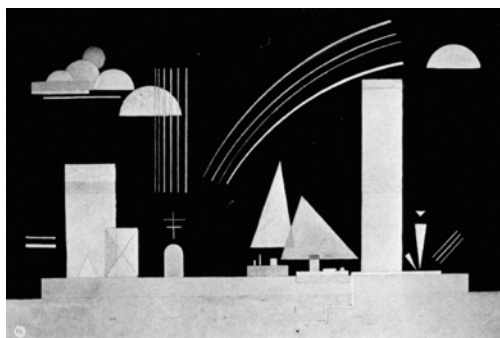


Fig. 5: Wassily Kandinsky, *Ruhe/At Rest*, 1928, New York, Solomon R. Guggenheim Museum, in GOMBRICH 1963

cussions. Ernst Gombrich, who underlined the relativity and hence context dependency of aesthetic effects, addressed Osgood's method in 1963: "I have discussed this natural equivalence between emotional states and sounds, colours and shapes in various contexts, and have drawn attention, in this connexion, to the experimental tool which C. E. Osgood has developed for their exploration."⁹⁸ In this context, Gombrich discussed Kandinsky's *At Rest* (fig. 5), a painting in which the painter had tried to create a peaceful and still impression by arranging the shapes and colors in a particular manner, at least according to the art historian Leopold Ettlinger.⁹⁹ Gombrich however, stated that he doubted that "even the most sensitive beholder would feel 'at rest' in

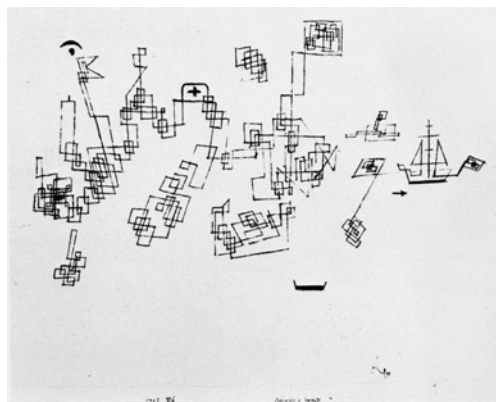


Fig. 6: Paul Klee, *Aktivität der Seestadt/Bustling Harbor*, 1927, Bern, Zentrum Paul Klee, in GOMBRICH 1963

front of this picture."¹⁰⁰ Nevertheless, he believed "that we can restore a clear expressive meaning to Kandinsky's composition by placing it in its historical context."¹⁰¹ He contrasted *At Rest* with Klee's drawing *Bustling Harbour* (fig. 6). In comparing the two pictures, the beholder might find that the overall effect of *Bustling Harbour* was "restlessness" and that *At Rest* then appeared calm.¹⁰² Gombrich's polemic was also a direct attack on slogans such as "abstract art as world language", popular after World War II, suggesting that abstract art is understandable to everyone and far more universal than figurative art.¹⁰³

98 E. H. GOMBRICH, *Expression and Communication*, in: Idem (ed.), *Meditations on a Hobby Horse: and other Essays on the Theory of Art*, London 1963, pp. 55–69, p. 58.

99 L. D. ETTLINGER, *Kandinsky's "At Rest"*, New York/Toronto 1961. *At Rest* was also the main topic of a lecture by Ettlinger (*Kandinsky and the psychology of perception*), quoted by Gombrich. Ettlinger makes a link to Humbert de Superville: "'At Rest' with its strong emphasis on the horizontal, would be a fitting illustration in any reprint of Superville's fascinating essay", p. 20.

100 GOMBRICH, *Expression and Communication* (cit. n. 98), p. 67.

101 Ibid., p. 68.

102 Ibid.

103 S. BANN, *Abstract art – a language?*, in: Tate Gallery (ed.), *Towards a New Art. Essays on the Background to Abstract Art 1910–20*, London 1980, pp. 125–231; A. HOORMANN, *Die Entfernung der Inhalte aus dem Bild. Zur Konstruktion des „unschuldigen Sehens“ im Abstrakten Expressionismus*, in: HOORMANN/D. BURDORF, *Medium und Material zur Kunst der Moderne und der Gegenwart*, München 2007, pp. 262–275; S. LEEB, *Abstraktion als internationale Sprache*, in: S. BARRON (ed.), *Kunst und kalter Krieg. Deutsche Positionen 1945–89*, exh. cat. Germanisches Nationalmuseum Nürnberg und Deutsches Historisches Museum Berlin, Köln 2009, pp. 119–133; and for an empirical approach to this question H. BRINKMANN/L. COMMARE/H. LEDER/R. ROSENBERG, *Abstract Art as a Universal Language*, in: *Leonardo* 47, 2014, pp. 256–257.

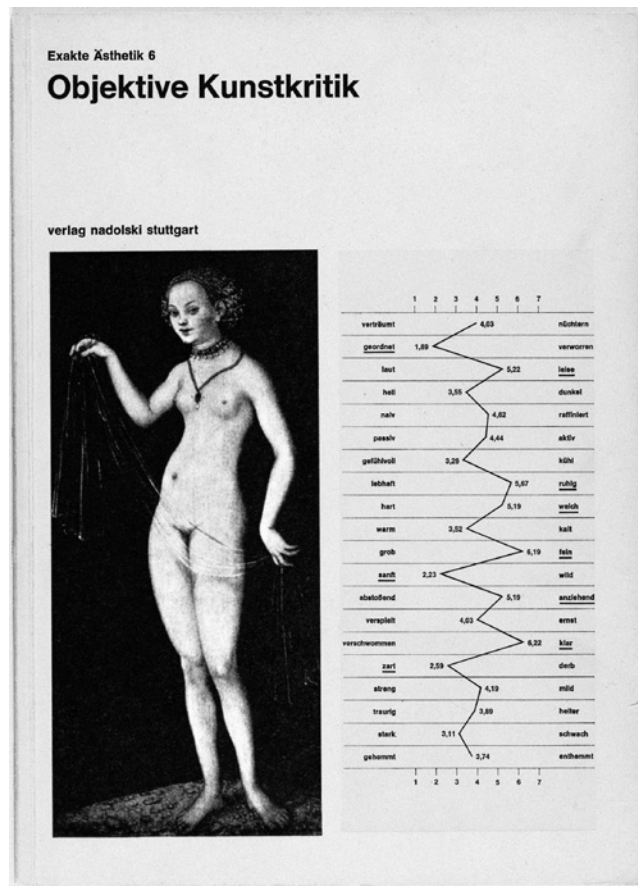


Fig. 7: Cover of the journal *Objektive Kunstkritik*, *Exakte Ästhetik* 6, 1969

The semantic differential was also used in a series of interdisciplinary aesthetic congresses organized by the psychologist William E. Simmat. These congresses were paired with the journal *Exakte Ästhetik*, edited by Simmat, who was also head of *Galerie d*, an association for modern art in Frankfurt. In 1966, the psychologist Peter Hofstätter published an article in this journal on the semantic differential as a method to capture “Anmutungsqualitäten”.¹⁰⁴ Three years lat-

er, in 1969, the journal focused on “objective art criticism” (“Objektive Kunstkritik”).¹⁰⁵ Simmat suggested to use the semantic differential for this purpose, an idea illustrated on the cover showing a *Venus* by Cranach next to a semantic differential (fig. 7).¹⁰⁶ He proposed that art appreciation requires a specific set of terms. The key challenge for him was to find a list of reasonable adjectives for this purpose. Ideally, these adjectives would be derived from art literature, but

104 P. R. HOFSTÄTTER, Objektive Methode zur Erfassung von Anmutungsqualitäten, in: W. E. SIMMAT (ed.), *Exakte Ästhetik* 3/4, Stuttgart 1966, pp. 47–65.

105 W. E. SIMMAT (ed.), *Exakte Ästhetik* 6: Objektive Kunstkritik, Stuttgart 1969.

106 W. E. SIMMAT, Das „semantic differential“ als Instrument der Kunstanalyse, in: SIMMAT, *Exakte Ästhetik* (cit. n. 105), pp. 69–88, esp. p. 69; and TUCKER, Experiments in Aesthetic Communications (cit. n. 96).

due to time limitations, he composed a list of 20 adjectives through pilot studies.¹⁰⁷ In his actual experiment, he asked 27 students to rate 10 artworks by different artists. The terms that the group selected were then presented to another group (without the paintings), where the subjects had to match the paintings to the terms.

This was thus a reversal of the process undertaken by the first group and all participants were able to identify the “correct” matches. Herewith, Simmat concluded that the semantic differential was an adequate tool for measuring aesthetic qualities of art.¹⁰⁸

5. AN EXPERIMENTAL STUDY AND OUTLOOK ON FUTURE RESEARCH

The review sketched in this paper demonstrates that many authors of art literature and art history on the one side and of psychology on the other side share three central assumptions: the importance of lines and colors as basis of aesthetic effects, their polar description, and the often tacit assumption of a high degree of universality of such effects. For instance, Goethe did not consider the possibility of individual differences, Humbert de Superville claims in the ti-

tle of his book to reveal the “unconditional signs in art” and Fechner assumed that at least “western” humans collectively share specific object associations with the color yellow.¹⁰⁹ Authors from the arts and psychologists came together at several moments of history. However, up until now, they remain far apart from each other. With some exceptions, as discussed in this text, art historians generally do not consider the use of empirical methods to objectify their assump-

	Bipolar (Semantic Differential) <i>cold-warm</i>
Unipolar Same (e. g. <i>cold</i>)	-0.84
Unipolar Different (e. g. <i>warm</i>)	0.85
Selection Same (e. g. <i>cold</i>)	-0.75
Selection Different (e. g. <i>warm</i>)	0.77

Table 1: Correlations between the bipolar (semantic differential) and the other scales

Note: Correlation is a measure of the relationship between two variables. “Same” in the unipolar and in the selection method means that if the bipolar scale goes from cold to warm (i. e., a low rating reflects the idea that the painting is cold) then “unipolar same” refers to the rating on the unipolar scale of cold (not cold – cold) and “selection same” refers to the selection of the word “cold” as relevant to describe the effect of the painting. In contrast, “different” refers to the rating on the unipolar scale of warm (not warm – warm) and “selection different” refers to the selection of the word “warm” as relevant to describe the effect of the painting. The table thus shows the expected pattern for consistent ratings with high negative correlations between the bipolar scale, the “unipolar same” and the “selection same” on the one side and high positive correlations between “unipolar different” and “selection different” on the other side.

107 In German: Verträumt-nüchtern, geordnet-verworren, laut-leise, hell-dunkel, naiv-raffiniert, passiv-aktiv, gefühlvoll-kühl, lebhaft-ruhig, hart-weich, warm-kalt, grob-fein, sanft-wild, abstoßend-anziehend, verspielt-ernst, verschwommen-klar, zart-derb, streng-mild, traurig-heiter, stark-schwach, gehemmt-enthemmt.

108 Ibid., p. 77.

109 FECHNER, Vorschule der Ästhetik (cit. n. 24).



Fig. 8: Wassily Kandinsky, *Untitled*, 1934, watercolor, ink, 31,6 x 24,6 cm, Paris, Centre Georges Pompidou

tions, and psychologists generally do not take history into account.

The central aim of our project, which joins psychologists and art historians, is to question the claim of universality of aesthetic effects. We use empirical methods from psychology to test assumptions from art literature, that are often used until today in art history. Our first step was to establish a suitable instrument to measure aesthetic effects. Due to the long tradition of polar description of aesthetic effects and to its acceptance in psychology, the semantic differential described in section four was our natural choice. To evaluate this choice, we compared the semantic differential with two other methods: a unipolar scale where every term was rated separately (i. e.



Fig. 9: Paul Klee, *Blick aus Rot / Be aware of red*, 1937, pastel on cotton on wallpaper glue on jute, 47 x 50 cm, Bern, Zentrum Paul Klee, donation Livia Klee

one rating for “cold” and another for “warm”), and a selection method, in which participants could mark every term they found appropriate to describe the artwork. Based on the analysis of historical texts described in section two of this paper, we composed a list of 20 adjective pairs of terms that occurred most frequently among the authors examined so far.¹¹⁰ As stimulus material, we chose 10 abstract pictures.¹¹¹ In total, 118 people participated in our study (96 female, mean age=21.36, SD age= 3.7). We divided them in three groups and asked each of them to rate 10 artworks, but only on one of the three scales: 38 people rated the artworks on the semantic differential, 37 on the unipolar scales and 43 used the free selection method.

The comparison of the three groups shows that the type of scale did not affect the choice of terms in the ratings. In fact, the three methods strongly correlate (Table 1). This indicates

110 These are (in German): positiv – negativ, aktiv – passiv, lebhaft – ruhig, spannungsvoll – locker, beruhigt – aufgeregt, statisch – dynamisch, aggressiv – friedlich, weich – hart, stark – schwach, heiß – kalt, schwer – leicht, sanft – grob, körperlich – geistig; maskulin – feminin, aufdringlich – zurückhaltend, protzig-bescheiden, harmonisch – disharmonisch, ausgewogen – unausgewogen.

111 Two works from each of the following artists: Wassily Kandinsky, Paul Klee, Fritz Winter, Joan Miró and Richard Mortensen.

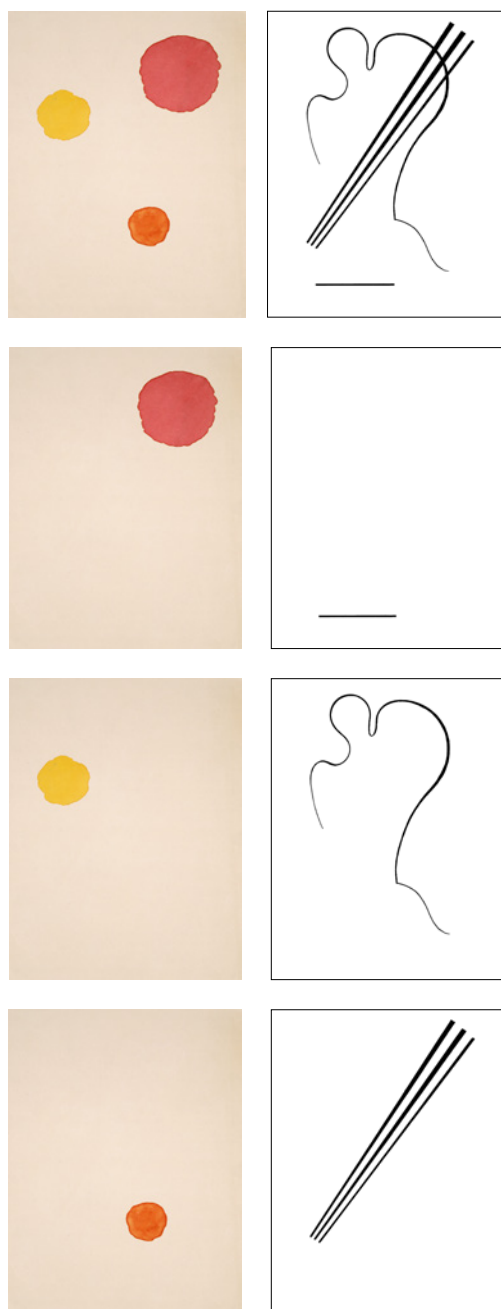


Fig. 10: Separation of lines and color of Kandinsky's watercolor

that if a painting is rated warm rather than cold on a semantic differential, it will also be rated warm and not-cold on the unipolar scales. Moreover, we found a high correlation ($r = -.81$) be-

tween the ratings of the unipolar scales. Specifically, if people rated cold and warm on two different unipolar scales, these unipolar scales together result in one bipolar scale. The semantic differential is based on the binary opposition of terms and forces participants to rate their impression of any given stimulus on a scale. This raises the question how the midpoint of the scale can best be interpreted as it can mean 'neither' and/or 'both equally'. The result of our comparison shows that when people rate a painting on the midpoint of a bipolar scale, this is because they find neither of the terms relevant (rather than finding both equally relevant). This interpretation of the results is strengthened by the low occurrence of the selection of two opposing terms in the selection method. In total, people selected two opposites only in 3.1% of all cases. Thus, we can conclude that all three scales function similarly and produce similar and comparable findings.

A closer look at the ratings of an untitled watercolor by Kandinsky (fig. 8) and Klee's *Be aware of red* (fig. 9) will clarify this point. On average, participants rated Kandinsky's work as positive, warm and feminine. Klee's painting was also rated warm and positive, but on average at the midpoint between feminine and masculine. Did participants find *Be aware of red* both feminine and masculine, or neither feminine nor masculine? The method-comparison showed that this work was perceived as not-feminine and not-masculine and, thus, confirmed that mid-scale ratings are used by participants when they want to indicate that both concepts are not suitable rather than indicating that both apply.

In the future, we will use these images to test to what extent the overall effect of paintings is a result of the sum of the effect of separable elements (lines and colors) and to what extent the claim of universality applies. In order to tackle these questions empirically, we disentangled three abstract pictures (by Kandinsky, fig. 8, Klee, fig. 9, and Miró) and isolated "their" single

lines and color patches (fig. 10). Kandinsky's watercolor shows a combination of lines with different qualities – there is a bundle of three diagonal lines and a curvy one. Is the horizontal line perceived as calm and still and the line bundle as dynamic? And what about the curvy line? Will it appear more feminine? Behind the black lines we detect three dots of color. Theoretically, these colors are all on the “warm” side of the spectrum. Similarly, Paul Klee's *Be aware of red* shows black lines and contains “warm” colors. A curved line above a dot on the right side of the picture reminds of a human eye and may be the reason for the German title of this work *Blick aus Rot*, literally *Gaze from red*. How do these various elements contribute to the overall impression of the artworks? Can we obtain information on the relations between these elements from analytical experiments with beholders?

Universität Wien

Acknowledgement

We would like to thank Stephanie Sailer for her assistance, Armin Plankensteiner, René Steyer, and Karl Pani for their support in the image editing process, Tanja Jenni, Nirmalie Mulloli and Klaus Speidel for revisions of the manuscript, and the WWTF for financing this research.

Image credits: Fig. 1: *Dekorative Kunst* I, 2, 1898, p. 121. – Fig. 2: from HARRISON/WOOD (eds.), *Art in Theory* (cit. n. 41), p. 65. – Fig. 3: from LUNDHOLM, *The Affective Tone of Lines* (cit. n. 75), p. 45. – Fig. 4: from OSGOOD/SUCI/TANNENBAUM, *The Measurement of Meaning* (cit. n. 88), p. 69. – Fig. 5 and Fig. 6: from GOMBRICH, *Expression and Communication* (cit. n. 98), figs. 33 and 34, Photo courtesy of Universität Wien, Institut für Kunstgeschichte. – Fig. 7: Cover of the journal *Exakte Ästhetik* 6, 1969, Photo courtesy of Universität Wien, Institut für Kunstgeschichte. – Fig. 8: Paris, Centre Georges Pompidou, MNAM-CCI, DIST. RMN-Grand Palais. – Fig. 9: Bern, Zentrum Paul Klee. – Fig. 10 and Fig. 11: Armin Plankensteiner, René Steyer, and Karl Pani, Institut für Kunstgeschichte, Universität Wien.

2.2.4

Boddy, J., Brinkmann, H., Forster, M., **Specker, E.**, Leder, H., & Rosenberg, R. (*manuscript in preparation*). Universality, sensitivity and measurability in discourses on the aesthetic effects of lines and colors: Knowledge transfer between art and psychology.

Summary:

This paper specifically deals with the notion of universality of aesthetic effects. It starts with a discussion of the discourse of universality. It shows that whereas in the eighteenth and early nineteenth century theorists like William Hogarth and David Pierre Humbert de Superville had looked for the “line of beauty” and “unconditional grammar of forms” in the realm of universals and ideals, by the end of the nineteenth century theoretical expositions came to be intertwined with research in “experimental” aesthetics, based on scientific experiment. The art historian Heinrich Wölfflin (1886), for example, explained that while the association of specific aesthetic effects with distinct lines and colors was based on his personal observation, it was vindicated on psychology literature (he referred to Wilhelm Wundt). Solid empirical evidence was never found, perhaps leading to the contention in psychology theory and practice that concrete statements about stimulus-response relations had no general validity, but it did not stop discussions about the aesthetic possibilities and cultural potential of abstract line and color in art. By 1900, an image of the psychological significance of line and color had been established in the minds of artists and theorists, and in the minds of their audience. From this the paper shows how a notion of sensitivity was introduced to account for individual variation. Sensitivity involved training through systematic evaluation of abstract lines and colors, which was grounded in an evolving sense that would make an abstract aesthetic system (such as Abstract Art) possible and was simultaneously required for attending to such an aesthetic system. The concept of aesthetic sensitivity remained popular throughout the 20th century. For example, the development of The Visual Aesthetic Sensitivity Test (VAST) by psychologists Hans Eysenck and Daniel Berlyne with the artist Karl Otto Götz. Nevertheless, a lack of scientific evidence supporting the assumed connection between a stimulus and a corresponding response in observers remained, leading Götz and his wife Karin to consider the hundred-year-old undertaking of empirical aesthetics as a failure (Götz & Götz, 1972). Finally, the paper gives a discussion of the findings of Specker et al. (*under review*, section 2.2.1) from an art historical perspective. This empirical study found sobering results for the notion of universality too. However, in the current paper the results are also discussed on an image-by-image level for proper art historical analysis though this too did not find any indications of high agreement.

My Contribution:

In this case the concept and idea underlying the manuscript was developed by HB, JB, and RR. My contribution lies in section 3 of the manuscript “Measuring agreement among observers today: an experimental approach”. This reports on findings of Specker et al. (*under review*; section 2.2.1). Though not responsible for this re-analysis of my data I did advise on the data interpretation along with MF and HL. In addition, I provided feedback on consecutive versions of the article by my co-authors.

Universality, sensitivity and measurability in discourses on the aesthetic effects of lines and colors

Knowledge transfer between art and psychology

Jane Boddy, Hanna Brinkmann, Michael Forster, Eva Specker, Helmut Leder, Raphael Rosenberg

Introduction

By the end of the nineteenth century, there was a longstanding tradition of writing about the aesthetic effects of specific colors, of straight and curved lines, shapes and the composition of light-and-dark. Theoretical expositions rested on the assumption that lines and colors are among the most important components of art.¹ Whereas in the eighteenth and early nineteenth century theorists like William Hogarth and David Pierre Humbert de Superville had looked for the »line of beauty« and »unconditional grammar of forms« in the realm of universals and ideals, by the end of the nineteenth century theoretical expositions came to be intertwined with research in »experimental aesthetics«, based on scientific experiment.² When Henry van de Velde, for example, declared in 1890 that » [...] soon the meaning of lines will be revealed to us«, he expected that scientific investigation would unravel the human response to line. This was not an isolated thought: artists, theorists, and art historians were resourceful in classifying line and color in psychological terms in theoretical expositions that were purportedly on the verge of laboratory problems.³

¹ Rosenberg 2007; Hanna Brinkmann/Jane Boddy/Beatrice Immelmann/Eva Specker/Matthew Pelowski/Helmut Leder/Raphael Rosenberg, Ferocious Colors and Peaceful Lines. Describing and Measuring Aesthetic Effects, in: *Wiener Jahrbuch für Kunstgeschichte* 65, 2018.

² The reference to »experimental aesthetics« comes from Fechner 1872.

³ The rise of experimental methodologies in psychology was an important impulse, but the manner in which artists and critics engaged with this research cannot be easily summarized: there were those who developed individual theories, some extended on experimental and empirical research, while others conducted experiments of their own. [→] Mark Micale (ed.), *The Mind of Modernism. Medicine, Psychology, and the Cultural Arts in Europe and America, 1880-1940*, Stanford, 2004. [→] Beatrice Immelmann etc. Kandinsky Questionnaire. [→] Jutta Müller-Tamm, Henning Schmidgen, Tobias Wilke (eds.), *Gefühl und Genauigkeit. Empirische Ästhetik um 1900*, München 2014. [→] Magdalena Bushart, »Form« und »Gestalt«. Zur Psychologisierung der Kunstgeschichte um 1900, in: O. Oexle (ed.), *Krise des Historismus – Krise der Wirklichkeit*, Göttingen 2007, 147-179. [→] Bernadette Collenberg-Plotnikov, Zoologen und Physiker als die berufensten Forscher in Sachen der Ästhetik? Zur Bestimmung der experimentellen Ästhetik in der Allgemeinen Kunstwissenschaft, in: *Zeitschrift für Ästhetik und allgemeine Kunstwissenschaft*, 58/1, 2013, 11-34. [→] Bitte auch prüfen, ob *Zeitschrift für Ästhetik und Allgemeine Kunstwissenschaft* Band 61. Heft 2 Berlin 1913 - Paris 1937. Ästhetik und Kunstwissenschaft im Zeitalter der Kongresse für dieses Argument (teilweise) relevant ist.

In this essay, we will assess the question whether there is an empirically identifiable basis for the claim of a »universal« aesthetics of line and color and thus emphasize the need for empirical evidence. Whereas the historical accounts often exhibit a distinctive notion of aesthetics effects that assumes a natural connection between line, color, and their impact on beholders, we emphasize instead its embeddedness in knowledge. Specifically, the aims of this essay are to explain and assess: (1) the link between different historical notions of »universal« aesthetics of line and color in art literature and the issue of solid empirical proof; (2) how differences between people were explained with the help of a notion of aesthetic sensitivity; and, (3) the possibility of measurement in experiments.⁴

1. Universal aesthetic effects of lines and colors

Art theorists writing around 1900 formulated a distinctive understanding of aesthetic effects that grounded the ability to apprehend pure line and color in the human body. The implication was that since everyone has a body, everyone has the potential to perceive and differentiate aesthetic effects. The art historian Heinrich Wölfflin, for example, asserted in 1886 that the »abrupt thrusts of the zigzag immediately bring to mind a burning red, while a soft blue suggests a gentle, wavy line«.⁵ This was a clear statement about the distinctive aesthetic effects of types of lines on the human psyche. Wölfflin clarified the difference between the zigzag and the wavy line along cross-modal terms and offered a psychological explanation of »analogies of linear awareness« which he understood in reference to Wilhelm Wundt as »the connections that we perceive between sensations of the different senses«.⁶ This involved the entire human body – its structure, the muscles and the skin – and not just the sense of sight. Wölfflin paired warm versus cold, burning versus soft, abrupt versus gentle and, to describe the difference between the zigzag and the wavy line, red versus blue.⁷ He explained that while the association of specific aesthetic effects with distinct lines and colors was based on his personal observation, it was supported by psychology literature. When Wölfflin went on to speak of warm and cold lines, he stated this as if it were a fact of perception rather than his

⁴ The paper relates the central results of the project »Universal aesthetics of lines and colors? Effects of culture, expertise, and habituation«, jointly conducted by art historians and psychologists from 2016 to 2019, and financed by the WWTF (CS15-036).

⁵ Heinrich Wölfflin, *Prolegomena to a Psychology of Architecture*, Michael Selzer (trans.), Colorado Springs 2017 [1886], 17. German original: »Das hastige Auffahren des Zickzacks führt unmittelbar die Erinnerung an brennendes Rot mit sich, während das sanfte Blau einer weichen Wellenlinie sich zugestellt.« Heinrich Wölfflin, *Prolegomena zu einer Psychologie der Architektur*, Ph.D. Dissertation: Universität München 1886, 14.

⁶ Wölfflin [1886] 2017, 17. German original: »Analogien der Linienempfindung« and »die Verwandtschaftsverhältnisse, die wir zwischen den Empfindungen disparater Sinne anzunehmen pflegen« (Wölfflin 1886, 13). The reference to Wundt is *Grundzüge der physiologischen Psychologie*, Leipzig 1874, vol 1, 486ff.

⁷ About the use of polar opposite terms to describe aesthetic effects, see Brinkmann et al. 2018.

individual impression.⁸ In his view, the ability to perceive aesthetic effects was ultimately a manner of opening up the appropriate channels of observation; once this was sufficiently developed in people, the difference between a zigzag and a wavy line would be immediately clear to everyone.

In the view of Wölfflin, every line and every color generates effects in spectators in one specific, pre-determined way and this was sufficient reason for attending to these aspects of art. At that time, art theorists often adopted a language of effect (*Wirkung*), evoking a cause-and-effect relation between stimulus and response. The designer and aesthetician August Endell famously stated about his artistic work, »The goal is aesthetic effect (*ästhetische Wirkung*), not narration or instruction about plants or animals«.⁹ Pure line and color were the source of such effects. Endell insisted, »The material of the artist consists of form and color. Like all things that become conscious, both induce specific feelings in us, indeed, *every color and every form cause the same feeling in all people*«.¹⁰ Nevertheless, the precise »code« for the link between line, color and feeling was still imperfectly understood. In an effort to formulate this code, Endell stipulated the existence of a mental capacity in humans responsible for the ability to apprehend the emotive effects of line and color, independent from figurative depiction, knowledge and culture.¹¹ A central question in literature at that time was whether aesthetic effects are the result of immediate sensation, or are rather embedded in conceptual systems of understanding (such as memories and associations).¹² According to Endell, aesthetic effects are the result of lines and colors working directly on the human sensorium. In particular, he thought that feelings (i.e., aesthetic effects) are the outcome of two facets working on the nerves: tension and tempo of perception.¹³ He utilized an argument about the physiology of eye-movements to justify that the every line produces one distinct feeling in all people, explaining that »[w]hen several people observe the same form, and really give themselves over to it, the accompanying feeling is the same for everyone«.¹⁴

⁸ Wölfflin 1886, 14. »Ebenso springt man von warmen und kalten Linien, von den warmen Linien des Holzschnitts z.B. und den kalten des Stahlstichs, Gegensätze die sich wiederum decken mit den Druckempfindungen: hart und weich«.

⁹ August Endell, Architektonische Erstlinge, in: *Dekorative Kunst* 6, 1900, 297–318, here 315. »Der Zweck ist eben ästhetische Wirkung, nicht Erzählung oder Belehrung über Pflanzen und Tiere«.

¹⁰ August Endell. *Um die Schönheit. Eine Paraphrase über die Münchener Kunstausstellungen 1896* (München, 1896), 8. »Das Material des bildenden Künstlers ist Form und Farbe. Beide rufen wie alle Dingen, die bewusst werden, in uns bestimmte Gefühle hervor und zwar ruft jede Farbe und jede Form bei allen Menschen das gleiche Gefühl wach«. Emphasis added.

¹¹ Endell 1896, 29.

¹² REF

¹³ August Endell, Formenschönheit und dekorative Kunst. II. Die gerade Linie, *Dekorative Kunst* 2, 1898, 119–121, here 120.

¹⁴ Endell 1896, 8. »Aber man hat die Gleichartigkeit der Wirkung auf den Menschen bezweifelt und darauf hingewiesen, dass ein und dasselbe Bild auf verschiedene Menschen auch einen total verschiedenen Eindruck mache. Das ist zweifellos, aber in diesen Menschen geht durchaus nicht das dasselbe vor, sie sehen gar nicht dasselbe. ... Wenn aber mehrere dieselbe Form betrachten, und sich ihr wirklich hingeben, dann ist das begleitende Gefühl auch bei allen dasselbe«.

Different art theorists and artists named and described aesthetic effects of lines and colors. A commonality was that people subscribed to psychological explanations of aesthetic effects, explaining it as a function of human perception. Theories of *Einfühlung* or empathy offered a philosophical-psychological underpinning for this idea. What these theories did not offer, however, was solid empirical evidence. In the years between 1890 and 1905 various theories of empathy circulated.¹⁵ Overall, such theories based on generalizations of personal, introspective evidence. This rested on the assumption that »universal« laws could be extrapolated from one's own experience.¹⁶ The major proponents of empathy – Theodor Lipps, Johannes Volkelt, Karl Groos, Stephan Witasek and Vernon Lee – relied on introspection, not laboratory experimentation.¹⁷ Yet, as they came to different observations, the problem emerged how diverging observations might be unified in a universally applicable doctrine.¹⁸ There were many different explanations for empathy, but without evidence from numerous people, there was no clear ground for favoring one theory over another.¹⁹ One solution for ensuring that theorists were not merely describing subjective experiences was to replace introspective evidence by data of larger numbers of people. There were also rigorous experiments carried out in this domain, but these experiments – such as by Oswald Külpe, Jacob Segal and Emma von Ritoók – either failed to give evidence for empathy or opened up new psychological problems altogether.²⁰ The psychologist Edward Bullough addressed this issue as he pointed out that Segal had found how »[t]he same subject found oblique straight lines sometimes pleasant and sometimes unpleasant, occasionally on one and the same day.«²¹ In fact, the lack of experimental evidence was a source of impatience. At the Congress for experimental psychology, held in Innsbruck in 1910, a respondent to Moritz Geiger's presentation made the following remark:

¹⁵ Harry Francis Mallgrave/ Eleftherios Ikononou (eds.), *Empathy, Form and Space: Problems in Modern Aesthetics, 1873–1893*, Santa Monica, 1994.

¹⁶ Juliet Koss makes this point in On the Limits of Empathy, in: *The Art Bulletin* 88.1, 2006, 139–157.

¹⁷ → Check

¹⁸ John Fizer, *Psychologism and Psychoaesthetics. A History and Critical View of Their Relations*, Amsterdam, 1981, 51–52. Juliet Koss, *Modernism after Wagner*, Minneapolis/London, 2010, 77–83.

¹⁹ Cf. Theodor Lipps, Psychologie und Ästhetik, in: *Archiv für die gesamte Psychologie* 9, 1907, 91–116, here 96–97.

²⁰ Oswald Külpe, Ein Beitrag zur Experimentellen Aesthetik, in: *The American Journal of Psychology* 14.3/4, 1903, 215–231, here 227. Jacob Segal, Beiträge zur experimentellen Aesthetik. I. Über die Wohlgefälligkeit einfacher räumlicher Formen. Eine psychologisch-ästhetische Untersuchung, in: *Archiv für die gesamte Psychologie* 7, 1906, 53–124. Emma von Ritoók, Zur Analyse der ästhetischen Wirkung auf Grund der Methode der Zeitvariation, in: *Zeitschrift für Ästhetik und allgemeine Kunstwissenschaft* 5, 1910, 356–407, 512–544, here 544.

²¹ Edward Bullough, Recent Work in Experimental Aesthetics, *The British Journal of Psychology* 12.1, 1921, 76–99, here 93. Cf. Segal 1906: 100. The subject Bullough refers to is the psychologist Ernst Meumann.

»When I arrived, I expected to hear something about experiments in the field of empathy. But what have I heard except plenty of old theories. No experimental results. But this is no philosophical association. I believe the time is ripe for those who want to present these kinds of theories, to show us whether or not they can be confirmed by experiments«. ²²

What was still missing was solid evidence from scientific psychological aesthetic investigation. The lack of evidence perhaps led to the contention in psychology theory and practice that concrete statements about stimulus-response relations had no general validity, but this did not stop discussions about the aesthetic possibilities and cultural potential of abstract line and color in art. By 1900, an image of the psychological significance of line and color had been established in the minds of artists and theorists, and in the minds of their audience. Against the background of the, albeit sparse, empirical evidence against the »universality« of experience, however, the question arises how the claim for »universality« was salvaged. One possibility was to claim that natural associations between lines and colors (or »universal« aesthetic effects) rest in everybody, but people need a certain level of sensitivity to be able to »see« them. »Aesthetic sensitivity« was therefore an important concept for accounting for individual differences between people's responsiveness to pure line and color. In the next section, we will examine how it helped deal with variation between people without giving up a universalism and what doubts and questions emerge about the »universality« claim from this notion.

2. Aesthetic sensitivity for line and color

Artists and theorists developed different notions of »aesthetic sensitivity«. ²³ The general idea can be described as follows: it named the ability to perceive, differentiate and name aesthetic effects, argued to be the fundament of any aesthetic comprehension. Whereas *potentially* everyone could perceive aesthetic effects, in practice this only happened if one had sufficiently developed his or her sensitivity.

A developmental aspect was thus central to the idea of sensitivity. In fact, theorists formulated a distinct understanding of an evolving sense for apprehending abstract line and color that drew from psychology theory and practice. The Spanish born American philosopher George Santayana outlined an aesthetic theory that incorporated findings of psychological investigation in

²² Moritz Geiger, Über das Wesen und die Bedeutung der Einfühlung, in: *Bericht über den IV. Kongress für experimentelle Psychologie*, Leipzig: 1911, 29–73, here 66. »Als ich hierher kam, habe ich erwartet, etwas zu hören über die Experimente in dem Gebiet der Einfühlung. Aber was habe ich eigentlich gehört – lauter alte, uralte Theorien. Nichts von Experimenten in diesem Gebiet. Dies ist keine philosophische Gesellschaft. Es schien mir, dass die Zeit gekommen ist, dass derjenige, welcher solche Theorien hierher bringen will, zeigen sollte, ob sie durch Experimente bestätigt sind«. Translation cited in Dan Zahavi, Empathy and Other-Directed Intentionality, in: *Topoi* 33, 2014, 129–142, here 129.

²³ ABOUT THE TERM

The Sense of Beauty (1896), especially leaning on the idea that aesthetic effects result from sensory stimulation working on the nerves.²⁴ Santayana drew consequences for how variation in nervous stimulation conditioned analogies between different modalities of sensory perception, resulting in affinities between aesthetic effects across the senses.

»These affinities escape many minds; but it is conceivable that the sense of them should be improved by accident or training. There are certain effects of colour which give all men pleasure, and others which jar, almost like a musical discord. A more general development of this sensibility would make possible a new abstract art, an art that should deal with colours as music does with sound«.²⁵

Color sensitivity was not an invention of Santayana; it was theorized in psychology literature as the ability in humans to perceive color harmonies and to discriminate colors.²⁶ Nevertheless, interesting about Santayana's account is that it described the possibility of building up an aesthetic system of »abstract art« by training long neglected faculties. On Santayana's admission, »We have not studied these effects, however, with enough attention, we have not allowed them to penetrate enough into the soul, to think them very significant«.²⁷

Despite such doubts, by the beginning of the twentieth century, the significance of pure form and color had long been established in the minds of artists, critics and their audience. Some brief examples will help clarify that this was not just a matter of physiological nerves stimulation, but rather involved the cognitive powers of the mind. According to Oscar Schmitz, for example, aesthetic pleasure was not only conditioned by nervous stimulation and bodily sensations, but also by the extent to which the rhythm of line, color and tone corresponded to someone's personality.²⁸ »Wir suchen eigene Nervenreizen, aus deren sinnlicher Ordnung wir gleichsam, wie aus Symbolen, ein tieferes, unbedingteres Leben ahnen«.²⁹ Schmitz viewed the emotional impact of Arnold Böcklin's painting *Villa am Meer* independently from narrative or representation: »Die Schicksale berühren uns nicht, nur die Linie oder der Ton. Wir haben nie gefragt, was die verhüllte Frau vor der Villa am Meer bewegen mag. Doch fühlen wir die Wehmuth ihrer Linie und die endlose Trauer der Landschaft«.³⁰ The lines in Böcklin's painting make such a strong impression, he thought,

²⁴ George Santayana, *The Sense of Beauty. Being the outlines of aesthetic theory*, New York/Chicago/Boston 1896. For a discussion of Santayana in the context of the history of abstract visual art, see: Sabine Flach, *Die Wissenskünste der Avantgarden. Kunst, Wahrnehmungswissenschaft und Medien 1915-1930*, Bielefeld 2016, 56–57.

²⁵ Santayana 1896, 74–75.

²⁶ [→] Grant Allen

²⁷ Santayana 1896, 75.

²⁸ Oscar A.H. Schmitz, Über das Empfinden der Landschaft, in: *Wiener Rundschau* 1, 255–260, here 255.

²⁹ Ibidem, 258.

³⁰ Ibidem.

because their rhythm corresponded to the »spiritual rhythm« of a »feine Race von Sensitiven«.³¹ Schmitz thus appears to link the sense for abstract line and color to the question of ethnicity. In his view, the outer material had not changed objectively, but people had deepened their sensitivity towards their surroundings, making them more receptive to the emotive power of pure line and color. This was itself not an isolated thought; however, not everyone linked this to questions of ethnicity. Anton Lindner viewed a sensitivity for line and color as an instinct that connected people to the Absolute. In the essay »Farbe und Linie« from 1900 he argued that lines and colors have the tendency to behave in space, similar to all other forms of life.³² He thus viewed line and color through the lens of biological thinking and connected it to metaphysical ambitions. Similarly, Karl Scheffler adopted a religious-spiritual tone when he wrote that a scientific understanding of the psychological impact of line and color might fill the void that Christianity left behind. »Christianity is dead, and a new form of universal concepts of the world, as this would have to emerge from the consequences of scientific research, is scarcely felt in its soft beginnings«.³³ As the examples above indicate, theorists opted for a reorientation on human feeling through a heightened sensitivity for line and color.

Could this prepare new abstract art? In the reception of Kandinsky's abstract painting, for example, commentators frequently addressed the state of an evolving color sense that was purportedly necessary for appreciating and understanding his work. On the one hand, this entailed discursive apologies from critics trying to make sense of new art forms that confronted audiences; on the other hand, it entailed scientifically oriented art criticism that questioned whether color sensitivity was a plausible basis for explaining Kandinsky's art. The psychologist Helge Lundholm, who met Kandinsky in Stockholm in 1916, carefully considered Kandinsky's stipulation of the abstract expression of emotions in painting in an art critical essay that was published in *Flamman* in 1917.³⁴ He had distilled this stipulation from *Über das geistige in der Kunst* (1912, *On the Spiritual in Art*). By comparing Kandinsky's theoretical statements with the experience that his work offered, Lundholm sought to put theory to the test of experiment. Specifically, on the basis of Kandinsky's

³¹ Ibidem, 255.

³² Anton Lindner, Farbe Und Linie. Ein Versuch, in: *Wiener Rundschau* 4 (1900), 2–6, 31–36, here 3.

³³ Karl Scheffler, Die Konventionen der Kunst (Aphoristisch), in: *Dekorative Kunst* 11, 1903, 300–311, 398–400, 438–439, here 305. »Das Christentum ist tot, und eine neue Form universaler Weltbegriffe, wie sie aus den Konsequenzen der naturwissenschaftlichen Forschungen hervorgehen müsste, ist kaum in leisen Anfängen zu spüren«.

³⁴ Helge Lundholm, Kandinsky, in: *Flamman. Tidskrift för modern konst* 5, 1917, n.p. Lundholm is especially known for his experimental work in psychology in the United States, see: Helge Lundholm, The affective tone of lines: experimental researches, in: *Psychological Review* 1921, 43–60. The reference for Lundholm meeting Kandinsky is from Vivian Endicott Barnett, *Kandinsky and Sweden*, Malmö/ Stockholm 1989: 41. Kandinsky was in Stockholm from December 1915 to March 1916. Margareta Tillberg, Kandinsky in Sweden - Malmö 1914 and Stockholm 1916, in: Hubert van der Berg etc. (eds.), *A cultural history of the avant-garde in the Nordic countries*, Amsterdam/New York 2012, 325–336, here 331, 333.

Composition VI (1913, fig. 1), he examined whether color was a means to transmit emotions from »soul to soul«, that is, between artists and public.³⁵

Composition VI had been on show at the *Baltiska Utställningen* (Baltic exhibition) in Malmö in 1914, where it was presented as »ultramodern« art, and at the Gummesson Gallery in Stockholm in 1916, where Lundholm saw it.³⁶ The painting received considerable attention in the printed press; especially Kandinsky's argument that color and form express emotions, similar to tone and rhythm in music, was a topic of controversy.³⁷ Lundholm, too, examined *Composition VI* as painted music. By the time of his essay, the music-color analogy must have been a familiar trope in discussions of abstract art in Sweden. Interesting about Lundholm's treatment of the topic is that he questioned whether people's color sensitivity was enough developed to understand an aesthetic system of abstract visual art, on a par with music. In fact, he did not think that the time was ripe. »[P]urely psycho-physiologically, I doubt that there is a somewhat complete connection between the emotional state and the color tone, and even if there is such a connection between the emotional state and the color«. Lundholm continued, »I doubt strongly that even the most artistically educated observer will be sensitive enough, to let the color tones in him evoke the state of the soul, that scientifically corresponds to the chord«.³⁸ Lundholm further drew on the analogy with music as he explained that music had needed a long time to develop into an abstract aesthetic system: »instrumental music needed thousands of years to evolve from the primitive human beating a drum, to its present high level«.³⁹ He did not doubt that there existed correspondences between sensory perception of color and emotional states, but he wondered whether people were sufficiently receptive, at least for the time being. Consequently, he considered Kandinsky premature in claiming to have grasped universally applicable correspondences between forms, colors and emotions, and using this as a basis for a new abstract art.

³⁵ Lundholm 1917, n.p. Lundholm distilled this from Kandinsky's theory. He explained: »The color and the shape become a sign language between the artist and the observer, a sign language whose sole criteria of perfection is the ability to convey a feeling from soul to soul«. Translation by Maria Billinger.

³⁶ Hans Roethel/Jean Benjamin (eds.), *Kandinsky. Werkeverzeichnis der Ölgemälde. Band I, 1900-1915*, München, 1982, 461, cat. nr. 464. Endicott Barnett 1989: 57. The reference for »ultramodern« is from Bengt Lärkner, 'Humbug' or 'genius'? When Kandinsky came to Sweden, in: Anders Rosdahl (ed.), *Nya perspektiv på Kandinsky. New perspectives on Kandinsky*, Malmö 1990, 23–35, here 27.

³⁷ For a discussion of its reception in Sweden, see Lärkner 1990. Kandinsky was receptive to reactions, as he responded with the short essay »Konsten utan ämne« (Art without subject), reemphasizing the analogy of abstract art and music. Wassily Kandinsky, Art without Subject, in: Kenneth Lindsay/Peter Vergo (eds.), *Kandinsky. Complete Writings on Art. Volume One (1901-1921)*, London 1982 [1916], 419–420. Cf. Wassily Kandinsky, On the Artist, in: Lindsay/Vergo 1982 [1916], 407–418, here 416.

³⁸ Lundholm, 1917, n. p. Translation: Maria Billinger.

³⁹ Ibidem.



1 Wassily Kandinsky, *Composition VI*, March 1913, oil on canvas, 195x300 cm, State Hermitage, St. Petersburg. [photo/image]

Color sensitivity was an aspect of perception that could be exercised; it helped explain that every human being has the potential to develop their sensorial experience and deepen their observations. A similar understanding of sensitivity was formulated for lines. According to Endell, for example, there was still a large number of people insensitive to aesthetic effects, but, fortunately, the ability to perceive them was a matter of »firm education of vision for the effects of form and color« and could therefore be amended.⁴⁰ By learning to detect differences between lines and perceive contrasts between them – for example between a wavy and a zigzag line – artists progressively sharpen their sensitivity and learn to be affected by more and more elements in their surroundings.⁴¹ For this, a »systematic knowledge of forms and colors« was crucial.⁴² He explained: »Art is exclusively labor and requires only comprehensive, detailed knowledge of artistic effects, it demands unconditional, passionate devotion and awe. Art is not the product of erratic genius, let

⁴⁰ August Endell, Originalität und Tradition, in: *Deutsche Kunst und Dekoration* 9, 1901-1902, 289–296, here 295. »Kurz, eine sichere Schulung des Auges für Formen- und Farb-Wirkungen«.

⁴¹ Cf. Endell 1896, 12. »Indem sich nun unsere Sensibilität steigert, werden wir allmählich gewahr, dass die Charaktere des Sichtbaren in gewissen Gegensätzen zu einander stehen: das Anmutige steht dem Gewaltigen, das Heitere dem Erhabenen gegenüber«.

⁴² Endell 1901-1902, 294. »eine systematische Formen- und Farben-Kenntnis«

alone unexpectedly passing 'moods'.⁴³ It certainly does not rely on a doctrine of »male and female art«, the cosmos, the world creation or any other mystical achievement.⁴⁴ He continued, »we need clear, plain answers to the questions that the work brings us. How can we make a line hard, how can we make it soft, calm, gently, smooth, elegant, how can we make an ornament appear light or heavy«.⁴⁵

In an effort to improve sensitivity, Endell described an understanding of abstract line and color that contained a developmental account of perceptual experience: greater sensitivity for line and color meant that more aspects of the surroundings are perceived. The better people learn to differentiate line and color, and the more sensitive they become to differences, the greater their emotional responsiveness will become.

Sensitivity thus involved training through systematic evaluation of abstract lines and colors, which was grounded in an evolving sense that would make an abstract aesthetic system (such as Abstract Art) possible and was simultaneously required for attending to such an aesthetic system.⁴⁶ This idea remained popular throughout the 20th century, especially in connection to abstract art. Making a leap to the mid-twentieth century, in cooperation with the psychologists Hans Eysenck and Daniel Berlyne, the artist Karl Otto Götz developed a test instrument designed to differentiate levels of aesthetic sensitivity in beholders.⁴⁷ The *Visual Aesthetic Sensitivity Test (VAST)* is based on an image comparison, in which beholders are asked to judge 52 pairs of black-and-white abstract images (fig. 2).⁴⁸ The assumption behind this test is that certain images are objectively more beautiful, balanced, harmonious etc. than others.⁴⁹ People with a high level of aesthetic sensitivity presumably make »correct« judgements about which of two similar images is aesthetically better.⁵⁰ Aesthetic sensitivity is thus the ability to detect difference between two contrasting images. In fact, the *VAST* it is built around pairs of abstract images. Each image pair consists of two variations

⁴³ Ibidem, 293–94. » Kunst ist ausschliesslich Arbeit und setzt einzig und allein eine vollständige detaillierte Kenntnis der künstlerischen Wirkung voraus, sie verlangt unbedingte leidenschaftliche Hingebung und Ehrfurcht. Sie ist aber nicht das Produkt fahrigem Genietums oder gar unverhofft hereinbrechender 'Stimmungen'.«

⁴⁴ Ibid.: 294. »von männlicher oder weiblicher Kunst«

⁴⁵ Ibidem, 294–95. »Wir brauchen vielmehr klare, nüchterne Antwort auf die Fragen, die die Arbeit uns bringt. Wie macht man eine Linie hart, wie macht man sie weich, ruhig, vornehm, glatt, elegant, wie lässt man ein Ornament leicht erscheinen oder wie macht man es schwer«.

⁴⁶ Cf. Stacy Hand, *Embodied Abstraction: Biomorphische Fantasy and Empathy Aesthetics in the Work of Hermann Obrist, August Endell and Their Followers*, Ph.D. Dissertation, Chicago 2008, 189. »It is exactly this desire to train and heighten the public's sensitivities, which took shape in strategies informed by scientific aesthetics, that opened paths for abstract art in the twentieth century«.

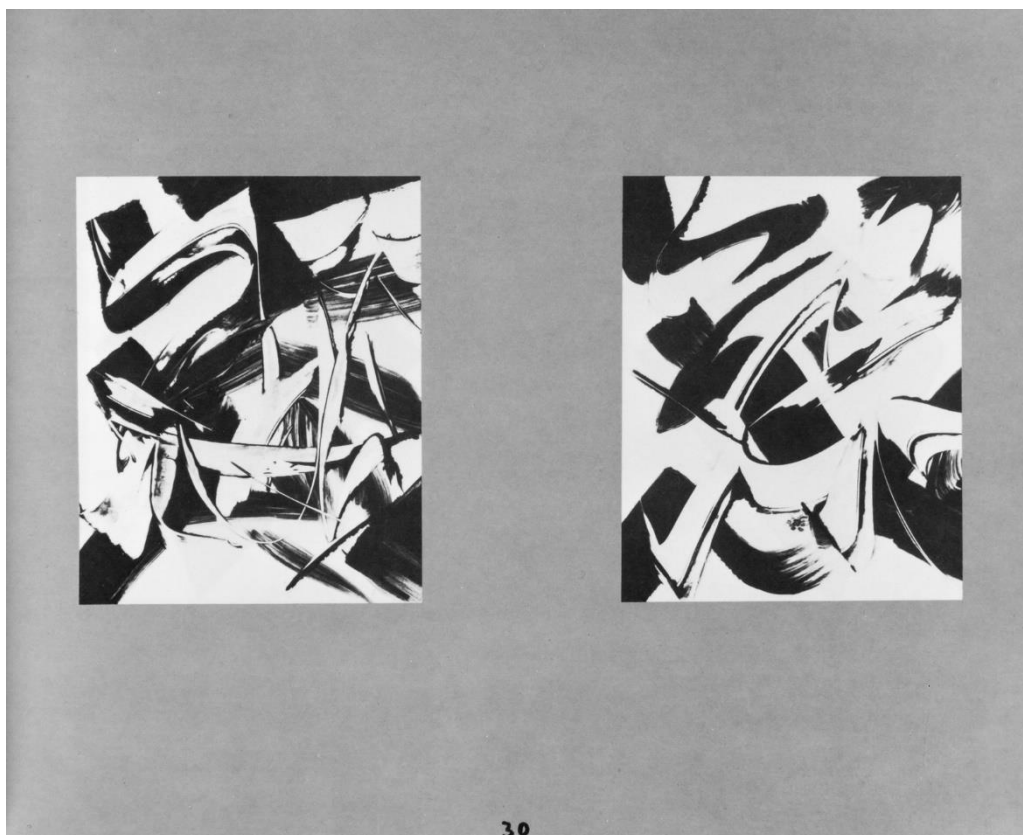
⁴⁷ VAST REFERENCE @ Hanna.

⁴⁸ Bamossy et al. 1985, 63. HIER VOLLE ANGABE, TITEL UND ORT FEHLEN @ Hanna.

⁴⁹ »Objectivity« here was established through expert ratings. @ Hanna.

⁵⁰ Again, the judgements are »correct« based on the expert ratings.

the same type. Götz designed the images himself; they hover at the edge of artistic images and psychological stimulus material. Crucially, within each pair one image is presumably aesthetically better than the other one. The task of the participants is to judge which one this is. In the instruction to the test, participants are explicitly discouraged from expressing their personal preference (preference judgements were and still are more common in experimental aesthetic investigation), but should instead accurately judge which image is better designed: »Do not form your judgement according to which of the two pictures you like better. This is not the point, because sometimes a picture that is less well designed may please you better, because you find it personally more interesting. Your task is to discover which picture has been better designed«.⁵¹



2. XXX

Thus, similar to Endell, for Götz sensitivity meant the ability to detect differences between contrasting entities, which could be exercised with the aid of pure line. Different from Endell, however, is that Götz worked with a notion of »correct« judgement about aesthetic characteristics, which he used to operationalize aesthetic sensitivity for measurement. A central aspect of Götz's

⁵¹ K.O. Götz/A.R. Borisy/R. Lynn/H.J. Eysenck, A new visual aesthetic sensitivity test: I. Construction and psychometric properties, in: *Perceptual and Motor Skills*, 49(3), 1979, 795–802. K. O. Götz, VAST: Visual aesthetic sensitivity test, (4th ed.). Dusseldorf 1985. K.O. Götz, Visual aesthetic sensitivity and intelligence, in: *Perceptual and Motor Skills*, 65(2), 1987, 422.

measure, then, is that it assumes an end-point: there is a pre-determined difference between the images and the participants may accurately perceive it, or fail to notice it.

Today, the *VAST* is still used in experimental aesthetic investigation.⁵² It is, as Fleck has noted, that aspect of Götz's scientific work that remains to receive attention in research.⁵³ Götz's scientific work was however not restricted to the *VAST*. He wrote several essays about art perception, the possibility of measurement and, importantly, expressed doubt about this possibility. In 1972, in a book that Götz wrote with his wife Karin, the authors pointed to the lack of scientific evidence supporting the assumed connection between a stimulus and a corresponding response in observers.⁵⁴ They took stock of empirical investigation in aesthetics – starting from Fechner up until 1972 – and came to a sobering conclusion: »Ihre Ergebnisse zeigen im Gegenteil, wie unterschiedlich Gefühlsreaktionen der Menschen vor Kunstwerken oder vor einfachen Form- und Farbelementen sind, ja wie unterschiedlich sogar Resultate derartiger Versuche ausfallen können, wenn sie wiederholt werden, und zwar mit denselben Menschen und mit demselben Stimulationsmaterial.«⁵⁵ In fact, they even went so far as to consider the hundred-year-old undertaking of empirical aesthetics as a failure.⁵⁶ Despite this disenchanted conclusion, Götz did not give up the idea of empirical research of art and developed the *VAST*. He remained to investigate the claim of »universality« empirically and, additionally, applied the *VAST* in different cross-cultural settings and varied with participants of different age groups.⁵⁷ In a study comparing children and adults from Japan and England, for example, Götz and colleagues [?] found no difference in visual aesthetics sensitivity [?] between participants from these countries.⁵⁸

Since then, psychologists developed various instruments that aim to differentiate aesthetic abilities of people. Generally, such instruments are premised on the assumption that aesthetic abilities are embedded in knowledge and training. The »Test of Aesthetic Judgment Ability«,

⁵² Nils Myszkowski/Martin Storme/Franck Zenasni/Todd Lubart, Is visual aesthetic sensitivity independent from intelligence, personality and creativity? 2013. ORT UND PUBL FEHLEN @ Hanna; Izard and Spelke 2009; Myszkowski and Storme 2017 VOLLE ANGABE @ Hanna.; Laura J. Summerfeldt/ J. Shaun Gilbert/Michael Reynolds, Incompleteness, aesthetic sensitivity, and the obsessive-compulsive need for symmetry, 2015. ORT UND PUBL FEHLEN @ Hanna.

⁵³ Fleck Rpbert KO Götz. West. Hundertwasser: 28–36. VOLLE ANGABE @ Hanna.

⁵⁴ Karl Otto Götz/Karin Götz, Probleme der Bildästhetik. Eine Einführung in die Grundlagen des anschaulichen Denkens, Düsseldorf 1972, 46.

⁵⁵ Ibid.

⁵⁶ Ibid., 47.

⁵⁷ @ Hanna. S. Iwawaki/H.J. Eysenck/K.O. Götz, A new Visual Aesthetic Sensitivity Test (VAST): II. Cross-cultural comparison between England and Japan, in: *Perceptual and Motor Skills*, 49(3), 1979, 859–862. J. Chan/H. J. Eysenck/K.O. Götz, A new visual aesthetic sensitivity test: III Cross-cultural comparison between Hong Kong children and adults, and English and Japanese samples, in: *Perceptual and Motor Skills* 1980, 1325–1326.

⁵⁸ Perceptual and Motor Skills 1980? 1325-26?? @ Hanna.

designed by Bamossy et al. (1983), measures »aesthetic ability in judging paintings« (using as stimuli images from Klee, Goya and Albright).⁵⁹ It rests on the assumption that aesthetic judgment develops over time and is learnable, in keeping with the discussion of aesthetic sensitivity around 1900. People with training in art or art history presumably have a higher sensitivity. In addition, the importance of formal training in art history for art appreciation was assessed in a recent study that investigated people's appreciation of Kitsch.⁶⁰ It showed that people with formal training are less likely to like Kitsch, compared to lay persons. In an effort to improve the differentiation of abilities of participants, which is necessary for such psychological experiments, the Vienna Art Interest and Art Knowledge Questionnaire (VAIAK, 2018) was introduced.⁶¹ It is a questionnaire designed to establish levels of interest in art and knowledge about art history and avoids a pre-determined distinction between formal training and no training. Thus, there exist an array of test instruments designed to differentiate aesthetic abilities. What these tests have in common is that they try to measure how accurate people are in making »correct« judgements about stimulus material, which makes it possible to differentiate abilities of beauty. What these tests are not designed to do, however, is measure how much people agree about aesthetic effects.

3. Measuring agreement among observers today: an experimental approach

Although the claim of »universality« results from controversial historical discourses, the idea that there should be agreement among people about the aesthetic effects of lines and colors is not only compelling, but also the basis of art historical description until now.⁶² Differentiating and naming aesthetic effects is an important skill pupils learn in art historical training.⁶³ Given its significance, it is highly relevant for art history whether concrete statements about aesthetic effects – such as, warm or cold colors and lines – are highly idiosyncratic and individual, or rather shared among observers. In an effort to assess this question, we conducted two empirical studies that test the assumption of »universality« by quantifying how much people agree in their description of aesthetic effects of lines, colors and abstract works of art.⁶⁴

First, and in keeping with the standards of psychological experimentation, we selected a relatively uniform set of three images: a watercolor and ink drawing by Kandinsky, an etching by

⁶⁰ Matthew 2019. The Kitsch Switch?—or (when) do experts dislike Kinkade Art? A study of time-based evaluation changes in top-down versus bottom-up assessment VOLLE ANGABE: @ Hanna

⁶¹ REF @ Hanna

⁶² REF @ Hanna

⁶³ REF @ Hanna

⁶⁴ For experimental work, see: Hanna Brinkmann, Laura Commare, Helmut Leder, and Raphael Rosenberg. Abstract Art as a Universal Language?, in: *Leonardo* 47.3 (2014), 256-257.

Joan Miró and a silk screen by Fritz Winter (fig. 3-5).⁶⁵ Relative uniformity means that the images are comparable on a visual level. Specifically, we chose artworks that (1) are abstract insofar that they contain no figurative elements; (2) consist of black lines and round-shaped colors, arranged as relatively discrete elements; and, (3) are works on paper with an original size between 50cm – 65cm that approximates the screen size of an [Z] computer.⁶⁶ These criteria allowed us to use artworks – without giving up control on our stimulus material – and to create new stimulus materials by digitally extracting discrete elements of these artworks (fig. 6).



3-5: XX

Specifically, we created new stimulus materials in order to show the test participants different types of colors and lines, which from the viewpoint of theory should have different characteristics. Kandinsky's watercolor is a particularly good case to illustrate the process of image extraction. In his image, he seems to have assembled the image by combining discrete elements in an additive manner. We, in turn, did the opposite: we extracted the elements from the image and disassembled it. We did not change the placement or size of the single elements, but isolated them.⁶⁷ The newly created images link nicely to statements that Kandinsky has made in his theoretical writings, especially as it is laid out in *Punkt und Linie zu Fläche* (1926, *Point and Line to Plane*). In this book, Kandinsky described the straight horizontal line as the simplest form of the line, which is associated with endless movement (fig. 6D).⁶⁸ Through the association of an endless movement, the character

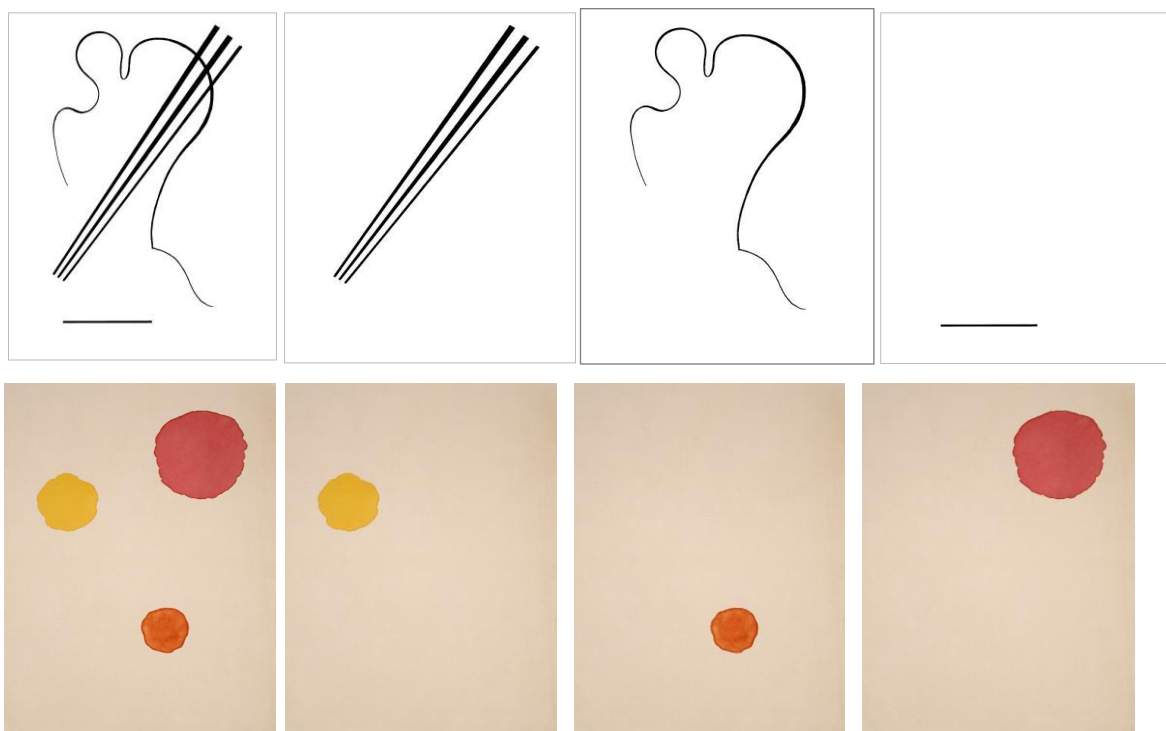
⁶⁵ Wassily Kandinsky, *Untitled*, 1934, watercolor and ink, 31.6 × 24.6 cm, Paris, Centre Georges Pompidou; Joan Miró, *Untitled*, 1961, color etching and aquatint, 31.7 × 24 cm, cat raisonnee: 292. Fritz Winter, *Siebdruck 6*, 1950, screen print, 50 × 65 cm.

⁶⁶ For the idea of disentangling lines and colors from artworks, see Jane Boddy, Disentangling the Aesthetic Effects of Formal Elements of Art, in: Véronique Alexandre Journeau/ Christiane Vial Kayser (eds.), *Notions Esthétiques. La perception sensible organisée*, Paris, 2015, 51-61. With the exception of one element in the Winter image that had a bow-like shape

⁶⁷ However, we did decide to alter the background color of the 'line' elements in order to ensure that all backgrounds were comparable.

⁶⁸ Wassily Kandinsky, *Point and Line to Plane*, New York 1947 [1926], 58.

of this line becomes calm and cold.⁶⁹ He linked the vertical line, by contrast, with warmth and activity. The free moving wavy line is described as temperamentvoll, energetic and with tension, the diagonal line is characterized as in between warm and cold.⁷⁰ Additionally, one could describe the diagonal line bundle, extracted from Kandinsky's drawing (fig. 6B) also as ascending or descending, depending on one's reading and writing direction. Kandinsky described this kind of diagonal as »nach oben«.⁷¹ Turning to the watercolor, in it Kandinsky combined these various elements in one image. The combination of all three line-types is also something Kandinsky elaborated in his writings. By contrasting different types of lines, he stated, their specific characteristics become particularly evident.⁷² In fact, the harmony of a composition is achieved by the highest contrast. Against the background of these considerations, the question emerged how observers today weight the balance between line and color in Kandinsky's image. Moreover, how do observers judge the lines and colors as discrete elements in and of themselves? In art historical literature, Kandinsky's watercolor has been described as having an explicitly graphical character, which might suggest that line dominates over color.⁷³ To what extend can such statements find support in empirical evidence?



6. A-H: XX

⁶⁹ REF @ Hanna.

⁷⁰ Beleg Seite@ Hanna.

⁷¹ @ Hanna. 156

⁷² REF@ Hanna.

⁷³ Georgina Illetschko, Kandinsky – Paris. Geschichte einer Beziehung, Bilder eines Jahrzehnts. PhD. Dissertation: Universität Wien, 1995: 161.

We selected descriptors that name aesthetic effects that are recurrent in art literature around 1900 (especially in Kandinsky's writings). Based on Brinkmann et al. (2018), we used the following terms: negative-positive, passive-active, lively-still, happy-sad, aggressive-peaceful, soft-hard, warm-cold, heavy-light, smooth-rough, bodily-spiritual, masculine-feminine, intrusive-cautious, dislike-like, and uninteresting-interesting.⁷⁴ These terms were presented to the participants as a bipolar semantic differential on a 7-point scale (ranging from X to Y). This scale captures trends in the direction no, low, middle and strong agreement.

We used »agreement« as a measure to operationalize »universality«. The question was to operationalize »agreement« for an empirical test. The collection of data of numerous people grounds in the assumption that people differ in their perception. In experimental research, there will never be a fully homogenous response. Even if people »agree« in their perception, their testimonies (i.e., ratings in a questionnaire or self-reports) are still likely to differ. This is because every measurement is a sum of the »true response« and measurement error. Experimental designs thus try to minimize the error and explicitly model it. Important factors influencing variations are the different ways in which people use rating scales, individual states of mind and attitudes. Thus, perfect agreement, so to speak, does not exist in actual experimental testing. However, there are test to establish relative XX. Specifically, to measure »agreement«, we applied the Beholder Index (»bi«).⁷⁵ The »bi« method estimates variance components that can be interpreted as the observed variance attributed either to the participant or to the stimulus.⁷⁶ In order to do this, participants need to rate each stimulus twice. The repeated measure allows for estimation not just of how much participants agree with each other on a rating of a stimuli (shared evaluation), but also how much participants agree with themselves on the rating of a stimulus (private evaluation). These variance components can then be combined into the beholder index. The higher the »bi«, the more variation can be explained by private, individual taste. Consequently, the lower the »bi«, the more participants agree in their aesthetic response.

The study was conducted in 2017 at the EVA-Lab at the Faculty of Psychology of the University of Vienna. In a controlled laboratory setting, participants saw high quality digitized images and were asked to rate these on a semantic differential. The sample consisted of 100 participants (24 male, age range = 19-48, M = 24.4, SD = 4.7) with balanced groups of 50 art history

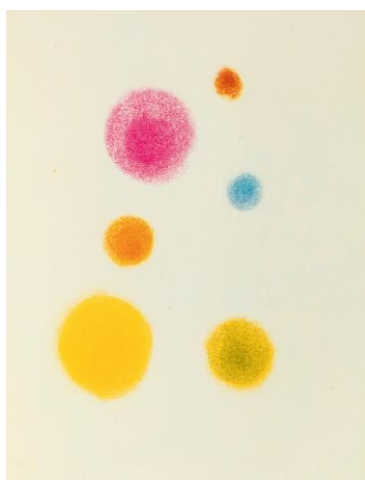
⁷⁴ Brinkmann et al. 2018. Seite?

⁷⁵ J. Hönekopp, Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness, *Journal of Experimental Psychology: Human Perception and Performance*, 32(2), 199–209.

⁷⁶ For a comprehensive description of the estimation of the variance components, see Leder et al., 2016, 2.

students (9 male, age range = 19-48, M=25.8, SD=6.3) and 50 psychology students (24 male, age range = 19-43, M= 25.1, SD=5.6).⁷⁷

The results of our study show significantly lower agreement in the effects of artworks and their single elements than expected. We controlled for art expertise, assuming that training in art history will influence the ratings. However, art expertise did not play a significant role and agreement was low on dimensions usually of interest in empirical aesthetics (e.g., like-dislike).⁷⁸ The statistical analysis collapsed responses to all stimulus material and all rating scales into three groups of image-types, specifically, images with single elements, images with line or color combinations, and the intact artworks. This means that image-specific differences are lost in the analysis. Yet, the analysis of Kandinsky's watercolor already showed that the qualities of the discrete lines differ enormously. We therefore also analyzed the ratings for each individual image in a descriptive and explorative manner by examining the distribution of ratings and the mean values. We interpreted a narrow distribution of ratings as strong ›agreement‹, as a narrow distributions means that the ratings center on a relatively small area. A wide distribution, by contrast, means little ›agreement‹. Based on this standard, we found that the ratings on only 4 out of the 37 images of the study could be considered as showing strong ›agreement‹ among the participants, and only on some of the 14 aesthetic effects. These were figures X-Z (horizontal, diagonal bundle and wavy line), extracted from Kandinsky's watercolor, and a color combination image derived from Miró (fig. 7).



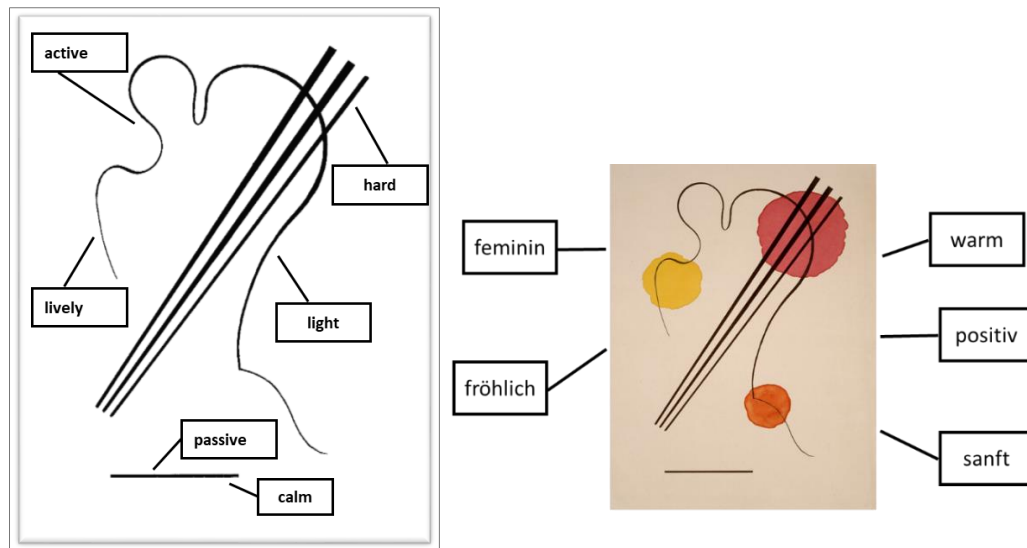
7 XX

Specifically, for the Kandinsky-images, participants agreed on the terms ›calm‹ (experts and lay persons) and ›passive‹ (lays) for the horizontal line, the term ›hard‹ (experts and lays) for the

⁷⁷ All participants received a compensation of 30€, for their participation at the end of the second session. The experiment was carried out in accordance with the Declaration of Helsinki and in a procedure, which was approved by the local ethical committee of the University of Vienna.

⁷⁸ SPECKER. Specker et al. Is Red Warm for everyone?

diagonal bundle and ›lively‹, ›active‹ and ›light‹ (experts) for the wavy line (fig. 8A). This comes as no surprise, as it aligns with Kandinsky's theoretical work outlined above. Interestingly, the watercolor itself was rated as ›feminine‹, ›happy‹, ›warm‹, ›positive‹ and ›smooth‹ (fig. 7B). This shows clear differences in the aesthetic effects of the discrete elements. Moreover, the distribution of ratings was relatively high, which means that there was relatively little agreement among participants.⁷⁹



8 A-B. XX

[Difference between image and parts] [Questions and doubts] Turning to Götz, he XX:

»Das, was von einem Bild wahrgenommen wird, sind nicht etwa die einzelnen, isoliert betrachtet oder gemessenen, Elemente, die bei der Bildanalyse benannt werden. Es ist auch nicht die Summe aller Elemente, denn ein Bild ist nur physikalisch betrachtet ein Aggregat, Psychologisch kann es nur als „Ganzes“ erfaßt werden, nämlich als das Resultat all der vielen wechselseitigen Veränderungen, die sich zwischen den Elementen abspielen, so wie unsere individuelle Informationsverarbeitung sie uns zum Bewußtsein bringt. [...] Ersetzt z.B. ein Maler an einem fertigen Bild ein Element durch ein anderes, so weiß er aus Erfahrung, daß dieser Wechsel sich auf das ganze Bild spürbar auswirken kann, selbst dann, wenn es sich um einen relativ kleinen Fleck handelt.«⁸⁰

One limitation of our first study is the selection of terms to measure aesthetic effects. The terms are drawn from historical sources that originated in a particular historical milieu, in which the aesthetic effects of lines and colors were considered to be significant for communication and

⁷⁹ Interestingly, for the color combinations of the Miró image, people agreed on very similar terms: ›happy‹ (experts), ›positive‹ (experts), ›light‹ (experts), ›warm‹ (experts and lays) and ›soft‹ (experts). Yet, for the intact image we could not find a high agreement on the aesthetic effects.

⁸⁰ Götz 1972, 112.

were part of the art-talk in the printed press.⁸¹ Although the question of universality of aesthetic effects is still highly relevant, the historical terms used to describe aesthetic effects might not be suitable anymore for today's beholders, where other forms of communication are perhaps more prominent. Moreover, since the highest agreement was found in the Kandinsky images, it might also be the case that the terms are especially well suited for his work specifically.⁸² Another limitation is that the terms can be considered as jargon, that is to say, they are part of a specialized language that art historians use to describe artworks. In a second study, we investigated this ›art-history-bias‹ by examining the adequacy of the terms to describe aesthetic effects.

Our starting assumption was that art historians are trained to use a specific and highly specialized language for describing artworks. [Explain: rationale of the study] We asked 30 art historians of the University of Vienna to judge how adequacy they thought the 14 adjective pairs (introduced above) are for describing the 37 images of our stimulus set.⁸³ The study was conducted in the controlled laboratory setting of the the Lab for Cognitive Research in Art History (CReA). Here, participants saw the images and the semantic differential on a high resolution computer screen (cf. study 1). They received the following task: »Wir bitten Sie, die Bilder nach 14 verschiedene Kriterien auf einer Skala von 0-100 zu bewerten. Uns interessiert, für wie passend Sie diese Kriterien halten, um die gezeigten Bilder zu charakterisieren.« Thus, participants were not asked to rate the images, but instead were instructed to rate the terms as descriptors for aesthetic effects. They could freely choose a point on a scale from 0% - 100%, thus indicating whether a term was absolutely adequate as descriptor of the aesthetic effect of a given stimuli (100%), or not adequate at all (0%).

The results showed little agreement among the art historians. There was no mean rating above 70% and the standard deviation was high [numbers]. This means that the participants [X]. In fact, based on the ratings, some terms were considered not suitable at all. These were »warm-cold« (for single lines) and »masculine-feminine« (for almost all images). [Another doubt: methodological problems: use of scale].

⁸¹ Bann 1980.

⁸² An examination of a possible ›Kandinsky-bias‹ will be part of future research.

⁸³ PhD students and MA-students at the stage of writing the thesis.

Conclusion: universal aesthetic effects of line and color?

We looked for an empirically identifiable basis for the claim of a »universal« aesthetics of line and color. [To be filled].

To recall Wölfflin's statement describing lines as being warm and cold. He would later come to downgrade this characterization by writing »nonsense« in his own copy of the book (between pages 14 and 15), underlining the words »warm and cold« and »hard and soft«.⁸⁴ When precisely Wölfflin made the notes is unclear.⁸⁵ However, [To be filled]. Similarly, Endell denied that principles of »male and female art« , [To be filled].

⁸⁴ Wölfflin 1886, 14. »warmen und kalten« and »hart und weich«. Notes between pages 14 and 15.

⁸⁵ The annotated copy is kept in the Getty Research Institute, Special Collections. See: Andrew James Hopkins, Heinrich Wölfflin's Own Annotated Books, *Getty Research Journal* 7, 2015, 177-184.

2.3 Field Research

Articles

2.3.1

Pelowski, M., Leder, H., Mitschke, V., **Specker, E.**, Gerger, G., Tinio, P. P. L., ... Husslein-Arco, A. (2018). Capturing Aesthetic Experiences With Installation Art: An Empirical Assessment of Emotion, Evaluations, and Mobile Eye Tracking in Olafur Eliasson's "Baroque, Baroque!" *Frontiers in Psychology*, 9, 1255.
<https://doi.org/10.3389/fpsyg.2018.01255>

My Contribution:

Frontiers requests you to note all authors contributions. As noted there:

MP, HL, VM, EV, TB, and AH-A contributed to the study design and the execution of the research study.

MP, VM, GG, ES, EV, and TB contributed to the data analysis.

MP, HL, VM, GG, ES, and PT contributed to the writing of the paper.

Concerning data analysis, I was responsible for all network analyses reported in the paper. Consequently, I was also the main author of this section of the paper. In addition, I provided feedback on consecutive versions of the paper.



Capturing Aesthetic Experiences With Installation Art: An Empirical Assessment of Emotion, Evaluations, and Mobile Eye Tracking in Olafur Eliasson's "Baroque, Baroque!"

Matthew Pelowski^{1*}, Helmut Leder¹, Vanessa Mitschke^{1,2}, Eva Specker¹, Gernot Gerger^{1,3}, Pablo P. L. Tinio⁴, Elena Vaporova¹, Till Bieg^{1,5} and Agnes Husslein-Arco⁶

¹ Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, Vienna, Austria, ² Department of General Psychology II, University of Würzburg, Würzburg, Germany, ³ Department of Psychology, Webster Vienna Private University, Vienna, Austria, ⁴ Department of Educational Foundations, Montclair State University, Montclair, NJ, United States, ⁵ Center for Technology Experience, AIT Austrian Institute of Technology, Vienna, Austria, ⁶ Museum Belvedere, Vienna, Austria

OPEN ACCESS

Edited by:

Nadia Bianchi-Berthouze,
University College London,
United Kingdom

Reviewed by:

Michał Przemysław Muszyński,
Université de Genève, Switzerland
Ursula Beermann,
Universität Innsbruck, Austria

*Correspondence:

Matthew Pelowski
matthew.pelowski@univie.ac.at

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 03 February 2018

Accepted: 29 June 2018

Published: 06 August 2018

Citation:

Pelowski M, Leder H, Mitschke V,
Specker E, Gerger G, Tinio PPL,
Vaporova E, Bieg T and
Husslein-Arco A (2018) Capturing
Aesthetic Experiences With
Installation Art: An Empirical
Assessment of Emotion, Evaluations,
and Mobile Eye Tracking in Olafur
Eliasson's "Baroque, Baroque!".
Front. Psychol. 9:1255.
doi: 10.3389/fpsyg.2018.01255

Installation art is one of the most important and provocative developments in the visual arts during the last half century and has become a key focus of artists and of contemporary museums. It is also seen as particularly challenging or even disliked by many viewers, and—due to its unique *in situ*, immersive setting—is equally regarded as difficult or even beyond the grasp of present methods in empirical aesthetic psychology. In this paper, we introduce an exploratory study with installation art, utilizing a collection of techniques to capture the eclectic, the embodied, and often the emotionally-charged viewing experience. We present results from an investigation of two pieces, both part of Olafur Eliasson's exhibition "Baroque, Baroque" held at the Belvedere museum in Vienna. These were assessed by pre- and post-viewing questionnaires focusing on emotion, meaning-making, and appraisals, in tandem with mobile eye tracking to consider viewers' attention to both installed artworks and/or to the museum environment. The data showed differences in participants' emotional states, appraisals, and visual exploration, which together paint a picture of the aesthetic reactions to the works. These differences also showed how viewers' appraisal strategies, meaning making, and physical actions facilitated relatively more or less deep engagement with, and enjoyment of, the art. The results are discussed in terms of their implications for museum studies, art education, and theory in empirical aesthetics.

Keywords: museum study, aesthetic emotion, mobile eye-tracking, installation art, art perception

INTRODUCTION

Installation art represents one of the most important, and one of the more empirically vexing, developments in the last 50 years of art's production and scholarship (Osborne, 2002; Bishop, 2005). Involving a monumental shift in emphasis from traditional

bounded sculptures or two-dimensional images placed before a viewer, installation artworks are instead comprised of site-specific pieces that *envelop* an individual, often incorporating aspects of the existing environment and designed to bring about complex sensory and emotional experiences (MOCA, 2002). As such, installation artworks can be particularly intriguing for perceivers who are confronted with a visitor-centered medium that often requires their active and mobile interaction (Reiss, 2001). Art installations are also a key emphasis for contemporary museums, where they have become centerpieces of new collections and monetary investment (Bishop, 2005). For the empirical researcher or art-interested psychologist, installation artworks are also particularly intriguing. They specifically possess features aligning with recent interest in art's ability to emotionally move and to conceptually challenge the viewer (Pelowski et al., 2017b), as well as raising importance of context (e.g., laboratory vs. gallery) or *in situ* ecologically valid conditions as a major component in the art experience (Pelowski et al., 2017a).

At the same time, despite this interest, installation art also poses particular challenges to its understanding and to its assessment, rooted in the very same aspects of the medium and the responses that it may elicit. Installation pieces often discourage traditional appraisals that involve some type of hedonic judgment ("it is a beautiful or pleasing object") or mimetic identification ("it is a picture of 'X'"). Instead, they require the visitor's consideration of the juxtaposition of space and artwork elements, and often their reflection on emotions, bodily sensations, as well as on ambiguity or confusion in such responses. These aspects are argued to be central to the artworks and perhaps key to their enjoyment. Particularly for the viewer with limited art knowledge, these experiences—as with other concept driven art—can lead to displeasure (Silvia, 2013), and are even given as a major reason for why individuals may not visit contemporary museums or enjoy contemporary art (Eaton and Moore, 2002; Goldie and Schellekens, 2009). This also poses a conundrum for museum directors—as well as for artists—with better understanding of how and why individuals do react to installation pieces of key importance for pragmatic, educational, and curatorial decisions. It is equally important to know how certain responses might lead to more or less appreciation, or even the range of responses that viewers might have to such art. For researchers who would do empirical studies, however, and thus who might begin to answer these pragmatic and scientific questions, because it cannot be collapsed into a static image with a seated viewer, and involves a temporally extended, complex, site-specific interaction, installation art has to-date almost never been empirically considered (Minissale, 2012; Weingarden, 2014; Kranjec, 2015).

In this paper, we report on a first exploratory study designed to begin capturing and partially differentiating the viewer experience of museum-based installation art. This was conducted at the exhibition, *Baroque, Baroque*, by the artist

Olafur Eliasson, at the Belvedere Museum in Vienna, Austria, and built around a unique opportunity to collaborate with curators and staff of the museum. By selecting two distinct rooms within the exhibition, and employing a combination of behavioral and eye-tracking methods, we examined the interplay amongst viewers' emotions, artwork evaluations, meaning interpretations, and perception or attention patterns as they engaged with the art. Although designed from the output to be an exploratory test case, this study, through its unique blending of approaches, is hoped to provide both a range of means and supporting evidence for future work, unlocking the unique experiences and key engagement features with these complex and intriguing artworks. Below, we also briefly review the main issues and the unique qualities of installation art as well as past studies, in tandem with a handful of research questions, which we feel give a good starting point for considering some of the most salient issues and our selected methods of analysis. We conclude with a brief discussion of implications and ideas for future research.

REVIEW—INSTALLATION ART: RESEARCH QUESTIONS, ISSUES, AND APPROACHES

Installation artworks might in fact be said to present a particularly perfect intersection of issues at the forefront of suggested new directions in empirical/psychological aesthetics. Beginning with the work of Alan Kaprow (e.g., 1950s; see Martinique, 2016) and becoming more common throughout the 1970s and to today, this medium again is defined by a process whereby an artist takes an existing space, often in public areas, in nature, or a room/gallery, and designs the entire environment to create conditions for a cohesive, unique, interactive experience. This can of course be done in any number of ways with different foci or materials. The earliest Kaprow works literally filled a room with various objects such as old tires, or sheets of plastic, colored cloth, etc. often paired with electronic sounds or music. Christo and Jeanne-Claude famously created numerous outdoor installations by for example wrapping the Berlin Reichstag in cloth or placing cloth gates throughout New York's Central Park through which one might pass while walking (Martinique, 2016). Other artists such as Bruce Nauman or James Turrell use colored lights or words in gallery spaces to confront or effect the mood of a viewer. Ai Wei Wei recently filled museum rooms with millions of porcelain sunflower seeds. Damien Hirst has installed glass cases with bisected animals; and the artist of the present study, Olafur Eliasson, converted a gallery in Denmark to reproduce a riverbed in his native Iceland or filled an entrance hall at London's Tate Modern with a looming, setting sun.

Such artworks might lead to multiple ways of engaging and participant responses, much as with other art media. However, certain factors are potentially most prominent and thus key candidates to empirically consider: First, installations tend to

cause and/or emphasize the foregrounding of some aspect of the viewer's affective responses, especially felt emotion. This has to do with the designed enveloping nature of the medium, which coincides with an expectation for the presence of a perceiver who, by engaging within the space and interacting with all senses, in a way completes the artwork. Viewers may have a basic awareness and appreciation or remembrance of certain sensations—such as the warmth and intensity or even melancholy of squinting and staring into a sun or walking in an icy barren land (Eliasson). They may feel the weight and leaning sensations from huge blocks of metal mirrored in their own body (Richard Serra; see Kapoula et al., 2011). Bruce Nauman's Green Light Corridor is known for negative responses, evoking a feeling of discomfort, disgust or claustrophobia (Griswold et al., 2013). Installation artists might in fact often be said to design their works specifically to evoke specific reactions, considering how their environments will impact the viewer's mood and body. The “participatory” nature of this medium (Novitz, 2001; see also Griswold et al., 2013) may in turn directly require some active awareness of emotions by a viewer and reflection on the nature and importance of their affective experience, and which may also be one artistic *point of* installations.

By creating an encompassing space often with a moving viewer, installation artworks may also act to “walk an individual through” an evolving encounter, and individuals might experience several emotions or evaluations within one experience. For example, a space may be designed to slowly reveal new features. Artworks may also require some acclimation or dawning awareness of different senses. Artists may also anticipate juxtapositions which could lead to mixed (positive and negative) or changed response. One might feel an initial discomfort or oppression, say an overly bright light as in Eliasson, or disgust from being confronted by an embalmed animal (Hirst), followed by relief, interest, or aesthetic appreciation.

This also raises another key aspect, regarding some amount of incongruity, insight, learning, or changed understanding. By foregrounding certain features, installations may serve to force an awareness or deeper consideration. An artist may cause an individual to consider how responses are typically taken for granted, or individuals may even come to deeper appreciation by juxtaposing visual and proprioceptive sensations (awareness of being in a gallery). For example, as noted by Sherman and Morrissey (2017), the works of James Turrell—denoted by rooms flooded with differently colored light—have been “described... as transformative. The immersive light environments cause one's own perception to become the object of reflection and led many to a deeper understanding of themselves and their relationship to the external world, whose access to the world is mediated by visual perceptual faculty with particular features, limitations, and abilities.” This may also directly involve a sense of discrepancy between the “art” and the pre-existing environment. By framing certain artistic interventions within, say, a gallery setting, artists may cause individuals to rethink their own responses, to consider different viewpoints (i.e., “aesthetic” vs. more descriptive or every-day

evaluations) or even create some confusion regarding what is art or the environment. Installations may also utilize a more overtly cognitive interpretations or meaning-making. The *Sunflower Seeds* by Ai Wei Wei, in addition to presenting a visually compelling stimulus, was also handmade by 1600 artisans and had the goal of commenting on mass production and consumption in conjunction with aesthetic experience (Martinique, 2016). The embalmed animals by Damien Hirst may confront viewers with their own mortality or cause them to question their aestheticizing of death (see Pelowski et al., 2012 for other examples).

These processes, as they relate to meaning interpretations are also argued to be a key element for the appreciation of installation art—suggesting a movement away from more traditional responses or attempts to identify the mimetic content or provenance of a work (who is it by, what is depicted, what is it made of) to a meaning relating the artwork significance to its ability to create the conditions for a reflective, insightful, or emotional experience. Such an “experience-based” meaning, although of course not unique to installation examples (e.g., see Pelowski and Akiba, 2011 for discussion with mimetic or abstract paintings), are an increasingly key aspect of contemporary conceptual and postmodern art (Bourdieu, 1979/1984; Goldie and Schellekens, 2009), and in fact have been argued to represent a more general goal of art viewing or art education, representing a personalized, deepened art experience (e.g., Dewey, 1980; Tucker, 2004; Pelowski and Akiba, 2011).

It is also such features, in conjunction with cognition, reflection, and body/affective responses, that may pose particular difficulty with viewers, who may expect to be faced with requirements for formal appraisals or understanding of content, and may thus not be prepared for an insightful, introspective, or shifting experience. It may also be that certain responses may not coincide with all art examples—for example Christo argued that his works should not have deeper meanings beyond the sensory experience (Martinique, 2016). These may of course differ between viewers, with only certain responses leading to more or less enjoyable, engaged, or meaningful responses.

All of these aspects—diverse emotions, awareness and reflection, meaning making, as well as a tendency to juxtapose environment/expectations and even insight—mark compelling research questions. They also mark the present limitations in empirical approaches. This can be traced in both study approaches as well as the theoretical models for processing art. A recent review of main theoretical bases for assessing art (Pelowski et al., 2016) specifically noted their lack of focus on the above features. Primarily this is due to a theoretical basis in prior models of basic vision and attention to low-level artwork features (color, line, form, content, etc., see e.g., Leder et al., 2004), which tend to then only consider responses to these aspects in art processing experience. This in turn connects to empirical studies, which also focus on basic (positive/negative) appraisals or responses, often with less than 5 s viewing durations (Pelowski et al., 2017a for review). This tends to omit changes or evolutions in art experience.

The above limitations also relate specifically to assessing emotion. While models and studies do often include affect, as well as understanding and appraisals, as major components, these are considered only in the context of basic hedonic pleasure/displeasure, arousal, or visual interest. There are presently few attempts in empirical study to take a simultaneously more detailed and broad perspective. Especially, researchers argue for needed attempts to map and record complex, shifting, and negative varieties (e.g., Silvia, 2009), as well as intense aesthetic responses such as being moved, transformed, or feeling resonance with art. Similar arguments are made for needed analyses of mixed or conflicting emotions (Hosoya et al., 2017). This is also important for insight, discrepancy, or even more top-down reflection, and meaning making, which are also obscured in current models and studies (Pelowski et al., 2017b), which also tend to focus on basic identification or understanding of artwork content, and thus do not consider the more personalized reactions described above. In turn, researchers have argued for the need of a broader focus on these different factors, capturing aesthetic responses arising from the interplay of emotion, cognition, expectations, more open-ended interactions (Tallis, 2008; Brown and Dissanayake, 2009; Augustin et al., 2012; Cross and Ticini, 2012).

Installation art also of course raises empirical issues relating to setting. Unlike many other art examples, installations require a site-specific viewing. This is in contrast to most empirical assessments which typically move the viewer to a laboratory and a computer screen for purposes of control and experimental focus (Cela-Conde et al., 2011; Pelowski et al., 2017a). However, this itself might lead to several issues. From the founding of psychology of art as a scientific field (e.g., Fechner, 1876), scholars have agreed that a major component of art's impact and character involves the tangible, immediate, and "real" experience of original artworks (Tschacher et al., 2012). Scholars also argue that current experimental approaches, dominated by laboratory studies, may lead to differing types of reactions or may not be able to elicit the emotions and juxtapositions found especially with *in situ* art (Brown and Dissanayake, 2009).

Presently, there are very few studies even of art within galleries or museums (e.g., Pelowski et al., 2017a for review). These primarily have focused only on general appraisals such as liking or beauty, asking participants to both assess a succession of artworks (paintings or photographs) on a monitor and/or also to assess the same artworks within a gallery. Results have in turn shown higher ability of real examples to evoke liking and interest (Locher et al., 2001), longer durations of viewing (Brieber et al., 2014), as well as arousal (e.g., Brieber et al., 2015 which also included sculptures as well as paintings) or emotional intensity (Specker et al., 2017). Certainly, this issue is not unique to even visual art (see also Scherer and Coutinho, 2013 for a similar argument with music), but, in conjunction with the general focus on evoked emotion and body responses, may be particularly salient in installation art study.

The need for more on site, in depth studies has only grown with the increasing awareness of the importance of context

in aesthetic judgments. Researchers are becoming concerned with the need to consider art reception as a complex interplay of the expectations of the viewer, the characteristics of the art object, and the conditions in which it is experienced. Installation art also adds the unique aspect of the mobile viewer and the inclusive environment, with an individual who may potentially engage with any aspect of the spaces as part of their encounter. Here too, although some emerging studies have shown the ability to generally track patterns of movement or body responses in the gallery (i.e., heart rate and standing patterns when looking at more or less interesting art, Tröndle et al., 2014a) this has generally not been considered in conjunction with complex emotional/cognitive experience (see Mitschke et al., 2017 for discussion of the key intrigue and need for such studies in empirical aesthetics). Although not specifically necessary, researchers would also benefit from access to curators and artists, especially when considering their making decisions and specific goals for evoking certain response, which may also be useful in forming hypotheses and assessing the efficacy of the works.

We are aware of only two examples of empirical studies that have used installation artworks: Tröndle et al. (2014b) considered reactions to a conceptual installation ("A Label Level, 2009" by Nedko Solakov), composed of black marker comments written on the museum walls and commenting on the gallery's other displayed paintings, drawings, and sculptures. In this case, the researcher, however, were interested in whether participants would find the installed writing to be an "artwork," finding a correlation between positive answers and higher ratings for importance, beauty, artistic skill, composition, and even curatorial quality, as well as interpersonal differences in regards to art classification, and frequency of museum visits, expectations that the exhibition would be thought-provoking, would touch all senses, and would have famous artworks, and with general appreciation of other video, performance and installation art. Kapoula et al. (2011) also report intriguing evidence from a study of Richard Serra towers installed in a gallery. Although mentioned only briefly as a pilot assessment within the context of a different study, they fitted viewers with a belt accelerometer while they walked through the installation and found that individuals unconsciously adopted the same degree of lean with their own bodies. These studies are compelling in highlighting unique responses and in suggesting that Installations can routinely be assessed even with lay museum viewers, but again only tangentially touch the most compelling aspects of the medium and the resulting experience.

An Integrative Approach for a Pilot Study of Installation Art Experience

To begin our assessment, it was thus necessary to combine several approaches and range of assessments which have been employed in our, and others', previous gallery studies, involving assessments of the more descriptive/global assessment of art experience. This also involved our decision to focus on a handful of features—emotion, aspects of meaning making, general appraisal, and basic

patterns of looking—which we felt to be most salient from the above review, and a good beginning point for this analysis. The study utilized a paradigm developed in our previous work in museums, in which participants are asked to interact with works of art, individually contained within one room, and with minimum prompting or intervention. This is then followed by a post-viewing survey.

The primary means of assessment involved soliciting participant reports of their felt emotions using a list of terms meant to capture a range of aspects or outcomes. This followed previous work in appraisal theory (e.g., Silvia, 2005a; Scherer et al., 2006) that suggests that specific emotions might be tied to the personal relationship between viewer and art and through mapping their general incidence may provide a broad qualitative sketch of personal experience. In turn, use of a list of emotions has been recently employed in art studies or with other aesthetic media, often collecting the range of responses felt in one's total museum interaction or hypothetically related to aesthetic media (e.g., Zentner et al., 2008 with music; Hosoya et al., 2017; Schindler et al., 2017 for general aesthetic responses), but also utilized in specific interactions with particular works of art (Specker et al., 2017).

The emotions employed here were based on a list produced as part of an art processing model (Pelowski and Akiba, 2011; updated in Pelowski et al., 2017b) which did attempt to lay out general main outcomes of art experience. This included a range of responses differentiating between largely facile/unemotional, negative/discrepant or highly positive response, as well as terms recently connected to installation or emotionally resonant art—insight, transformation, joy, thrill, being moved etc. This was used in a previous gallery study (Pelowski et al., 2012; Pelowski, 2015), which approached an installation setting—the Rothko rooms, consisting of series of paintings and individual spaces designed by or highly controlled by the artist—showing ability to meaningfully distinguish between broad varieties of experience and especially to identify particularly resonant or insightful encounters. In order to more fully paint a picture of viewers' personal/emotional combinations, we also utilized a network modeling technique (e.g., Epskamp and Fried, 2016; Hosoya et al., 2017 for use with art). This method can visually and mathematically group and connect individual emotion responses via their partial correlations, and thus visualizes the strength and interconnections in an emotional space. It is especially useful for descriptive/exploratory studies, clustering items without an underlying assumption of higher-order factors, providing information regarding the centrality, interconnections, and specific importance of items in defining the global affective experience.

The above emotions were also combined with a list of artwork appraisals and also with an assessment of meaning using the open-ended question, "What did the artwork mean?" This has been shown to be effective in augmenting emotional and evaluative scales (see Pelowski et al., 2012; Pelowski, 2015). Specifically, answers can be divided into three general types that can further be used to demarcate the experience: (1) descriptive statements about the formal, historical, or semantic content of an artwork (e.g., it is "a picture of X," or it is "a painting by Y");

(2) statements regarding an artwork's lack of meaning; and (3) statements in which the meaning or interpretation of an artwork is related to personal experience, emotions, or bodily sensations (e.g., "it reminded me of Y," or "it made me feel X"). This allowed us to both consider the general approach of viewers—regarding focus on content/semantic information or emotion—in their interpretations of the artworks' meaning and significance, as well as to identify particular responses described in the above review.

The typical strategy, used by especially naïve art viewers when assessing paintings or sculptures, again often involves responding with descriptive/semantic interpretations (e.g., focus on the content, materials, physical spaces, or places, etc., Leder et al., 2014). In past studies (Pelowski et al., 2012), more direct descriptive responses have also been shown to coincide with approaches typically taken by (often novice) viewers, and, for the present artworks, are also those responses argued to perhaps not be suitable for installation art. Similarly, finding no meaning or feeling a work to be meaningless often coincides with generally negative emotional and cognitive experience. On the other hand, the latter so-called "experience based" interpretation (e.g., see Pelowski and Akiba, 2011)—which can often include both basic recognition of emotional/cognitive responses as well as insight or even metaphorical meaning assumed to be intended by the artist—was argued to be vital for appreciating installation art. Such interpretations have also been shown, in previous museum studies with abstract or more conceptual pieces (e.g., Pelowski, 2015), to correlate with more positive and rewarding assessments.

By combining these measures, we assessed the following research questions: (1) Do installation artworks produce distinct emotional, evaluative, perceptual, and meaning responses that may be captured and that may differ within-subjects? In addition, (2) How do viewers generally respond to the art, and particularly are specific factors linked to more understanding and enjoyment? More specifically in reference to this second question, we considered if certain emotional responses, or general magnitude of emotion, might lead to more positive appraisals. This was in keeping with the above-argued importance of evoked emotion. We also assessed if similar patterns might be found between experience-based interpretations and/or focus on affective experience in meaning, rather than identifying semantic content, and relatively more enjoyment. (3) We also combined the above behavioral measures with use of mobile eye-tracking in the museum installation space. This approach is only emerging in use with art viewing (Heidenreich and Turano, 2011; Walker et al., 2017). However, the technique, which can record looking patterns of a mobile, walking viewer, is especially useful for an installation environment where any part of the environment might be a point of attention. This study followed our recent work in real-world art interaction contexts, specifically Mitschke et al. (2017) which employed the same technology as viewers walked along a riverside path with a number of installed sculptures, graffiti art, in conjunction with the ambient visual environment. This study showed good ability to record general patterns of looking and attention on the different objects.

Due to lack of previous study and specific hypotheses, we were primarily interested in assessing general areas of attention (by placing attention on eye fixations), and considering whether viewers appreciated the installed elements as stand-alone artworks or as components that interact with the ambient environment. This latter approach was hypothesized to be an important aspect of installation artwork appreciation. We also considered how visual exploration itself (time spent looking at specific areas) relates to the experience of installation pieces. Time of total fixations is considered an important feature of psychological experience, representing attention and salience (Tatler, 2007), and has been shown to positively correlate with the appreciation of art (Brieber et al., 2014; Leder et al., 2016). We expected that longer fixation times on the artworks (as opposed to other elements of the rooms such as floors and ceilings) would coincide with more positive or emotional experiences. At the same time, because other aspects of the rooms might play a role in the experience, this technique allowed us to assess whether participants were assessing these features and how attention to these might also coincide with certain appraisals or outputs. By using a matched group of individuals with and without the glasses, we could also offer one more assessment of the efficacy of this technique itself in the art gallery.

MATERIALS AND METHODS

Stimuli

The setting for our study came about through a unique opportunity to collaborate with the staff, artist, and curators at the exhibition, *Baroque, Baroque*, by the artist Olafur Eliasson (November 2015–March 2016), at the Belvedere Museum in Vienna, Austria.¹ The exhibition consisted of a series of rooms, each containing one installation. According to its curators (Zyman and Codognato, 2015), the exhibition had the purpose of “establishing a dialog between” the baroque architecture and the artworks, creating “surprising affinities between Eliasson’s works and their temporary settings” or exploring “the relation between object and viewer” (Belvedere and Thyssen Bornemisza Art Contemporary, 2015, p. 2). We focused on two pieces (**Figure 1**), which themselves provided two potentially distinct viewing experiences while also combining, potentially, the above emotional and cognitive aspects. The first was *Eye See You* (Eliasson, 2006), a relatively small room ($\sim 3 \times 4$ m) that contained a sun-like sculpture made from a prefabricated mirror-polished solar cooker and a sodium lamp mounted at its center, which flooded the space with golden light. Two additional dichromatic glass disks changed colors according to a viewer’s position and movement. The installation provided a sense of staring into a warm sun, with a slightly shimmering “moiré effect” (Eliasson, 2017), suggested by the artist and critics to potentially elicit positive and/or melancholic emotion. The second installation, *Wishes vs. Wonders* (Eliasson, 2015), was

located in a slightly larger room ($\sim 10 \times 5$ m) and containing a number of painted panels on all walls (original to the building) depicting landscapes and battles. The artist subdivided the room with a tall mirror on which a metal half ring (5 m in diameter) was embedded and which hovered above the parquet floor. According to critics, this installation was expected to provide a “perception of dizzying completeness” or “produc(e) a virtual ring that appeared to float or pass from the actual space... uncannily traversing the surface of the mirror image” (Husslein-Arco and Habsburg, 2015), but also potentially calling to mind cognitive aspects of war or the human condition as reflected in the design.

Participants

The study involved 51 participants (all female, $M_{\text{age}} = 22.7$ years, $SD_{\text{age}} = 4.7$, range: 18–49) who viewed both rooms in a counterbalanced order. Participants were psychology students from the University of Vienna. They had normal or corrected-to-normal vision and participated for course credit. All participants were novice art viewers with no previous training in art making or art history (assessed via a post-test survey). None of the participants had previously seen the installations. The sample was divided into two groups: those who used the eye-tracking glasses ($N = 24$) and those who did not ($N = 26$).

Procedure

The procedure again followed that reported in our past museum studies (Pelowski, 2015). Participants were met at the entrance of the exhibition hall in groups of two (after having previously signed up for the study and given a scheduled time to make their visit). They were guided to a small room, which was separated from the galleries but still located within the museum. They were asked to provide informed consent and complete a pre-viewing questionnaire (see below). The participants were then separated and one individual from each pair was equipped with eye-tracking glasses that were calibrated (see below, following procedure of our previous art-viewing study by Mitschke et al., 2017). This participant was allowed to briefly acclimate to the use of the glasses. Before starting the gaze recordings, glasses-equipped participants were informed that their gaze behavior and audio were being recorded. Before viewing, participants also completed a series of Likert-type questions assessing their attitudes about art and past art knowledge or experience (7-point, 1 = “Does not apply to me at all”; 7 = “applies absolutely”). This was used as a means of ensuring their novice status.

Each participant was then escorted by a researcher to one of the two rooms containing the installation artworks, which were visually separated (**Figure 1**). Before entering, the participants were instructed as follows: “Please enter this room and stay for as long as you like. You have no time constraints or task requirements. When you are done viewing, please come back to me” (translated from German). After returning from the installation rooms, participants were guided to the testing room and given a post viewing survey about their experience

¹Held in the Winter Palace of Prince Eugene of Savoy, Vienna; officially a cooperation between Belvedere, Vienna, Thyssen-Bornemisza Art Contemporary Collection, Vienna, and The Juan and Patricia Vergez Collection, Buenos Aires.

²<https://www.yatzer.com/olafur-eliason-baroque-baroque>

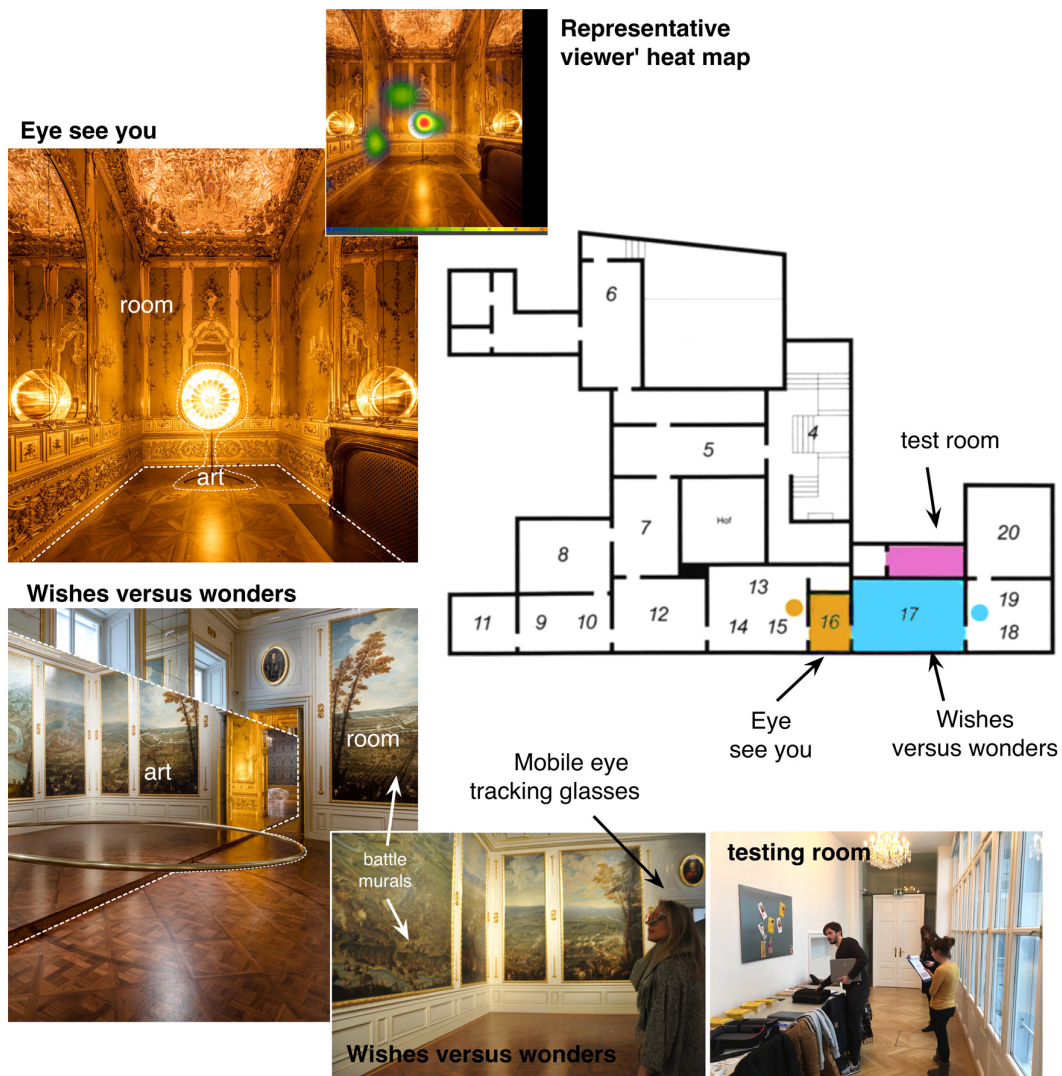


FIGURE 1 | Two installation art rooms, exhibition layout, and details regarding the testing procedure (Note, written informed consents was obtained from all depicted individuals for the publication of the images).

with the specific artwork. The same procedure was then repeated for the second room/installation piece. The amount of time that each participant spent inside each room was recorded. All participants were debriefed upon completion of the study. Participants in the eye-tracking group used the glasses for both rooms, with recalibration done following the first room.

Note also that the inclusion of a glasses and no-glasses equipped group was done both in order to provide a matched control for assessment of potential modulation by wearing glasses themselves, and also for pragmatic reasons, due to museum-imposed limitations on total time for testing. Based on our previous use of this technology (Mitschke et al., 2017) we did not expect any issues to arise from the glasses, and thus it was expected that both glasses and no-glasses groups could be combined in the other analyses.

Behavioral Self-Report Measures

To assess overall aesthetic experience, we used Likert-type questions related to the following dimensions: Once again, as a main means of assessment, after viewing each room, participants reported incidence and relative magnitude of 20 emotions (see **Figure 5** for all items), based on a similar list and procedure used in a museum study by Pelowski (2015). This has been shown to provide a good range of responses and ability to generally describe the global experience, as well as to meaningfully distinguish between broad varieties, differentiating between largely positive/negative and more or less profound/facile response (see also Ekman et al., 1980; Carstensen et al., 2000 for use of similar lists and Likert-type scale measures in non-art contexts). The list asked participants to report on their emotions via 9-point scale [i.e., “while I was inside the room, I experienced (emotion)”; 0 = “not at all” to 8 = “extremely”].

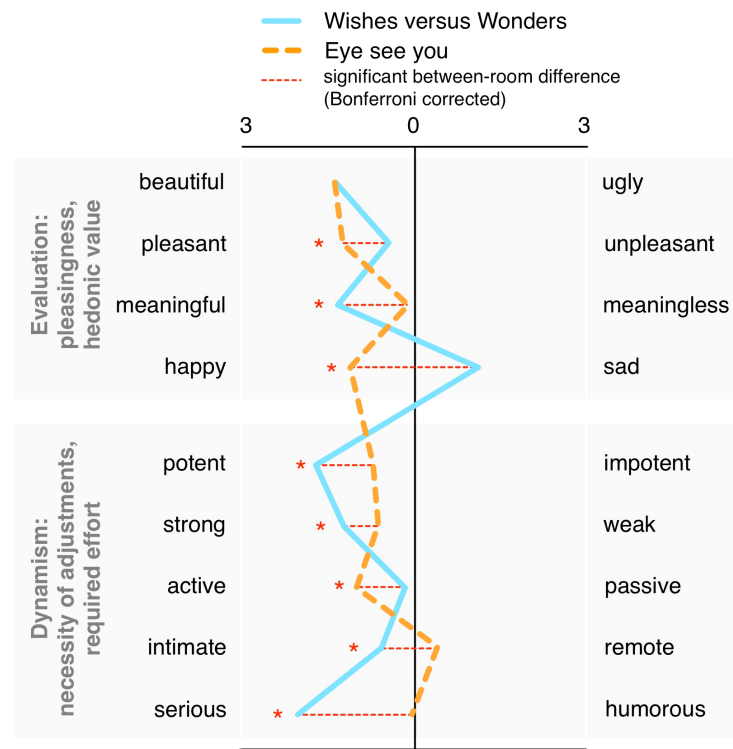


FIGURE 2 | Appraisals (post-viewing) for two rooms of installation art. * and dotted line indicate significance at $p < 0.05$, Bonferroni corrected (see also legend above).

Participants then rated the artwork using a series of nine bipolar scales (7-point scale, e.g., 1 = “extremely ugly” – 7 = “extremely beautiful,” all items listed in **Figure 2**). The items, based on the semantic differential technique (Osgood et al., 1957), were selected from a previous museum study (Pelowski, 2015) and combined both hedonic assessments (e.g., liking, pleasing, beauty) as well as terms relating to underlying stimulus potency (strong-weak) and activity (intimate-remote). These latter items—which can also be combined into a more general “dynamism” dimension especially in ratings with art (Kumata and Schramm, 1956; Berlyne, 1974)—have been argued to indicate a relative need for effort/adjustments and thus in conjunction with the above hedonic ratings a relative measure of engagement and/or processing difficulty (Carroll, 1959; see also Berlyne, 1974). In addition, participant’s answers to the hedonic scale “pleasing-displeasing” were expected to be used as a general assessment of art preference or enjoyment (e.g., Silvia, 2005b).

Following this, participant understanding of artwork meaning was assessed using the open-ended question, “What did the artwork mean?” and with the expectation of coding into the three levels of descriptive, no-meaning/no-understanding, and experience-based or emotion/introspection-centered interpretations (following Pelowski et al., 2012; Pelowski, 2015, described in the introduction above).³

³Note that this study also employed a number of other questionnaires (available upon request). These were used for a separate assessment of museum interaction

Mobile Eye-Tracking

Participants in the eye-tracking group had their fixations and fixation durations measured using a pair of lightweight mobile eye-tracking glasses (iViewETG; SensoMotoric Instruments, Teltow, Germany) that was connected to a small data recording unit (approximately the size of a cellular phone) carried inside a bag worn around the waist and that stored data at a rate of 60 Hz (monocular). The glasses were calibrated (prior to the viewing of each room) inside the testing room with a three-point calibration method using the provided software (following Mitschke et al., 2017).⁴

Ethics Statement

The study was approved by the ethics committee of the University of Vienna and by the Belvedere Museum. All participants gave informed written consent.

RESULTS

The results are reported in two sections that are aligned with the research questions: (1) We begin with a general

conducted by a collaboration partner, administered last after the above reported procedure and questions and will not be referred to in our analyses.

⁴We also recorded total fixations and total fixation duration. However, these measures are typically highly correlated. It is in turn suggested that duration may be a clearer measure of attentional focus.

description of the art experience in both rooms that considered viewers' evaluations, emotion, meaning-making, eye fixations, and between-room differences. This section also includes a technical assessment of eye-tracking vs. no eye-tracking groups (research question 3), as the results of this assessment were important for treatment of the other data. (2) We then consider a further breakdown between viewers reporting generally positive/negative experiences

and tie these responses to reported emotion and looking patterns.

All participants completed all aspects of the study. Inspection of the histograms and Q-Q plots for the emotion and artwork appraisal items suggested, in general, approximately normal distributions, however, with some having a rightward or leftward skew due to the unidirectional nature of some scales. No outliers were detected (based on boxplots). Given

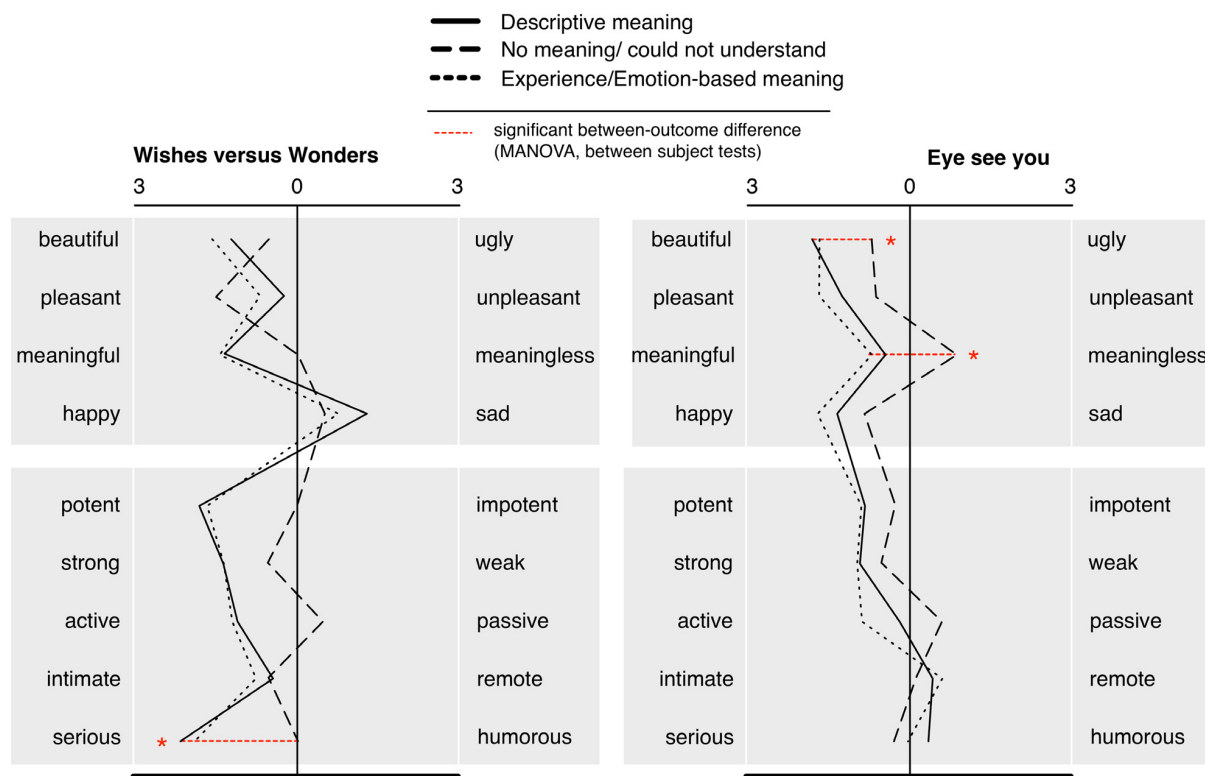


FIGURE 3 | Differences in appraisals (post-viewing) between different types of answers to “what was the meaning of the art?” for the two rooms of installation art. * and dotted line indicate significance at $p < 0.05$ (see also legend above).

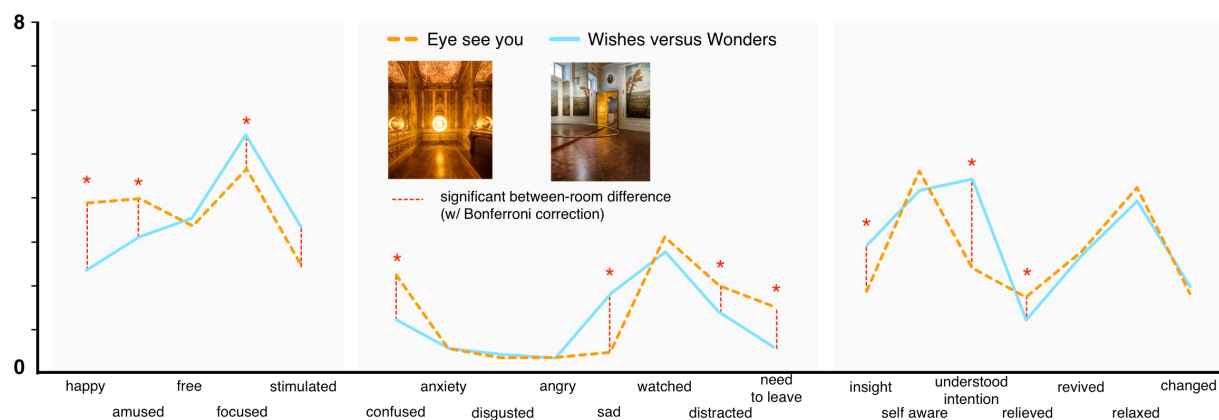


FIGURE 4 | Reported emotions (post-viewing) in two rooms of installation art. All reported significant differences are based on univariate analyses with Bonferroni correction for multiple comparisons. * and dotted line indicate significance at $p < 0.05$ (see also legend above).

our sample size and scale-based data, as well as the expectation of primarily using Analysis of Variance assessments which are robust to minor deviations, statistical tests for normality of distributions (e.g., Shapiro–Wilk) were not conducted (see Altman and Bland, 1995; Field, 2009). Assessment of homogeneity of variance for univariate comparisons are noted at the relevant locations below. In cases of violation of sphericity (repeated measures ANOVAs), Greenhouse-Geisser corrections were applied. Viewing time and eye-tracking data showed no violations in normality of distributions and variance homogeneity.

Comparison of participants wearing the eye-tracking glasses vs. those without glasses also did not reveal notable differences. This was assessed via repeated measures MANOVAs (Room as within subject factor, eye-tracking (Y/N) as between subject factor) considering either the nine artwork evaluation scales or the 20 emotion terms. Results showed no significant main effect of wearing glasses (Evaluation $F(9,41) = 1.20, p = 0.32$; Emotion $F(20,25) = 1.07, p = 0.43$), and no significant Room $\times$ Glasses interaction [Evaluation $F(9,41) = 2.04, p = 0.06$; Emotion $F(20,25) = 0.72, p = 0.78$]. Univariate tests for each individual scale also revealed no differences between Glasses conditions (Levene's test for homogeneity of variances for all items *ns*). A repeated measures ANOVA for time spent viewing also showed no main effect of wearing eye-tracking glasses [$F(1,45) = 0.19, p = 0.66$] or a Room $\times$ Eye-tracking interaction [$F(1,45) = 0.55, p = 0.46$]. Therefore, all participants were combined in the following analyses.

General Behavioral Responses

The general responses to and appraisals of the artworks are shown in **Figures 2 and 3** (evaluations), **Figure 4** (reported emotions), and **Tables 1, 2** (reported artwork meaning). Taken together, these data show notable differences in responses to the two installation works.

Artwork Appraisals and Time Spent Viewing

Participants spent roughly twice as much time viewing *Wishes* vs. *wonders* (hereafter “*Wishes*”) ($M = 4.24$ min, $SD = 3.17$, min 1:00, max 20:10) than *Eye see you* (hereafter “*Eye*”) ($M = 1.76$, $SD = 1.56$, 0:43 to 10:52), with a *t*-test indicating that this difference was significant [$t(46) = 7.75, p < 0.001$]. This result may have partially been due to the larger dimensions of *Wishes*, but may have also been tied to the detailed murals on the three walls of the installation (see also related findings on artwork meaning, **Table 1**). We also found differences for appraisals. To analyze differences in the ratings, we conducted a repeated measures MANOVA with Room (2) as the within subject factor while including the nine evaluation scales. This revealed a significant main effect of Room [$F(9,42) = 19.42, p < 0.001, \eta_p^2 = 0.81$], and with individual univariate comparisons showing that, with the exception of “beautiful-ugly,” all scales differed significantly between rooms (all *ps* < 0.04 , with Bonferroni adjustment for multiple comparisons). Generally, as shown in **Figure 3**, *Eye* was rated more highly on the hedonic measures (happy, pleasant), and was seen as more active, whereas *Wishes*

was rated as more sad, but also more serious, meaningful, and generally more potent.

Artwork Meaning

The above differences in evaluations could also be tied to viewers' understanding of the meaning of the artworks. To analyze this, we first classified all short meaning answers into one of the three categories described in the Methods above: (1) direct or descriptive; (2) no meaning or assumed meaninglessness; (3) experience- or emotion-based descriptions. Meaning was classified, using the original German answers, by three independent raters trained in identifying the three levels. Raters showed a high level of initial agreement in their classifications, coming to the same classifications for $>95\%$ of answers; disagreements were settled by majority decision. All answers (English translations) are shown in **Table 1** (*Wishes*) and **Table 2** (*Eye*). Between rooms, there appeared to be notable differences in the use of the three types of meaning answers, with higher incidence of descriptive meanings in *Wishes* (64.7%), followed by experience-/emotion-based meaning (31.4%), and very low use of not understanding (4%). In contrast, *Eye* showed relatively less descriptive meaning (33.3%), higher experience-based meaning (35.3%) and higher lack of understanding (31.3%). A Chi square test (3 meaning types $\times$ 2 rooms) revealed these ratios to differ significantly between rooms [$\chi^2(N = 102, 2) = 16.1, p < 0.001$].

We also considered the likelihood that the same individuals would give the same type of meaning for both rooms despite the overall between-room differences. Individuals gave the same meaning type 49.0% of the time [well above a chance level of 33.3%; $X^2(51, 1) = 5.7, p = 0.02$], suggesting between-room within-participant consistency.

Meaning $\times$ Appraisals

The relation between the meaning answers and the artwork ratings (reported in **Figure 3**) was analyzed via two MANOVAs (meaning type set as between subjects factor) for each of the nine evaluation scales for each of the two rooms, conducted separately. No significant main effects were detected [*Wishes* $F(18,82) = 1.02, p = 0.45, \eta_p^2 = 0.18$; *Eye* $F(18,82) = 1.15, p = 0.32, \eta_p^2 = 0.20$]. Univariate tests for between-subjects effects for specific scales (with Bonferroni correction) revealed a significant difference in *Wishes* for “humorous-serious” [$F(2,48) = 4.45, p = 0.02, \eta_p^2 = 0.16$]. In *Eye*, differences were found for “beautiful-ugly” [$F(2,48) = 3.16, p = 0.05, \eta_p^2 = 0.12$] and “meaningful-meaningless” ($F(2,48) = 5.95, p = 0.005, \eta_p^2 = 0.20$; homogeneity of variances (Levene's test) *ns*). No significant differences were detected for time spent with the art (ANOVA, meaning category between subjects, conducted separately for each room).

Emotion

Figure 4 displays the mean reported emotions experienced in the two rooms. Once again, to analyze between-room differences we conducted a repeated measures MANOVA with Room as the within-subject factor while including all 20 emotion scales. This resulted in a significant main effect of Room [$F(20,26) = 3.87, p = 0.001, \eta_p^2 = 0.14$]. Univariate comparisons between rooms

TABLE 1 | What was the artwork meaning? Wishes vs. wonders.

Meaning type	Answer, translated from original German (participant number)
(31.4%) Experiential/emotion-based ^a	(3) To (the viewer) the pictures seem terrifying (5) It represents the cycle of life (8) I think it represents the connection between modern and classical art. The half ring is closed through the mirror and builds an unity in its reflection, also with the classical pictures on the opposite walls. Maybe it should represent a cycle that is linked to the acts of war that are shown in the pictures. (13) Modern art (optical illusion) linked to classical art. Transition. Infinity. Unity. (14) floating weightless sphere in space. No existence of space and time and gravity. The question of the persistence in life? Being exists only in the here and now? Reminds of the universe- > orbit of the sun. (16) It shows the human being as a part of the masses and it shows that its death is just one of many others and nothing special. (17) The artwork is astonishing. One feels drawn in to the situation by the mirror, like an undertow. It brings one down to earth again, how good our life in fact is, far away from battle, blood and death. But will this change in the future? (22) And again I don't know, presumably: the history of the mankind is an eternal cycle. Again and again war and destruction. (30) Sadness, calm atmosphere (32) Cycle of war. The war cannot bring peace, but only new conflicts. (35) The blurring and interaction of the link between illusion and reality. (37) I think the intention was to show how the form alters depending on the angle from which the artwork is viewed. The circle/oval changes from each perspective (41) paintings of battles were shown and I think the meaning of the mirror was to lead the recipient to believe that everything was bigger and wider than depictable on the pictures. Although I found the circle in the middle was very impressive, because it seemed to float in the air, I could not think of any meaning (42) little, the optic circle reminded me a bit of the "doom", the never-ending up and down of wars (with reference to the paintings) (43) In the battles many lose their lives and take the bullet for their country. Nature gets destroyed as well. General destruction. But also new, good things can emerge from that. (48) It plays at least with the optical illusion of a levitating object. It might advert to closed-ness or endlessness due to the imagery and the mirror.
(64.7%) Descriptive meaning	(1) detailed depiction of battles (2) moderate meaning in the art sector (6) Depiction of battles to provide documentation (7) The artworks are historical and show various battle scenes. They (the artworks) tell the visitors how and where battles were fought in the past and maybe they serve as a general reminder of the cruelties of war. (9) It shows war on several locations (10) Maybe the circle denotes a connection between the artworks (repetition of war) (11) To represent various battle scenes. How spatially extended they could be. (12) Several battles were depicted. (15) Mid-level meaning, because it depicts historical events, but without being a remarkably popular art piece. (18) paintings of various battles around the year 1700, historical, detailed (19) War (20) To me it was like an orbit (21) Seeing a piece of history (23) pictures of battles (24) The artworks were about battles at various places. (25) War theater, battle (26) Depiction of battlegrounds, tries to unify to a great artwork. (27) Depiction of a war scene. The pictures in a chronological order, beginning with gunfire etc. (28) the artist obviously had a preference for battles. I am impressed by the details of the paintings and I am wondering the whole time how long it took him to paint these artworks. Ha ha. It was surely exciting for the contemporary people to look at pictures of heroic battles. But I mostly could not figure out who was fighting whom. (29) A battle was depicted, that should be captured. (31) The presentation of various battles (historical moments), perhaps as achievements (33) It illustrates diverse battles of the 18th century. (34) Sundry presentations of battles. (36) I saw pictures of battles, these types of artworks have no meaning to me.

(Continued)

TABLE 1 | Continued

Meaning type	Answer, translated from original German (participant number)
	(38) historical meaning (demonstration of power on one side) – uplifting-floating lightness on the other one
	(39) Depictions of several battles/wars, demonstration of power/efficacy of war → Dead people, winnings, losses, changes etc.
	(40) battle fields, wars of the past
	(44) Depiction of historical battles.
	(46) Depiction of battles. Soldiers back then (in contrast to these days) were hailed as heroes, though, and thus their fall was dramatic. It looks rather heroic than war pictures nowadays, which meant to act as a deterrent.
	(47) painted war scenarios on several pictures, were imposing, but they all showed the same The artwork with the mirror (did not affect) me.
	(49) The paintings in the room depicted battles.
	(50) Battle, raid into a village, war
	(51) Showing several war-/battle strategies and depictions of “a whole”
(3.9%) No Meaning	(4) No idea
	(45) No idea

categorization based on previous argument by Pelowski and Akiba (2011), Pelowski (2015). Assignment to categories done by three independent raters with disagreement decided by majority decision.

for individual scales showing significant differences for happy, amused, and relief—all higher in *Eye*—and for feeling “focused,” “stimulated,” and “sad”—all higher for *Wishes*. Finally, looking to more cognition- or understanding-related scales, *Eye* had significantly higher reported feeling of distracted, confusion and “need to leave,” while *Wishes* had higher reported feeling of focused, insight, and “understanding the artist intention” (all $ps < 0.05$, following Bonferroni adjustment for multiple comparisons). Comparing the relation between emotion and artwork ratings and the two rooms, note also that the differences between the installations themselves would appear to explain much more of the artwork evaluations (81% as assessed via effect sizes) than the specific emotions experienced while viewing (14%).

Network Model of Emotion Terms

To investigate the emotional experience of viewers in more depth, we then used a Partial Correlation Network Model Analysis (Epskamp et al., 2012; Epskamp and Fried, 2016). This again assesses the partial correlations between items without an underlying assumption of higher-order factors. It can provide information regarding the centrality, interconnections, and specific importance of items in defining the global experience. Our main aim here was to investigate how the emotions related to each other, as assessed via their connections, as well as to consider between-room differences.

Analysis was conducted using the R package *qgraph* (Epskamp et al., 2012) with “graphical LASSO” regularization assessing the same 20 underlying emotions as above, conducted separately for each room. The hyperparameter (γ) was set to 0.2 for both rooms’ models. This determines the sparseness of the network (the number of connections), by eliminating spurious connections. Although literature is only now emerging for modeling procedures, previous publications have suggested a typical hyperparameter at around 0.25 for hypothesis-driven research (a higher total is more conservative; e.g., see Barber and Drton, 2015). However, Epskamp and Fried (2016)

also suggest that researchers should make an informed decision to adjust the parameter, typically toward a less conservative number, based on their own study conditions. Therefore, due to the exploratory nature of the study, we determined that 0.2 gave an optimal balance between conservatism while still retaining some key connections between items.

The final models are shown for the two rooms in **Figure 5**. The individual emotions are shown as circles; green lines connecting the circles indicate a notable positive partial correlation (surviving the conservative regularization procedure); red lines indicate a negative partial correlation. Thickness of lines indicates strength of correlations, with no line suggesting that a relation is minimal. The colors of the circles were assigned by ourselves and correspond to the basic expected nature of the emotion (negative, positive, cognitive/meaning-related, self-awareness). We also report the calculated “closeness” (relative centrality within the entire model, calculated as the inverse sum of shortest distances from all other emotion nodes to one focal node), “betweenness” (degree to which a node is on the shortest path between other nodes), and “strength” (sum of the weights of connections that a node has) for all emotion terms in **Figure 6**. The centrality measure of closeness also represents the inverse sum of shortest distances from all other nodes (emotions) to one focal node.

The main result appeared to be that generally more positive emotions (yellow circles) and negative emotions (red circles) were located on opposite sides of the network with a general negative correlation between these items (red lines), suggesting a low likelihood of reporting both sets of emotions in one experience (see also the generally low “Closeness” measures, **Figure 6**, for both emotion types). At the same time, mediating these groups were more cognitive responses (blue circles)—confusion, insight, and especially self-awareness. In both artwork cases, in order to move between a negative and a more positive responses (tracing a pattern along green connecting lines), required a movement through self-awareness. Self-awareness, in turn, fed into “understanding intention,” relief, and insight.

TABLE 2 | What was the artwork meaning? Eye see you.

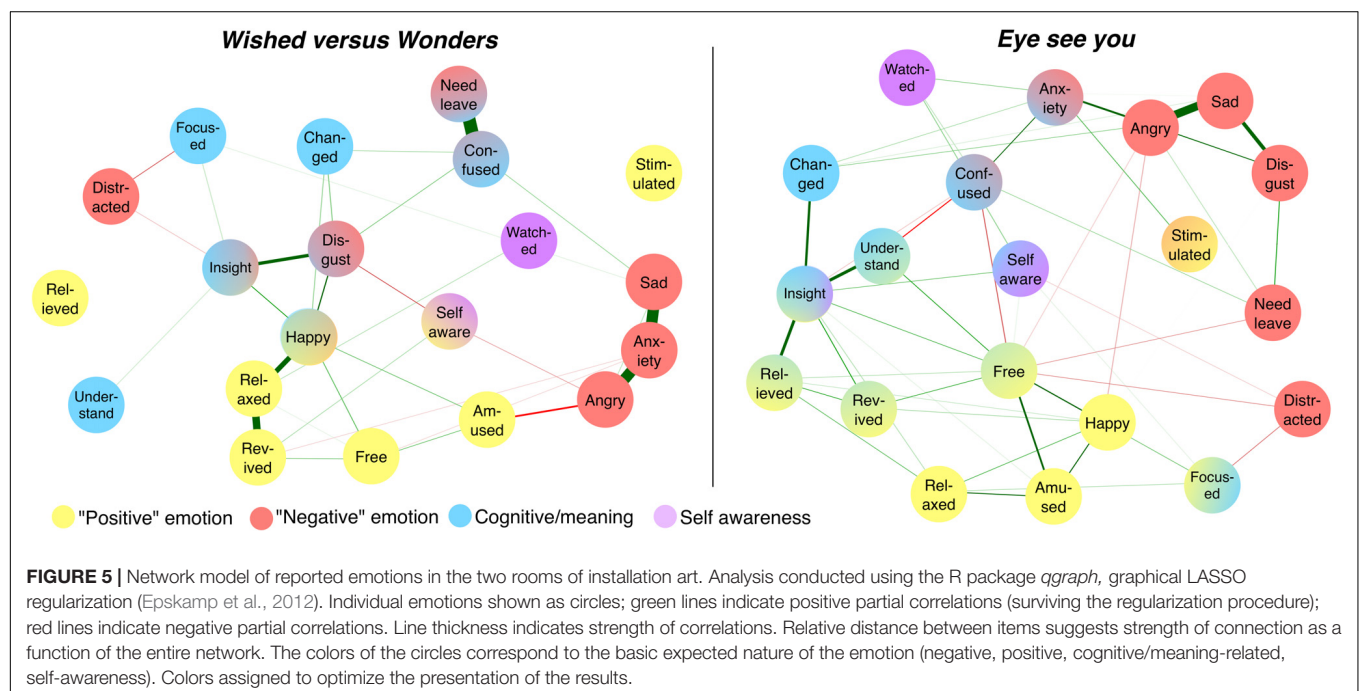
Meaning type	Answer (participant number)
(35.3%) Experiential/emotion-based	(3) double meaning of life (5) To learn and understand other perspectives (6) magnificent ambience (7) The sort of “sun” bathes the room in an orange, very bright but pleasant light. Through the mirror one can see oneself everywhere a 100 fold. The meaning is very abstract and can certainly be interpreted in different ways. Maybe it wants to increase the self-confidence or just show that creating a pleasant and free feeling in such a small room is possible (8) I think it is supposed to show the limits of our visual perception and that we can still win by means of these new impressions. By playing with the complications and perceiving unfamiliar things, which are hard to interpret at first sight (like the other color perception) (11) Creating a world without color (14) seeing eye. Sees itself through all the mirrors. Insight question and recurring/continual. To understand oneself? (16) It resembles a sun and makes one feel more cheerful and more alert. (17) The angels on the ceiling look like they are going to put the flower wreath on one's head. Impressive wall ornaments and one can see how impressive optical illusions have always been to people – the mirrors seem to reach ad infinitum. They carry one into a different world. (21) Humor. The artwork was very funny to me. (29) The warm light with the beautiful room composition made one happy (30) calming effect (34) Very garish light, it was hard to concentrate on the room, strongly distracted from the mirror image. (35) An extreme representation of width and infinity of spaces. Every apparent barrier (walls etc.) can bear additional space. (40) The influence of the yellow light on the experience of a joyful room. (42) strong light effect, oneself as “sinking” (head in the middle of the lamp with royal surroundings (46) One sees herself in a mirror with (awesome) light reflections that makes you look fancy. Congenial to the fancy room with the mirror. And in general congenial to the whole Winterpalais – apparently Mr Prince Eugen went for fancy things – and because it is still cool, a mirror was positioned. In this way we are all a bit prince/princess when we sparkle so nicely in our reflection. (47) Gorgeous ceiling painting with angels, it conveyed calmness, yellow light, that illuminated the room, that I found unpleasant, very beautiful ornament on doors and walls
(33.3%) Descriptive meaning	(1) reflection (2) dubious meaning of the positioned installation, rest of the room with very conspicuous beautiful adornments (13) Highlighting the room through light. Highlighting certain aspects of the room (18) splendid, historical (19) It is some Christian motive, but I do not know enough about this subject matter, to construe it in a right way. (20) It looks like one big eye, which can see everything because it lights the whole room. Perhaps a depiction of an “all-seeing eye” (23) golden room with angel (25) splendor, depiction of wealth, affluence (27) Abstract rebuilding of a sun (28) I felt watched by the angels with their dark eyes. Besides, in my opinion it shows that nobles in the past spent far too much money on art, just to boast. Apart from that I cannot say much about it. (31) I think it was just a well-lighted embellishment, that should demonstrate liking and pompous wealth, especially the flowers and tea cups/tankards cause wellbeing (38) a “beautified” picture of oneself (in the mirror) (39) Representation of sun, reflection in the mirrors (41) To me it looked like a living room with chimney and pleasant pictures of angels and flowers. The light had no meaning to me in this context and was incongruous. (43) Plenty of gold and the big lamp that spends “warmth”. The mirrors were very cool (similar to the hall of mirrors) (50) round reflecting sun (51) I associate the following with the artwork: warmth, sun, gold, beauty.
(31.4%) No Meaning	(4) no idea (9) ? (10) I cannot answer this question, to me it was not obvious. At least the light was pleasant. (12) No particular meaning, ornaments on wall and ceiling, bathed in yellow light (15) no big (meaning)

(Continued)

TABLE 2 | Continued

Meaning type	Answer (participant number)
	(22) I don't know. Maybe it was about highlighting something, drawing attention to something, to gain center stage?
	(24) the meaning is unknown
	(26) no idea
	(32) I don't know
	(33) Unfortunately, I am not aware of the intention of the (art).
	(36) I looked at the room, to me it had no special meaning and I did not feel aroused/stimulated by the room
	(37) I don't know
	(44) no idea
	(45) no idea
	(48) I heard how other visitors were talking about an "eye". But I may not have noticed.
	(49) Unfortunately, I am not absolutely sure about the meaning. There was a big lamp positioned that dominated the room.

categorization based on previous argument by Pelowski and Akiba (2011), Pelowski (2015). Assignment to categories done by three independent raters with disagreement decided by majority decision.



This might especially be connected to the path toward an experience-based meaning evaluation as seen in *Eye*. The individual connections among the emotions can also be taken into account. In addition to line thickness and red/green color, the significance of such connections within the overall model is also evidenced by the measure of "Strength." As can be seen in the **Figures 5** and **6** in *Wishes* disgust, insight and happy are strongly connected. This suggests that people may initially feel disgusted by the artwork but then as they gain insight feel happy. Since happy is then strongly connected to relaxed (and relaxed to revived), the feeling of happiness may then persist and lead the person to feel relaxed and revived. Though the art experience may have started off negatively with disgust, as people gained insight they may have felt like they gained some understanding (insight) and had an overall positive experience.

In *Eye*, we do not find the above pattern. Even though Insight and revived are still connected, this connection does not run through happy or relaxed. In this case, insight mainly connects to feeling changed and feeling a sense of understanding of the art. What may be happening with *Eye* is that people gain insight into the art which makes them feel they understand it, as well as that they feel changed by it, and finally that they feel a sense of relief. The experience seems therefore more "cognitive" with a steady pattern of understanding the work than in *Wishes* where insight seems to work more as a pathway to go from feeling bad (disgusted) to feeling good (happy, relaxed, and revived). However, in both *Wishes* and *Eye*, sad, anxiety, and angry are also highly connected to each other but minimally connected to other nodes. Thus, if a viewer feels one of these three emotions, they are likely to feel all three during their experience, but are unlikely to report feeling anything else.

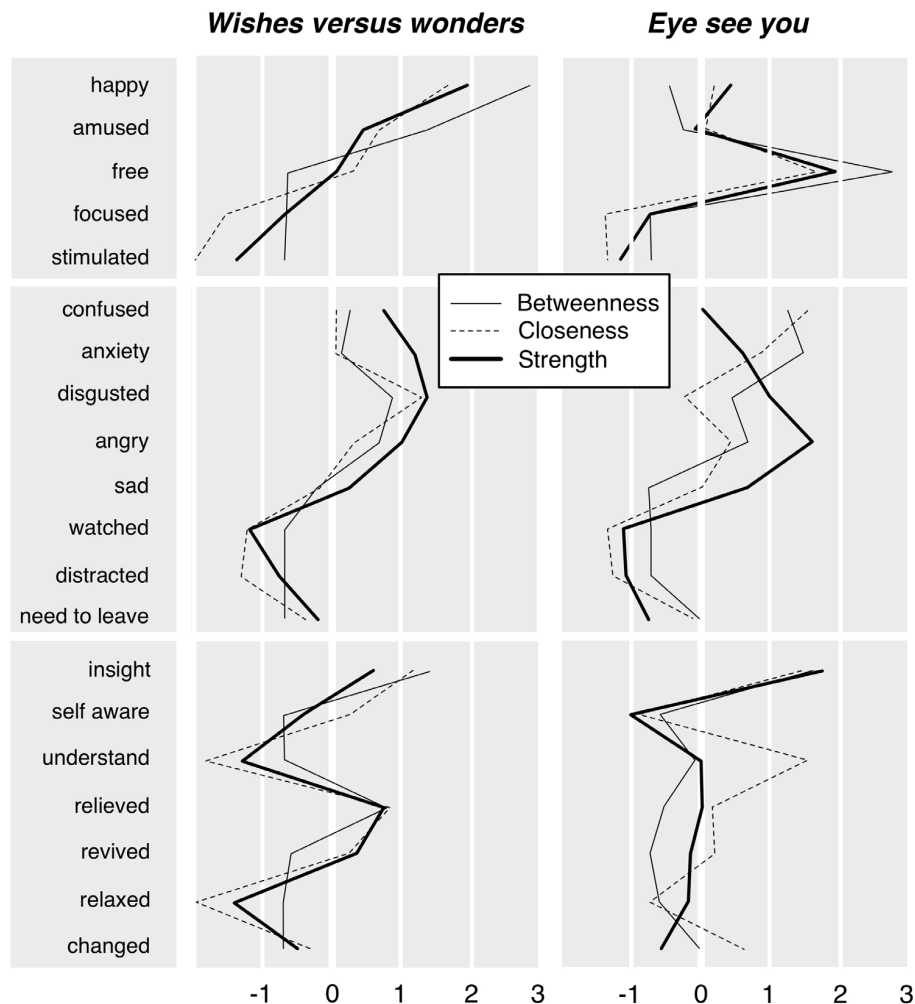


FIGURE 6 | Closeness, betweenness, and relationship strength of emotion terms in the network model.

Emotion × Meaning

We also assessed the potential differences in reported emotion between the three meaning types (MANOVAs conducted for each of the 20 scales × the two rooms, separately). Analysis again revealed no significant main effects for both rooms. Additionally, for *Wishes*, no notable differences in univariate tests of between-subjects effects within specific emotions were detected. For *Eye*, tests for between-subjects effects revealed one significant difference (with Bonferroni correction, $\alpha = 0.0025$) regarding “having understood the artist intention” [$F(2,48) = 9.18$, $p < 0.001$], which was highest with descriptive meanings, slightly lower for experience based, and lowest for no meaning. Without Bonferroni correction, differences were also detected for feeling “changed” [$F(2,48) = 5.92$, $p = 0.005$], and “insight” [$F(2,48) = 3.97$, $p = 0.02$], which were generally higher for an experience-based meaning. Differences were also found for “confused” [$F(2,48) = 5.91$, $p = 0.005$], higher for individuals giving “no meaning” answers (tests for homogeneity of variance for univariate comparisons, *ns*).

Eye-Tracking

Fixations and Fixation Duration

BeGaze software (version 3.5.101) was used for semantic gaze mapping. This allows frame-by-frame assessment and manual coding of fixation counts and fixation durations as they relate to researcher-defined regions of interest (ROI) within the recorded video. For the purpose of our analysis, and in order to give a general overview of where participants were looking in each room, we identified three main ROIs for each room. These included: (1) the artworks (defined as any component that had been installed by the artist); (2) the ambient environment (including walls, ceiling and ornamentation); and (3) the self (either viewed directly or as a reflected image).

Descriptive statistics for both rooms—with fixation counts, raw fixation times, fixation durations, and percentage of looking time as they relate to viewing either walls/ceiling, self, or artwork—are displayed in **Table 3**. Participants tended to look at the *Wishes* installation as well as all of its components longer (fixation time) than with *Eye*. They similarly showed more

TABLE 3 | Eye-tracking results.

	Fixation time (ms)		Fixation count		Fixation duration (ms)
	Mean (SD)	% Total	Mean (SD)	% total	
<i>Wishes vs. wonders</i>					
Artwork	27,299.6 (30,960.2)	34.5% (36.0)	104.8 (106.7)	25.4% (29.1)	221.5
Walls/ceiling	108,459.2 (115,086.3)	66.3% (35.9)	471.9 (464.5)	73.1% (29.9)	204.1
Self	3,071.6 (3,247.2)	4.4% (5.8)	10.7 (10.8)	2.2% (1.9)	295.6
Total	131,781.0 (108,643.9)		583.8 (431.2)		212.8
<i>Eye see you</i>					
Artwork	8155.9 (15,003.7)	11.6 (12.14)	28.5 (42.2)	11.5% (12.0)	225.3
Walls/ceiling	48,869.5 (44,718.4)	83.8% (13.4)	208.0 (172.7)	85.7% (12.9)	225.0
Self	3,303.8 (2,946.8)	7.0% (6.1)	5.4 (6.8)	2.8% (3.2)	378.4
Total	59,227.9 (54,724.7)		241.9 (189.2)		235.5

TABLE 4 | Correlation of artwork ratings and percentage of fixation time on artworks.

	Fixation time			Percent of fixations		
	Artwork	Wall	Self	Artwork	Wall	Self
<i>Wishes vs. wonders</i>						
Beautiful: ugly	-0.271	-0.264	0.212	0.225	-0.176	0.407
Active: passive	0.129	-0.377	-0.289	0.136	-0.106	-0.072
Pleasant: unpleasant	-0.226	0.258	0.368	-0.202	0.178	0.167
Happy: sad	-0.488*	0.359	0.271	-0.416	0.419	0.171
Strong: weak	-0.119	-0.242	-0.487	0.170	-0.094	-0.056
Intimate: distant	0.153	-0.012	0.717*	-0.131	0.080	0.309
Meaningful: meaningless	0.186	-0.508*	-0.329	0.457	-0.409	0.102
Serious: humorous	0.598*	-0.390	-0.194	0.634*	-0.574*	0.345
Potent: impotent	0.177	-0.319	-0.130	0.343	-0.262	0.229
<i>Eye See you</i>						
Beautiful: ugly	0.012	0.113	0.107	0.102	-0.136	-0.024
Active: passive	-0.284	0.153	0.008	-0.288	0.196	0.333
Pleasant: unpleasant	-0.045	0.141	0.127	-0.019	0.008	0.181
Happy: sad	0.133	-0.158	0.265	0.144	-0.272	0.309
Strong: weak	-0.267	0.208	-0.035	-0.320	0.323	0.068
Intimate: distant	0.231	-0.057	-0.031	0.443*	-0.373	-0.336
meaningful: meaningless	-0.100	0.169	-0.144	0.019	-0.034	-0.018
Serious: humorous	0.167	0.279	0.140	0.257	-0.296	0.028
Potent: impotent	-0.065	0.219	0.015	-0.069	0.027	0.025

Negative correlation implies relation to leftmost term on bipolar scale. Positive correlation implies relation to right term. *Indicates significance at $p \leq 0.05$, uncorrected. No correlations maintained significance following Bonferroni correction for multiple comparisons (familywise corrected alpha = 0.0028).

fixations. Following typical eye-tracking studies, these aspects were also themselves highly correlated (total fixations and total fixation time, *Wishes*, $r = 0.981$; *Eye*, $r = 0.968$). The individual fixation durations for *Eye* tended to be longer, although paired sample t tests for duration of looking for the different elements showed no significant differences.

When considering the percentage of total looking time/fixations devoted to artwork, self, or room elements (wall/ceiling), participants also tended to look at the artwork in *Wishes* (defined as both the room-length mirror and/or the embedded ring) roughly three times more (34.5%) than in *Eye* (11.5%). The majority of other looking duration/fixations were devoted to the environment, followed by roughly 5% of time

spent looking at the self. Univariate t -test comparisons for percentage of duration/fixations devoted to wall, self, or art (with Bonferroni correction, adjusted alpha = 0.0084, based on six comparisons) showed significant between-room differences for fixation time with the artwork [$t(17) = 3.50$, $p = 0.003$] and the wall [$t(16) = 3.02$, $p = 0.008$]. Fixation duration on self and all comparisons of fixation number comparisons were *ns*.

Eye-Tracking Results × Appraisals, Emotion, and Meaning

Assessment of correlations between fixation durations on the artworks, self, or room and the viewers' appraisals (shown in **Table 4**) did not show any significant relations for either room

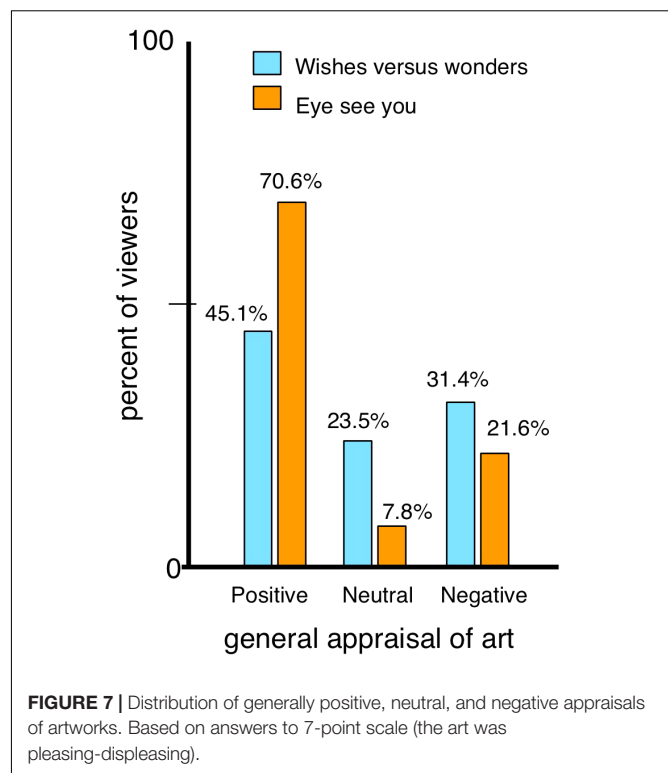
following Bonferroni correction. Without correction, trends were detected for *Wishes* regarding fixation duration on the artwork and “happy-sad” and “seriousness-humorousness.” Here, more focus on the art, and presumably not on the battlefields, resulted in a happier but less serious appraisal. This latter assessment of seriousness was also reflected in the percentage of attention given to either art or room (see Table 4). A notable correlation was also found between fixation duration on the room and feeling the art to be more meaningful. Moreover, focusing more on the self correlated to assessing the art as more distant. No significant/notable patterns were uncovered for *Eye*. For emotion as well, no significant correlations were detected for either room following correction for multiple comparisons.

An analysis of the relationship between the eye-tracking results and the meaning categories (ANOVAs for meaning type, conducted separately for the six fixation durations and percent of fixation duration assessments described above) did not lead to significant or notable findings. Note that the eye-tracking data was only conducted on half of the participants, leading to a quite small sample, which may not have detected more subtle relationships.

Positive Vs. Negative Experiences and Related Factors

To assess the factors positively influencing the viewers' experiences in more depth, we compared individuals who reported a positive vs. a negative experience in the two rooms. This was done by considering answers to the artwork evaluation “pleasing-displeasing,” considered as a general assessment of preference or enjoyment of the art (Silvia, 2005b). A breakdown of responses (separated between generally “pleasing,” score of “1–3”; “displeasing,” “5–7”; and “neutral,” score of “4” on 7-point scale) are shown for both rooms in Figure 7. *Eye* had more individuals giving a generally positive rating (70.6%) vs. *Wishes* (45.1%). *Wishes* in turn had more negative (31.4%) and neutral (23.5%) responses as compared to *Eye* (pleasing = 21.6%; neutral = 7.8%). This also suggested less ambivalence, as surmised from neutral scores, with *Eye*.

Table 5 shows the correlation between the “pleasing-displeasing” and the other artwork ratings. As expected, the ratings were generally highly correlated with other hedonic measures such as happiness and beauty. There was also general negative correlation with scales tied to dynamism (potency, activity), and especially “serious-humorous” (although not significant after Bonferroni correction). The same general results were also found when participants were broken into the three groups of those giving a positive, neutral or negative answer to liking the art. These were assessed with a set of two MANOVAs (pleasing-neutral-displeasing Appraisal as between subjects factor) conducted with all eight remaining scales for each room, separately. These showed a significant main effect of Appraisal in both rooms [*Wishes*, $F(\text{Pillai's Trace}, 16,84) = 2.58$, $p = 0.003$, $\eta_p^2 = 0.33$; *Eye*, $F(\text{Pillai's Trace}, 16,84) = 1.82$, $p = 0.04$, $\eta_p^2 = 0.26$]. Tests of between subject effects for each room revealed significant differences in *Wishes* for beautiful, happy, and humorous (higher for good vs. the other appraisals, all



$ps < 0.03$, corrected for multiple comparisons). Similarly, for *Eye*, differences were detected for more “beautiful” and more “happy,” as well as for “strong-weak” and “intimate-distant,” with the negative appraisal coinciding relatively more with the leftward terms on the scale (all $ps < 0.05$, corrected).

For emotion (see also Figure 8), a similar set of two MANOVAs were conducted for all 20 scales in each room, separately. Once again, these revealed a significant main effect in each room [*Wishes*, $F(\text{Pillai's Trace}, 40,58) = 1.76$, $p = 0.025$, $\eta_p^2 = 0.55$; *Eye*, $F(\text{Pillai's Trace}, 40,52) = 1.79$, $p = 0.025$, $\eta_p^2 = 0.58$]. In *Wishes*, individual tests of between subject effects showed differences for sad, angry (both higher for negative evaluation compared to positive evaluation), and revived (higher for positive, all $ps < 0.025$, corrected). In *Eye*, differences

TABLE 5 | Correlation between “pleasing-displeasing” artwork assessment and other art ratings.

	<i>Wishes vs. wonders</i>	<i>Eye See you</i>
Beautiful: ugly	0.398	0.601*
Active: passive	−0.286	0.198
Happy: sad	0.602*	0.629*
Strong: weak	−0.272	0.251
Intimate: distant	−0.302	−0.366
Meaningful: meaningless	−0.049	0.283
Serious: humorous	−0.328	−0.241
Potent: impotent	−0.077	0.149

*Significant correlation following Bonferroni correction for multiple comparisons (familywise). Adjusted alpha = 0.0031 for 16 comparisons.



FIGURE 8 | Comparison of Mean reported emotions among individuals giving positive, neutral, and negative general appraisals of artworks. * and dotted line indicate significance at $p < 0.05$ (see also legend above).

were detected again for angry (higher for negative), happy (higher for positive), as well as for confused and need to leave (higher for negative), for relieved, relaxed, and free (higher for positive), and for focused (highest for neutral but also relatively high for positive; all p s < 0.05 , corrected). Interestingly, in both rooms it appeared that those ending with a neutral

art rating tended to show the highest focus, possibly due to the need to deal with an ambivalent semantic or experiential situation.

Interestingly, when we looked to the general pleasing/displeasing artwork appraisals and assessed artwork meaning, we did not find a notable pattern. Rather, as shown in **Table 6**, in both bad and good appraisal cases, individuals tended to rely most heavily on descriptive meaning. Contrary to our working hypothesis, emotion-based meaning did not appear to tie to more positive assessments in either installation. Further, giving an answer that the art had no meaning or that an individual could not understand the art also did not appear to necessarily lead to negative assessments. This was especially true for *Wishes*, where no individuals giving a “no meaning” response coincided with a displeasing assessment (X^2 for meaning types $\times$ outcome type for both rooms: *ns*).

Time spent in the rooms and eye-tracking data (raw fixation times and percent of time spent looking at artwork, room, or self) also showed few significant or notable differences. When breaking down the ratings into the three groups of those giving a good, neutral, or bad rating, only one result showed significance (with or without correction). Fixation time on the room with *Eye* [$F(2,18) = 8.79$, $p = 0.002$] was higher among those giving a bad

TABLE 6 | Distribution of types of answers to “What did the artwork mean?” when viewers appraise art as generally good, neutral, or bad.

% Meaning type usage for good, bad, neutral artwork appraisals			
	Descriptive meaning (%)	No meaning (%)	Emotion-/experience-based meaning (%)
<i>Wishes vs. wonders</i>			
“good” appraisal	56.5	8.7	34.8
Neutral appraisal	75.0	0.0	25.0
“bad” appraisal	68.8	0.0	31.3
<i>Eye see you</i>			
“good” appraisal	33.3	27.8	38.9
Neutral appraisal	50.0	25.0	25.0
“bad” appraisal	27.3	45.5	27.3

appraisal of the art. Neutral and good appraisals showed roughly the same looking time.

DISCUSSION AND CONCLUSION

This study considered, in an exploratory fashion, the experience of installation art encountered in the museum. We investigated whether this unique art medium—which has become increasingly common yet is often difficult for both the interested researcher and for the novice viewer to assess or appreciate—can be meaningfully evaluated in terms of the emotional and semantic experience. We further considered how experiences may differ between specific artworks or in relation to viewers' positive/negative evaluations, as well as more technical questions related to how individuals approach the spaces, namely, as an integration of room and installation or as stand-alone pieces. Finally, we evaluated the efficacy of our self-report approach, collection and organization of reported emotions, as well as the use of eye tracking.

Regarding our first research question: Do the two artwork installations produce distinct emotional or evaluative responses that may differ within-subjects, and could these be captured? We found that indeed the two installation examples produced distinct emotional and evaluative responses. Participants, even when viewing both rooms in a counterbalanced order, tended to have a more positive and emotional experience with *Eye see you*, with them reporting significantly higher happiness, feeling of pleasantness, and active nature of the art. In contrast, for *Wishes vs. wonders*, participants reported more sadness, less pleasantness, but also more meaningful, potent, intimate and serious encounters, as well as more feeling of understanding the intention of the artist. These differences appeared to be due to the specific installation elements in conjunction with the pre-existing room features, as well as at least partially the artist intention. In *Eye*, the warm light cast by the artificial sun tended to drive the interpretations and emotions, whereas in *Wishes*, the paintings tended to lead to a focus on war and death that were tied to the presence of battlefield depictions. These were part of the pre-existing environment, but further highlighted by the artist's intervention.

These differences could themselves suggest two different responses to the art—that is, between a more “emotional/harmonious” and a more “cognitive/intellectual” engagement. Such a dichotomy in potentially positive but qualitatively different reactions has been recently posited in models of art processing (Pelowski et al., 2017b). This suggestion is also supported by the emotion network model, which shows a more central position for basic positive emotions such as happiness in *Eye see you* and a more central position for cognitive responses—insight, feeling changed, confusion—in *Wishes vs. wonders*. This is also suggested by the meaning interpretations. Visitors to *Wishes* tended to focus much more on the semantic level. Almost every participant in the descriptive category tied the meaning to the battle scenes. This result would also partially explain the above differences in seriousness and sadness attributed to the art.

The focus on room and art elements also suggests an answer to our sub-question of whether viewers would appreciate the installed elements as stand-alone artworks or as components that interact with, and may play off of, the ambient environment. Viewers did appear to specifically consider the interplay of these factors when making evaluations or as part of their perceptual/emotional experience. In both cases (and especially with *Wishes*) viewers tended to interpret the artist-installed portions in light of the ambient environment, arriving at a final interpretation that more or less successfully combined both elements. This suggests that—even if this approach may be a relatively newer way of engaging art made necessary with installation and other contemporary examples—individuals with even limited art knowledge seem capable of this response. Such interactions can also be empirically captured. Viewer interpretations of the installations also tended to align with existing art critical and curatorial discussions, showing that lay viewers could indeed pick up on and appreciate the often ambiguous or esoteric nature of the installations.

Especially in *Wishes vs. wonders*, it could also be suggested that the rather traditional, semantic content facilitated interpretation—thus, the very low incidence of not understanding. Note, however, for those who mentioned the mirror and ring, this was usually in the context of not truly grasping their purpose or trying to connect these to the endless cycle of war or the human condition. Those suggesting an experience- or emotion-based interpretation in turn suggested sadness, terror, calm, as well as infinite or forced questioning of perception or life meaning. In *Eye*, participants providing descriptive answers focused primarily on a “sun,” or provided a recap of the installation elements and/or the ornate walls and ceiling. Those focusing on experience/emotion noted feelings of warmth, alertness, joy, but also (presumably because of the mirror and bright light) forced consideration of vision. (No viewer mentioned melancholy, which had been suggested as a potential reaction by the artist). The higher incidence of “no meaning” answers presumably tied to this room's lack of more obvious mimetic elements. However, when looking at those individuals who had a generally positive vs. negative art experience, we did not find that the respective meaning interpretations—dividing between descriptive and more emotion-/experience-based answers—showed significant distribution differences. This in itself is quite interesting, as it tends to work against the argument that such latter interpretations are the “proper” way to enjoy installation art. It also is contrary to our own past findings in museum studies with abstract art (Pelowski, 2015) that meaning interpretations which focus on the experience rather than a recount of features or descriptive aspects tended to correlate with higher enjoyment and engagement.

The lack of connection between meaning assessments and ratings could be due to our use of lay viewers who often base their aesthetic judgment on emotionality (Leder et al., 2012, 2014). It could also be that the focus on the content rather than emotion for many viewers related to more overt (if largely subconscious) strategies used by participants to, for example, distance themselves from the emotion related to these depictions

by basing their interpretation toward descriptions. Similar results have been found by Kemp and Cupchik (2007); see also Cupchik (2013) who presented viewers with a range of positive and negative paintings and found that viewers who wanted to see the paintings with negative themes a second time did so primarily because they evoked thoughts rather than feelings. Whereas in *Eye*, where there was both less readily graspable semantic content and generally more positive emotions, viewers may have been more likely to focus on the ambiguity and/or their emotional experience. It may also be that our sample did not detect a tie to meaning due to the limited size.

At the same time, although not detected in the quantification of written meaning answers, the patterns found in the network models of emotion, especially for *Eye*, did tend to suggest at least some evidence for general evolutions or importance of reflection and insight. Again, in the relationship between emotion reports we found that positive and negative emotions tended to be separate, however, with the possibility for a movement between these responses mediated by confusion and especially self-awareness, which also fed into “understanding intention,” relief, and insight. Such a progression has again been specifically suggested in models of transformative art (Pelowski and Akiba, 2011). Such interactions between specific emotions are obvious key candidates for future more in-depth research. Studies could especially attempt to identify which artwork aspects might trigger which emotions or interpretations within a flow of experience. This also raises the potential importance of network models in study with art.

The emotion results, even in their present level of complexity, did on the other hand support our second research question regarding ability to divide reactions to the rooms into generally positive or negative experiences. While the approach to meaning was not a major driver of appraisal, nor was understanding, we did find that the general valence of emotions experienced showed the most important tie to the general positive/negative evaluation. In both rooms, those who evaluated the art as happy tended to have a positive viewing experience, whereas those reporting sadness had generally negative evaluations. Although this finding is more facile than we had expected, it may be an important point of departure for future research. Certainly, previous studies have found similar tie between positive/negative emotional valence and appraisals of art. Such a focus on positive feelings might be a useful candidate for setting up future museum programs or art education. It is also important to note that positive and negative emotions might also occur at the same time (e.g., Hosoya et al., 2017). These aspects, especially due to their emphasis on induced emotions, would also be a key candidate for future installation art research, and might be explored by, for example, network models in tandem with meaning analysis.

Finally, we also further supported the use of mobile eye tracking for museum studies, as there was no evidence that such use changed the emotional experience or appraisals of the viewers. This is in keeping with other real-world art viewing studies which also showed similar results (Mitschke et al., 2017). At the same time, although looking patterns did differ between artworks, they did not appear, at least at the level of our present

assessment, to play a notable role in differentiating the nature of the outcomes such as other emotions, types of meaning, or generally positive/negative appraisals. This also is in line with previous study by Gartus et al. (2015), which also considered art in conjunction with background context (gallery space vs. street backdrops) and which found that longer artwork viewing durations did not necessarily tie to more positive evaluations. Note again, however, that our sample was split in half for the eye-tracking manipulation (for experimental control). Thus, we may not have had sufficient power to detect subtle differences, and this topic remains a largely under-explored avenue for future empirical work.

This study also has other caveats and questions for future research. Future analyses should include larger samples, a wider range of art, and more focused hypotheses. Note also that we included a handful of particularly salient items. There are a wealth of other factors—more specific movement, body responses, other senses, that might and should be considered. Future study may also assess which of the above findings can generalize to all or most installation art or *in situ* art study examples. That said, we would argue for some key implications and conclusions that can be drawn from this research. This includes the ability to meaningfully assess people’s experiences of installation art and to uncover art/person differences. It is clear that many viewers can and do have meaningful reactions to installation art, driven largely by their generally positive or negative emotions, rather than where they look or their interpretations of meaning. We also thus support the sensitivity or efficacy of the included measures, laying out a paradigm that might be employed in future research. We also show that researchers should further examine the experience of installation artworks as they provide optimal contexts for more ecologically valid investigations of the experience of art.

AUTHOR CONTRIBUTIONS

MP, HL, VM, EV, TB, and AH-A contributed to the study design and the execution of the research study. MP, VM, GG, ES, EV, and TB contributed to the data analysis. MP, HL, VM, GG, ES, and PT contributed to the writing of the paper.

FUNDING

This project was supported by a grant to MP and HL by Marie Skłodowska-Curie (MSCA-IF-2014-EF: Individual Fellowships, 655379).

ACKNOWLEDGMENTS

We would like to thank the Belvedere museum and its staff for the wonderful support and unique access, as well as its director, and the artist, Olafur Eliasson. Their collaboration and willingness to work with us made this study possible, as well as an extremely rewarding experience for the authors. Data from the study is available from the authors upon request.

REFERENCES

- Altman, D. G., and Bland, J. M. (1995). Statistics notes: the normal distribution. *BMJ* 310:298. doi: 10.1136/bmj.310.6975.298
- Augustin, M. D., Wagemans, J., and Carbon, C. C. (2012). All is beautiful? Generality vs. specificity of word usage in visual aesthetics. *Acta Psychol.* 139, 187–201. doi: 10.1016/j.actpsy.2011.10.004
- Barber, R. F., and Drton, M. (2015). High-dimensional Ising model selection with Bayesian information criteria. *Electron. J. Stat.* 9, 567–607. doi: 10.1214/15-EJS1012
- Belvedere and Thyssen Bornemisza Art Contemporary (2015). *Olafur Eliasson: Baroque Baroque with Works from Thyssen-Bornemisza Art Contemporary (TBA21), Vienna and Juan & Patricia Vergez private Collections, Buenos Aires [Press Release]*. Available at: https://enfilade18thc.files.wordpress.com/2015/11/press_release_oe_long_final.pdf
- Berlyne, D. E. (ed.). (1974). *Studies in the New Experimental Aesthetics: Steps Toward an Objective Psychology of Aesthetic Appreciation*. Washington, DC: Hemisphere Publishing Corporation.
- Bishop, C. (2005). *Installation Art: A Critical History*. London: Tate Publishing.
- Bourdieu, P. (1979/1984). *Distinction: A Social Critique of the Judgment of Taste (R. Nice, Trans.)*. Cambridge, MA: Harvard University Press.
- Brieber, D., Nadal, M., and Leder, H. (2015). In the white cube: museum context enhances the valuation and memory of art. *Acta Psychol.* 154, 36–42. doi: 10.1016/j.actpsy.2014.11.004
- Brieber, D., Nadal, M., Leder, H., and Rosenberg, R. (2014). Art in time and space: context modulates the relation between art experience and viewing time. *PLoS One* 9:e99019. doi: 10.1371/journal.pone.0099019
- Brown, S., and Dissanayake, E. (2009). “The arts are more than aesthetics: neuroaesthetics as narrow aesthetics,” in *Neuroaesthetics*, eds M. Skov and O. Vartanian (Amityville, NY: Baywood), 43–57.
- Carroll, J. B. (1959). Review of the measurement of meaning. *Language* 35, 58–77. doi: 10.2307/411335
- Carstensen, L. L., Pasupathi, M., Mayr, U., and Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life span. *J. Pers. Soc. Psychol.* 79, 644–655. doi: 10.1037/0022-3514.79.4.644
- Cela-Conde, C. J., Agnati, L., Huston, J. P., Mora, F., and Nadal, M. (2011). The neural foundations of aesthetic appreciation. *Prog. Neurobiol.* 94, 39–48. doi: 10.1016/j.pneurobio.2011.03.003
- Cross, E. S., and Ticini, L. F. (2012). Neuroaesthetics and beyond: new horizons in applying the science of the brain to the art of dance. *Phenomenol. Cogn. Sci.* 11, 5–16. doi: 10.1007/s11097-010-9190-y
- Cupchik, G. C. (2013). “I am, therefore I think, act, and express both in life and in art,” in *Art and Identity: Essays on the Aesthetic Creation of Mind. Consciousness, Literature the Arts*, eds T. Roald and J. Lang (Amsterdam: Rodopi), 67–92.
- Dewey, J. (1980). *Art as Experience*. New York, NY: Perigee.
- Eaton, M. M., and Moore, R. (2002). Aesthetic experience: Its revival and its relevance to aesthetic education. *J. Aesthet. Art Crit.* 36, 9–23.
- Ekman, P., Friesen, W. V., and Ancoli, S. (1980). Facial signs of emotional experience. *J. Pers. Soc. Psychol.* 39, 1125–1134. doi: 10.1037/h0077722
- Eliasson, O. (2006). *Eye See You [Installation]*. Vienna: Winter Palace of Prince Eugene of Savoy. doi: 10.1037/h0077722
- Eliasson, O. (2015). *Wishes Versus Wonders [Installation]*. Vienna: Winter Palace of Prince Eugene of Savoy.
- Eliasson, O. (2017). *Eye See You [Personal website]*. Available at: <http://olafureliasson.net/archive/artwork/WEK100607/eye-see-you>
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., and Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *J. Stat. Softw.* 48, 1–18. doi: 10.18637/jss.v048.i04
- Epskamp, S., and Fried, E. I. (2016). *A Primer on Estimating Regularized Psychological Networks*. Available at: <http://arxiv.org/abs/1607.01367> [accessed November 06 2017]
- Fechner, G. T. (1876). *Vorschule der Aesthetik*. Leipzig: Breit- kopf & Härtel.
- Field, A. (2009). *Discovering statistics using SPSS*, 3 Edn. London: SAGE publications Ltd.
- Gartus, A., Klemer, N., and Leder, H. (2015). The effects of visual context and individual differences on perception and evaluation of modern art and graffiti art. *Acta Psychol.* 156, 64–76. doi: 10.1016/j.actpsy.2015.01.005
- Goldie, P., and Schellekens, E. (2009). *Who's Afraid Of Conceptual Art?* Abingdon: Routledge.
- Griswold, W., Mangione, G., and McDonnell, T. E. (2013). Objects, words, and bodies in space: Bringing materiality into cultural analysis. *Qual. Sociol.* 36, 343–364. doi: 10.1007/s11133-013-9264-6
- Heidenreich, S., and Turano, K. (2011). Where does one look when viewing artwork in a museum? *Empir. Stud. Arts* 29, 51–72. doi: 10.2190/EM.29.1.d
- Hosoya, G., Schindler, I., Beermann, U., Wagner, V., Menninghaus, W., Eid, M., et al. (2017). Mapping the conceptual domain of aesthetic emotion terms: a pilot-sort study. *Psychol. Aesthet. Creat. Arts* 11, 457–473. doi: 10.1037/aca0000123
- Husslein-Arco, A., and Habsburg, F. (eds) (2015). *Olafur Eliasson Baroque Baroque*. Vienna: Belvedere.
- Kapoula, Z., Adenis, M.-S., Lê, T.-T., Yang, Q., and Lipede, G. (2011). Pictorial depth increases body sway. *Psychol. Aesthet. Creat. Arts* 5, 186–193. doi: 10.1037/a0022087
- Kemp, S. W. P., and Cupchik, G. C. (2007). The emotionally evocative effects of paintings. *Vis. Arts Res.* 33, 72–82.
- Kranjec, A. (2015). Conceptual art made simple for neuroaesthetics. *Front. Hum. Neurosci.* 9:267. doi: 10.3389/fnhum.2015.00267
- Kumata, H., and Schramm, W. (1956). A pilot-study of cross-cultural meaning. *Public Opin. Q.* 20, 229–238. doi: 10.1086/266611
- Leder, H., Belke, B., Oeberst, A., and Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *Br. J. Psychol.* 95, 489–508. doi: 10.1348/0007126042369811
- Leder, H., Gerger, G., Brieber, D., and Schwarz, N. (2014). What makes an art expert? Emotion and evaluation in art appreciation. *Cogn. Emot.* 28, 1137–1147. doi: 10.1080/02699931.2013.870132
- Leder, H., Gerger, G., Dressler, S. G., and Schabmann, A. (2012). How art is appreciated. *Psychol. Aesthet. Creat. Arts* 6, 2–10. doi: 10.1037/a0026396
- Leder, H., Mitrovic, A., and Goller, J. (2016). How beauty determines gaze! Facial attractiveness and gaze duration in images of real world scenes. *I-Perception* 7:2041669516664355. doi: 10.1177/2041669516664355
- Locher, P. J., Smith, J. K., and Smith, L. F. (2001). The influence of presentation format and viewer training in the visual arts on the perception of pictorial and aesthetic qualities of paintings. *Perception* 30, 449–465. doi: 10.1068/p3008
- Martiniq, E. (2016). *10 Famous Installation Artists Whose Work You Have to Know. Widewalls*. Available at: <https://www.widewalls.ch/installation-artists/>
- Minissale, G. (2012). Conceptual art: A blind spot for neuroaesthetics? *Leonardo* 45, 43–48. doi: 10.1162/LEON_a_00324
- Mitschke, V., Goller, J., and Leder, H. (2017). Exploring everyday encounters with street art using a multimethod design. *Psychol. Aesthet. Creat. Arts* 11, 276–283. doi: 10.1037/aca0000131
- MOCA (2002). *The Museum of Contemporary Art, Los Angeles. Installation art. Collection: MOCA's First Thirty Years*. Available at: <http://moca.org/pc/viewArtTerm.php?id=18>
- Novitz, D. (2001). Participatory art and appreciative practice. *J. Aesthet. Art Crit.* 59, 153–165. doi: 10.1111/0021-8529.00015
- Osborne, P. (2002). *Conceptual Art: Themes and Movements*. London: Phaidon Press.
- Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.
- Pelowski, M. (2015). Tears and transformation: feeling like crying as an indicator of insightful or ‘aesthetic’ experience in empirical study of art. *Front. Psychol.* 6:1006. doi: 10.3389/fpsyg.2015.01006
- Pelowski, M., and Akiba, F. (2011). A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas Psychol.* 29, 80–97. doi: 10.1016/j.newideapsych.2010.04.001
- Pelowski, M., Akiba, F., and Palacios, V. (2012). Satori, koan and aesthetic experience: exploring the “realization of emptiness” in Buddhist enlightenment via an empirical study of modern art. *Psyke Logos* 33, 236–268.
- Pelowski, M., Forster, M., Tinio, P. P. L., Scholl, M., and Leder, H. (2017a). Beyond the lab: an examination of key factors influencing interaction with ‘real’ and museum-based art. *Psychol. Aesthet. Creat. Arts* 11, 245–264. doi: 10.1037/aca0000141

- Pelowski, M., Markey, P. S., Forster, M., Gerger, G., and Leder, H. (2017b). 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. *Phys. Life Rev.* 21, 80–125. doi: 10.1016/j.plrev.2017.02.003
- Pelowski, M., Markey, P. S., Luring, J. O., and Leder, H. (2016). Visualizing the impact of art: an update and comparison of current psychological models of art experience. *Front. Hum. Neurosci.* 10:160. doi: 10.3389/fnhum.2016.00160
- Reiss, J. H. (2001). *From Margin to Center: The Spaces of Installation Art*. Cambridge, MA: MIT Press.
- Scherer, K., Dan, E., and Flykt, A. (2006). What determines a feeling's position in affective space? A case for appraisal. *Cogn. Emot.* 20, 92–113. doi: 10.1080/0269930500305016
- Scherer, K. R., and Coutinho, E. (2013). "How music creates emotion: a multifactorial process approach," in *Series in Affective Science. The Emotional Power of Music: Multidisciplinary Perspectives on Musical Arousal, Expression, and Social Control*, eds T. Cochrane, B. Fantini, and K. R. Scherer (New York, NY: Oxford University Press), 121–146. doi: 10.1093/acprof:oso/9780199654888.003.0010
- Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M., et al. (2017). Measuring aesthetic emotions: a review of the literature and a new assessment tool. *PLoS One* 12:e0178899. doi: 10.1371/journal.pone.0178899
- Sherman, A., and Morrissey, C. (2017). What is art good for? the socio-epistemic value of art. *Front. Hum. Neurosci.* 11:411. doi: 10.3389/fnhum.2017.00411
- Silvia, P. J. (2005a). Cognitive appraisals and interest in visual art: exploring an appraisal theory of aesthetic emotions. *Empir. Stud. Arts* 23, 119–133. doi: 10.2190/12AV-AH2P-MCEH-
- Silvia, P. J. (2005b). Emotional responses to art: from collation and arousal to cognition and emotion. *Rev. Gen. Psychol.* 9, 342–357. doi: 10.1037/1089-2680.9.4.342
- Silvia, P. J. (2009). Looking past pleasure: anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychol. Aesthet. Creat. Arts* 3, 48–51. doi: 10.1037/a0014632
- Silvia, P. J. (2013). Interested experts, confused novices: art expertise and the knowledge emotions. *Empir. Stud. Arts* 31, 107–115. doi: 10.2190/EM.31.1.f
- Specker, E., Tinio, P. P. L., and Van Elk, M. (2017). Do you see what i see? An investigation of the aesthetic experience in the laboratory and museum. *Psychol. Aesthet. Creat. Arts* 11, 265–275. doi: 10.1037/aca0000107
- Tallis, R. (2008). The limitations of a neurological approach to art. *Lancet* 372, 19–20. doi: 10.1016/S0140-6736(08)60975-7
- Tatler, B. W. (2007). The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *J. Vis.* 7 4.1–4.17. doi: 10.1167/7.14.4
- Tröndle, M., Greenwood, S., Kirchberg, V., and Tschacher, W. (2014a). An integrative and comprehensive methodology for studying aesthetic experience in the field: merging movement tracking, physiology, and psychological data. *Environ. Behav.* 46, 102–135. doi: 10.1177/0013916512453839
- Tröndle, M., Kirchberg, V., and Tschacher, W. (2014b). Is this art? An experimental study on visitors' judgement of contemporary art. *Cult. Sociol.* 8, 310–332. doi: 10.1177/1749975513507243
- Tschacher, W., Greenwood, S., Kirchberg, V., Wintzerith, S., van den Berg, K., and Tröndle, M. (2012). Physiological correlates of aesthetic perception of artworks in a museum. *Psychol. Aesthet. Creat. Arts* 6, 96–103. doi: 10.1037/a0023845
- Tucker, M. (2004). "No title," in *Buddha Mind in Contemporary Art*, eds J. Baas and M. Jacob (Berkeley, CA: University of California Press), 75–86.
- Walker, F., Buckner, B., Anderson, N. C., Schreij, D., and Theeuwes, J. (2017). Looking at paintings in the Vincent Van Gogh Museum: eye movement patterns of children and adults. *PLoS One* 12:e0178912. doi: 10.1371/journal.pone.0178912
- Weingarden, L. S. (2014). "Installation art and experiential aesthetics: mediating art history and neuroscience research," in *Proceedings of the Twenty-third Biennial Congress of the International Association of Empirical Aesthetics*. Available at: <http://www.science-of-aesthetics.org/data/proceedings/IAEACongressProceedings2014.pdf>
- Zentner, M., Grandjean, D., and Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8, 494–521. doi: 10.1037/1528-3542.8.4.494
- Zyman, D., and Codognato, M. (2015). *Olafur Eliasson: Baroque Baroque*. Available at: http://press.tba21.org/News_Detail.aspx?id=45014&menuid=9361

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Pelowski, Leder, Mitschke, Specker, Gerger, Tinio, Vaporova, Bieg and Husslein-Arco. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

2.3.2

Pelowski, M., **Specker, E.**, Gerger, G., Leder, H., & Weingarden, L. (2018). Do you feel like I do? A study of spontaneous and deliberate emotion sharing and understanding between artists and perceivers of Installation Art, *Psychology of Aesthetics, Creativity, and the Arts*.
<http://dx.doi.org/10.1037/aca0000201>

My Contribution:

In this, my contribution lies in the data analysis. In addition, I provided feedback on consecutive versions of the article by my co-authors.

AQ: 1 Do You Feel Like I Do? A Study of Spontaneous and Deliberate Emotion Sharing and Understanding Between Artists and Perceivers of Installation Art

AQ: au Matthew Pelowski, Eva Specker, Gernot Gerger,
and Helmut Leder
University of Vienna

Lauren S. Weingarden
Florida State University

Humans appear naturally inclined to both broadcasting and to perceiving each other's emotional experiences. Especially in the area of empathy or emotion contagion, studies have routinely documented our ability to respond to others' affective states, often via faces or bodies. This can occur on an intellectual level of perceiving emotion signs or can involve actually feeling how others appear to be feeling or mirroring responses with our own brains and physiology. However, while well-documented in interpersonal studies, can this process occur in the absence of another human? Specifically, with art or music, a longstanding argument involves the potential for these media to act as a proxy for the artist, providing an interface whereby one can "feel in" to share an emotional connection. For visual art, however, this process has almost never been empirically recorded. Here, we introduce a paradigm wherein we assessed working artists' emotional experiences and intentions as they produced installation artworks and then assessed resulting emotions and understanding of viewers. We find that in some cases (2 of 3 artworks) viewers did consistently feel more emotions that the artist had intended, they also spontaneously reported feeling similar emotion patterns as those reported by the artists when making, regardless of whether these had been intended. With all 3 artworks, viewers also showed ability to guess the emotion-transmission intentions of artists. Success at emotion sharing and at feeling intended emotions also positively related to ratings (i.e., goodness) of the art.

Keywords: emotion sharing, empathy, visual art, emotion contagion, aesthetics

AQ: 2 *Supplemental materials:* <http://dx.doi.org/10.1037/aca0000201.supp>

To evoke in oneself a feeling one has once experienced, and having evoked it in oneself, then, by means of movements, lines, colors, sounds, or forms . . . so to transmit that feeling that others may experience the same feeling—this is the activity of art (Leo Tolstoy, 1896).

Art—probably since its inception—has been intimately connected with a belief that it should serve as a conduit for a shared experience. As put by the artist Eugène Delacroix (in Goldwater &

Treves, 1958, p. 230), an artwork is a "bridge between the mind of the [artist] and that of the spectator." Art can function as ritual or social performance, providing a space for individuals to meet and to share ethical, political, or aesthetic goals or aspirations (Dis-sanayake, 2008; Kreidler & Kreidler, 1972). Art can record actions, whereby the artist personality or intentions, left like handprint traces in the medium (Benjamin, 1969), can be retraced, and—in conjunction with mimetic signs—provide a vehicle or even a "language" (Gombrich & Saw, 1962) for transmitting meaning, perceptions, or impressions. AQ:3-4 AQ: 5

One of the most fundamental aspects for artist-viewer interactions involves emotion. From at least the days of Plato (Swanger, 1993), art was and is seen as allowing individuals to share "how it feels" or "what it is like" to be somewhere, in some time, or in the shoes of another, and to feel or appreciate their affective experience (Bachelard, 1958; Dewey, 1980; Swanger, 1993). Emotion sharing via art was in fact a major component in the theoretical formation of the 19th century idea of "Einfühlung"—and more recently "empathy" (Gerger, Pelowski, & Leder, 2017)—meaning "feeling into" the emotional state of another through their body or through an artwork (Lipps, 1903; Vischer, 1873).

Such connections are argued to be key to both making and to enjoying art. Numerous researchers suggest that the better one's ability to feel into an object, the deeper, the more sincere, and the

Matthew Pelowski, Eva Specker, Gernot Gerger, and Helmut Leder, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna; Lauren S. Weingarden, Department of Art History, Florida State University.

We thank Hannah Wier for her assistance with the data collection. We would also like to extend a very heartfelt thanks to the three artists who participated in the study, and who graciously made their artworks and their intentions available for our analyses. Without their support and willingness to collaborate at all stages, this project would not have been possible.

Correspondence concerning this article should be addressed to Matthew Pelowski, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, Liebiggasse 5, A-1010 Vienna, Austria. E-mail: matthew.pelowski@univie.ac.at

more pleasurable one's engagement (see Freedberg & Gallese, 2007; Lanzoni, 2009; Leder, Bär, & Topolinski, 2012; Vischer, 1873 for visual art; Eerola, Vuoskoski, & Kautiainen, 2016; Miu & Baltes, 2012; Zickfeld, Schubert, Seibt, & Fiske, 2017 for opera and music). Similarly, for artists, "successful art," notes Swanger (1993, p. 43; see also Gombrich, 1963; Kozbelt, 2006), "creates a connection between the participant's sensibility [and] the sensibility of the artist." The "perceiver, in an empathic sense, becomes one with the artist." Importantly, this might involve both intellectually understanding how another might be feeling, or even sharing, spontaneously or deliberately, an affective experience through the artwork medium. Equally interesting, such connections could be accomplished by both overtly depicted scenes or people, as in a realistic painting, but may work with even minimalist design or abstract images, with the media acting as a "proxy" for the body (Barwell, 1986; Frith & Frith, 2010; Gell, 1998; Sherman & Morrissey, 2017). This empathic engagement is suggested as an "automatic" response which "constitutes a basic level" of processing (Freedberg & Gallese, 2007, p. 202) and is argued to potentially be core to the development of art itself (e.g., see Eerola, Vuoskoski, Peltola, Putkinen, & Schäfer, 2018).

Emotion sharing—especially via visual media *sans human*—is also emerging as a fascinating topic for general brain science and psychology. This might be connected to automatic physical mimicry of emotions as they occur in interpersonal settings, for example by mirroring smiling and frowning expressions (Balconi & Bortolotti, 2012; Singer & Lamm, 2009), and relating to discussion of emotion contagion, social empathy, or mirror neurons (Hoenen, Lübke, & Pause, 2017; Sbriscia-Fioretti, Berchio, Freedberg, Gallese, & Umiltà, 2013; Umiltà, Berchio, Sestito, Freedberg, & Gallese, 2012), as well as constituting a major component of theory of mind or emotion regulation (Balconi & Bortolotti, 2012; Goldin, McRae, Ramel, & Gross, 2008; Singer & Lamm, 2009). Emotional engagement with artistic media—currently, dance, literature, music—is also argued to potentially enhance empathic ability in general (Bal & Veltkamp, 2013; Clarke, DeNora, & Vuoskoski, 2015; Goldstein & Winner, 2012; Harris, 2007) or cultural training and prosocial behavior (Herman, 2007; Menninghaus et al., 2015; Zickfeld et al., 2017). This is also suggested for the shared experience with visual art (Tinio, 2013; van Leeuwen, Verstijnen, & Hekkert, 1999).

However, despite this interest, the actual possibility for emotional sharing through visual art has almost never been empirically investigated (Kozbelt, 2006). Especially, due to the laboratory setting which can tend to minimize "natural" emotional and perhaps empathic responses (Pelowski, Forster, Tinio, Scholl, & Leder, 2017), such study has not involved visual images, which nonetheless formed the basis for much of the discussion of artist-viewer emotion connections itself. Similarly, there are almost no studies that incorporate the actual artist, which could document how these individuals felt or what they wanted to communicate to a viewer. These issues raise some very basic questions whose answers have long been implicitly, if not empirically, accepted: Do viewers routinely perceive specific emotions that were intended by an artist as part of the artwork design, or can viewers intellectually understand the artistic intent for emotion? Can viewers also "feel into" the emotion related to visual art and does this correspond to the artists own felt experience? How do relative sharing or under-

standing of emotion connect to the success or appreciation of art, or to interpersonal (e.g., trait empathy) differences?

In this article we assess the above questions within a unique paradigm employing the collaboration of both viewers and visual artists. By recording the felt emotional experiences and the emotion transmission intentions of three artists as they produced installation artworks with no obvious depiction of human bodies and matching these to a group of viewers as they interacted with the same works within a gallery setting, we considered whether emotions can be deliberately and/or spontaneously transmitted as well as the relation to art appreciation and interpersonal viewer differences. Directly below, we review in more detail the empathy or emotion transmission arguments as well as existing studies, and conclude with implications and questions for future research.

Background: How Might Emotion Be Shared Through Visual Art?

In setting up this study, a review of the literature (e.g., see Barwell, 1986 for a nice review in art criticism) suggests four main possibilities for how emotion could be transmitted to a viewer through art. These also provide a nice set of research hypotheses.

First, it is possible that the same emotions could be experienced by both artist and viewer when engaging with the same work. Also referred to as "emotion mirroring" (Freedberg & Gallese, 2007), "emotional contagion" (Singer & Lamm, 2009), or "emotion communion" in interpersonal psychological studies (Levinson, 1990), this would involve an isomorphic sharing of another person's affective state (de Vignemont & Singer, 2006; Eisenberg & Fabes, 1990; Sherman & Morrissey, 2017) rather than merely picking up emotion signs in someone else's actions. Looking to the topic of art-making, this could come about through deliberate artist-directed actions. Largely following the quotes in the introduction, this phenomenon—what Barwell (1986, p. 176) calls the "romantic view" of art-making—would suggest a case where an artist may have had an experience or a specific set of emotions which they then wished to transmit, reproduce, or evoke in a viewer as well. An artist could accomplish this by employing higher level metaphor or conceptual approaches—for example, evoking specific feelings or similar episodic memories via symbols or content (Juslin & Västfjäll, 2008; Marks, 1983). More subtly, artists might employ techniques such as specific colors, lines, or rhythm, which excite physiology in specific ways so as to evoke affective meaning (Goodman, 1968; Ramachandran & Hirstein, 1999). Such emotion sharing could also be a byproduct of the making process. An artist could choose a certain means of depiction or stop at a certain point by using their emotional response as a primary guide for their actions. For example, an artist might make an aesthetic decision because it 'feels' a certain way or by 'relying on their gut' (Schwarz & Clore, 2003; see Elkins, 1999 for discussion with art). This could theoretically also impact the viewer in the same way when they experience the art, leading to a shared emotional experience.

Second, it is possible that a viewer may have an emotional response, upon encountering an artwork, mirroring the emotions spontaneously felt by an artist when making but perhaps having nothing to do with their intentions. Much as individuals, when happy or sad, may reflect this on their face or in their actions even without overtly trying to do so, resulting in others mirroring or

sharing such feelings, the artwork may similarly serve as a “rubbing” or an “index” (Gell, 1998; see also Barwell, 1986) of the ambient emotions of the artist. This would not imply that the emotions would necessarily match the actual intentions of the artist, but may be an unplanned result of their making a work. Think of well-known works by artists who were undergoing life changes—divorces, marriage, the contemplation of suicide (see, e.g., Berger, 1972 for an example with Van Gogh and a painting of wheat fields in the summer). It is possible something of their emotional state could make its way into the final product where it could similarly modulate the emotions of a viewer.

Both the first and second types suggest a subjectively *felt* (i.e., “hot”) emotional experience (Schaefer et al., 2003), rather than mere recognition of symbolic emotional content. Third, on the other hand, a viewer could also have a more “cold” analytical or declarative response. One might engage a work and, as part of their perception or meaning-making, pick up on the emotional tones communicated by an artwork but without actually feeling them. This might also involve the understanding of assumed intention or the ability of art, as part of its design, to symbolize or refer to certain emotions or affective valence (Barwell, 1986). It is this variety of emotion “perception” that has in fact been a main target for much of the existing emotion-sharing aesthetic research (Zentner, Grandjean, & Scherer, 2008 for review). For example, early studies on appraisals of music asked individuals to describe music not in terms of how they felt, but in emotional terms relating to its construction (Collier, 2007; Wedin, 1972). This is also common in art critical discourse, which often explains visual effects of painting or other visual media in terms of emotion connotations (e.g., Kandinsky, 1947). Note, however, this still implies some empathic communication or connection between artist and viewer, and potential individual differences.¹

Fn1

Fourth, it is also possible that viewers may have an idiosyncratic emotional experience that has little to do with the artist’s intention or feelings. Each viewer may respond to the features of an artwork, connect these to their own expectations or history, and have a response which is unique to the individual (e.g., Barwell, 1986). This of course may still emphasize the ability of art to evoke emotion, which itself may be key in appreciation, but does not imply any covenant between viewer and artist.

Previous Studies and Outstanding Questions

The above possibilities of course raise numerous questions regarding whether and how these can be documented with visual art, as well as regarding the basic ability of viewers to pick up or share emotional content via artistic media at all. These are only beginning to be researched. To date, studies have predominantly involved nonvisual media or basic emotion-recognition topics.

Barbiere, Vidal, and Zellner (2007; see also Trainor & Trehub, 1992 for a similar study) had participants listen to four sound clips which they were then asked to associate with colors and depicted emotions (happiness, sadness, or anger). They showed that indeed participants were able to make consistent associations between music and both aspects, suggesting that a perceiver may be able to pick up, if only intellectually, emotions even with abstract musical content. Similar results, regarding interpersonal consistency in emotion ratings, have been reported with visual objects such as color patches or patterns (e.g., Mayer, DiPaolo, & Salovey, 1990; see

Gage, 1999 for review). Melcher and Bacci (2013) more recently showed ability of lay viewers to agree on the basic valence (happy or sad) of abstract paintings. Individuals showed generally the same consistency as in a similar study with faces. When a first positively or negatively valenced painting was shown for 5 s, it was also shown to modulate the valence appraisals of a second emotionally neutral target image. (see also van Paasschen, Bacci, & Melcher, 2015 for evidence regarding stability of emotion valence ratings even when comparing between novices and art experts).

Miu and Baltes (2012) also offered some evidence for both perceived and felt emotion in conjunction with empathic engagement. They instructed persons to either adopt an emotional focus while listening to opera performances (“imagine vividly how a performer feels in what is described in the music and try to feel those emotions”) or an objective focus asking individuals to consider the musical descriptions and to ignore intuited performer feelings. The empathic condition changed subjectively perceived emotion intensity in perceivers in relation to the specific emotion content. Happy music led to appraisals as more powerful and higher respiration; a sad aria led to higher reported nostalgia and to skin conductance decreases, an indicator for reduced arousal associated with this emotion (Dawson, Schell, & Filion, 2000), suggesting that individuals may be able to actively modulate their experience by trying to share or feel the emotions of an artist.

AQ: 6

Some evidence has also been reported for connections to artists at the level of the brain. Umiltà, Berchio, Sestito, Freedberg, and Gallese (2012) showed participants abstract artworks by Lucio Fontana, comprised of a slashed canvas, as well as computerized black and white line representations of the same slashing patterns, while measuring brain activity with EEG. They reported higher Mu rhythm suppression suggesting activation in cortical motor areas with the original works, which they attribute to viewers’ automatic mirroring of or response to the action traces of an artist. The originals were also rated as more liked (see Leder, Bär, et al., 2012 for similar behavioral findings with painting).

Although showing potential for the present paper’s research questions, these studies also themselves raise issues. Studies may not necessarily translate between domains—especially between static visual media and literature/music (Omigie, 2015). Most studies also have focused only on emotion reception or perceiving emotional content. Nor have studies moved beyond basic emotional valence. Similarly, studies such as that of Umiltà et al. (2012) dealt with a very specific type of response involving motor or action simulation- that may not - involve empathy or emotion itself. Another question involves the interrelation of the varieties of emotion sharing and understanding. While it is agreed that we might personally feel and/or more objectively recognize transmit-

¹ Note that all three emotion transmission/sharing possibilities, when framed within the practice of art-making and implicit arguments for what art *should* do or how it should be responded to, may tend to lead to what Barwell (1986, p. 176) called the “unpalatable consequence” whereby the “expressive properties of the artwork [can become] dependent upon the emotional state and intentions of its creator.” That is “any claim about the sadness, joyfulness, or melancholy of a work . . . can be falsified by the discovery that its creator was not sad, joyful, or melancholy and did not intend to express these emotions.” This is an interesting issue for art theory, but, we would argue, is only of consequence to art psychology in regards to differences in reception between artworks or individuals.

ted emotions, do they often coincide within one experience? Studies have shown that felt and perceived emotions can indeed differ (Zentner, Meylan, & Scherer, 2000 with music); however, again this is not often investigated in visual art, and omits the main question here involving the intentions of an artist.

To our knowledge, only two studies have explored the connection between viewers' reception and the actual emotion transmission goals of the producer of visual art. Takahashi (1995) had university art students produce nonrepresentational line drawings conveying one specific expressive meaning, as assigned by the authors—femininity, human energy, depression, illness, as well as anger, joy, and tranquility. These were ranked by a different group of viewers as to whether they conveyed the intended expressive contents. Although the authors' focus was primarily on the metaphorical response to line or shape qualities and not empathic connections themselves, their results showed agreement among judges (typically from 30% to 50% selecting the “correct” emotion label). More recently, Dubal et al. (2014) asked one artist to produce Chinese characters, depicting the same sign but with either low or high emotional expressivity (although there was no discussion of specific emotions themselves). These were then rated by non-Chinese-speaking participants—making the designs essentially abstract line drawings albeit with an obvious communicative component. The authors used a sensitivity measure (d') to match emotional intensity ratings of the art by viewers against the artists' classifications. They report as well that most of their participants showed a sensitivity above random chance levels. These studies, in conjunction with those above, provide evidence that there may both be a means of communicating intended emotions through visual art. Similarly, this might be routinely received and understood by viewers, with some interpersonal or interstimuli differences—albeit without a clear distinction between felt or perceived emotion, focus on rather simple drawings or characters, and with only one (happy/sad, high/low intensity) emotion intention rather than a complex mix of meanings and responses that might occur in a more natural art making and viewing experience.

Present Study

The present study follows on those by Takahashi (1995) and Dubal et al. (2014), utilizing a similar collaboration with artists, however to a much deeper level of ecological validity and granularity regarding the making and viewing task, the recording of intentions and emotion-transmission, and the distinction between felt and perceived emotional experience. This involved students and researchers within the Florida State University department of Fine Art and Art History and at the University of Vienna Faculty of Psychology, structured around a graduate level class on Installation art practice. Three Master of Fine Arts (MFA) candidates were asked to each produce one installation piece within an appointed gallery space. Following completion, the artists were assessed regarding their felt emotional experience when producing their pieces, using a number of emotion scales and Likert-type ratings. They were also asked to identify the artwork meaning and the specific emotions, using a binary yes/no method, that they desired to transmit through their artworks to a viewer. The emotion and meaning survey was then used for assessing a separate group of viewers' own felt emotions while interacting with the pieces, as well as their own understanding of artist intentions, both of which

could then be matched to and compared against the answers of the artists.

The art type—installation art—was also specifically suited to an emotion transmission and appraisal focus. One of the most important developments in the last 50 years of art production and scholarship (Bishop, 2005; Osborne, 2002), this involves the production of site-specific pieces which envelop a viewer and which are often designed to bring about complex sensory and emotional experiences (Reiss, 2001; Weingarden, 2014, 2015). This is often the aim of the work and perhaps key to its appreciation (Pelowski et al., 2018). At the same time, the installation pieces employed here (much as in the previous studies with abstract paintings or drawings) also had an absence of clearly defined semantic content or human bodies that might directly convey or describe emotions, thus providing a good test case for our main research aim. The chosen artworks also included a range of themes in terms of conceptual or emotional topics and expected positive-negative emotional valence.

We assessed five research questions: (a) Do artists and viewers routinely share a similar emotional experience when making and when assessing a work of art, thus suggesting the possibility that emotion might be evoked similarly or somehow crystalized within an artwork, potentially stemming from an artist's original felt experience? (b) Do viewers feel and report specific emotions that were intended by an artist—whether or not the artist had themselves felt that way personally? That is, can artists in effect “call their shots” or anticipate how their works will actually impact a viewer? (c) We also assessed whether viewers could guess what emotions artists were probably intending to evoke (following on the above-reviewed studies), presumably as a more cognitive or objective exercise in reading intentions, regardless of the viewers' own felt experiences. (d) Beyond these main targets, we considered whether specific interpersonal differences regarding viewer trait empathy, as well as its cognitive and emotional subfactors such as emotional contagion might allow for more felt emotion or better emotion sharing with artists. In social settings, studies have shown that higher trait empathy scores coincide with stronger correspondence to facial emotion displays by others (Dimberg, Andréasson, & Thunberg, 2011; Lamm, Porges, Cacioppo, & Decety, 2008; Sonny-Borgstrom, 2002; Sonny-Borgstrom et al., 2003), as well as to higher intensity of subjective emotional experience (Dimberg et al., 2011; Korb, With, Niedenthal, Kaiser, & Grandjean, 2014; Künecke, Hildebrandt, Recio, Sommer, & Wilhelm, 2014; Levenson & Ruef, 1992; Sato, Fujimura, Kochiyama, & Suzuki, 2013). With music, people scoring higher in trait empathy tend to show higher ability to accurately perceive and understand emotions and intentions (Baltes & Miu, 2014; Wöllner, 2012). The Mayer, DiPaolo, and Salovey (1990) study of color patches above also reported that higher trait empathy resulted in more congruent emotional reactions. Higher trait empathy is also connected to being more “moved” with music or film (Eerola et al., 2016; Zickfeld et al., 2017). Thus, we expected that participants with higher trait empathy scores might feel more emotion, might exhibit more congruent reactions to the emotional tone of artworks (Dimberg et al., 2011; Sonny-Borgstrom, 2002), and be more able to discern the artist intentions regarding evoked response. (e) Finally, we considered if any of the above aspects of emotion sharing coincided with more or less liking or enjoyment of the art. Here too, there are several potential hypotheses. Most basically, emotion

valence in the viewer, irrespective of artist intentions or feelings, could drive responses. In many art theories and investigations it is suggested that emotionally positive experiences tend to lead to more positive and negative to more negative judgments (Cupchik & Laszlo, 1992; Leder, Gerger, Dressler, & Schabmann, 2012; Pelowski, Markey, Luring, & Leder, 2016 for review). A similar argument might also be made here, with an emphasis especially on felt, but perhaps also perceived emotions in the viewer. Alternatively, taking an objective “aesthetic attitude” whereby one perceives but does not necessarily *feel* emotions, can lead to more pleasurable experience. Especially this may be true for emotionally negative content—for example, disturbing or disgusting images (Gerger, Leder, & Kremer, 2014; Menninghaus, 2003; Wagner, Menninghaus, Hanich, & Jacobsen, 2014). This might also allow individuals to better appreciate the formal features of a work leading to higher appreciation; it could on the other hand also lead to a rather shallow engagement (e.g., Sloboda, 1992). Both possibilities could also be further modulated by potential alignment with the artists’ own intentions and felt experiences. Feeling one is making a connection with an artist or perceiving the ability of art and of the artist to share feelings may lead to higher appreciation (see also the art historical arguments in the introduction; but Currie, 2011).

Method

The study again involved two parts: (a) first a process of recording the intentions and emotional experience of artists when making artworks; and (b) the exhibition of the pieces and recording of the corresponding emotions and understanding by a separate group of viewers. The methods for both parts will be discussed in order before moving to the results for the entire study.

Part 1: Art Making

Participants. The artists for Part 1 were three Master of Fine Arts (MFA) students at Florida State University (one male, two females). In the U.S., an MFA is typically considered a terminal degree for the professional academic training of artists. To enter this program, all individuals were required to pass a competitive portfolio review, and thus were considered to be skilled emerging artists.

Stimuli and art-making procedure. The stimuli consisted of three art installations created separately, one by each artist. These were located in three separate gallery spaces within the College of Fine Arts building on the Florida State campus. All gallery areas were cleared of other objects and were devoid of windows. The artists were assigned the gallery space and asked to complete a work within a 1-month period, with 3 weeks for planning and conception and 1 week for creating and installing the pieces. Artists were given complete freedom in the conception and execution of their artworks, with no prompting or explanation for the underlying study purpose.

F1 Artwork 1 (see Figure 1 for all artworks) consisted of hand drawings transferred to vinyl and attached to the walls, floor, and ceiling of a long rectangular hall. The drawings were executed to resemble long strands of curly hair typical of a woman with African American heritage, forming undulating masses or piles, and also seeming to blow in the air along the hallway. These could

be interpreted both as rather abstract, whimsical shapes providing a light happy tone—as suggested by the artist’s own stated meaning (see Procedure below)—“Art is life’s joy. It is what makes us human.” They could also be interpreted more conceptually, addressing life as a Black woman, or pointing out differences, for example, between social mores of beauty or normality in America.

Artwork 2 consisted of a large room with three projectors set side-by-side and each showing a different video onto a long and a short wall. Each video presented an exterior of a closed door with features unique to each location. These projections were arranged by the artist to reproduce the environs of cases where she had been sexually assaulted. In front of the scenes, crumpled clothes that the artist had been wearing at the time were strewn on the floor. Although the installation did not overtly show people or clues to the underlying significance to the artist, it was designed with the intention of reexperiencing emotions felt during those episodes as well as transmitting something of those feelings to a viewer. As put by the artist “The art piece was a way to come to terms” with events that happened in the past. “This piece was about evoking similar emotions [that] I felt during [that] time . . . general feelings of not knowing what you’re walking into and then increasing anxiety.”

Artwork 3 involved another hallway with two ventilation grates at floor-level along one wall. In the middle of the hall, the artist installed a large industrial canister with three orange extension cords emerging from the top. The canister was placed on a grass-covered platform with casters. The extension cords led into the ventilation system and to two TV sets behind the grates. As viewers walked down the hallway, they could see the flickering TVs playing personal home movies, a satirical political comedy,² and the 1992 Democratic National Convention. The art, according to the artist, was meant to both “split and draw off” or “bombard” the viewer with “chaotic discord of opposing sounds and mired imagery” of contemporary social media. It was also expected to “bring people together in an intimate . . . way” making them feel and/or consider “new methods of human socialization.” The placement of the greenery and extension cords was also meant to be “haphazard” and “amplif[y] the utilitarian characteristics . . . firmly rooted in the spirit of the discarded contemporary proletariat.”

Procedure: Recording artist emotions and intentions.

Immediately after completing the artworks, artists were then given a self-report survey to assess their felt emotional experience, as well as their intentions for transmitting specific emotions to a viewer and the general meaning or motivations behind their artworks. The survey consisted of 37 emotion terms assessed via an 8-point scale (for example, “While I was making the art, I experienced [factor]”). An answer of “0” corresponded to “no such feeling;” answers from 1–7 corresponded to some incidence as well as relative magnitude, with 7 = *extremely*. The emotion terms and scale procedure were based on previous art museum study (e.g., Pelowski, 2015; see also Ekman, Friesen, & Ancoli, 1980; Carstensen, Pasupathi, Mayr, & Nesselroade, 2000 for similar use in nonart related contexts). The specific emotions (see Figure 3 for full list) were selected due to their position as main experiential terms linked to the viewing of art (Pelowski, Markey, Forster,

² Aversano, Garefino, and Parker (2004).

Stimuli and Layout

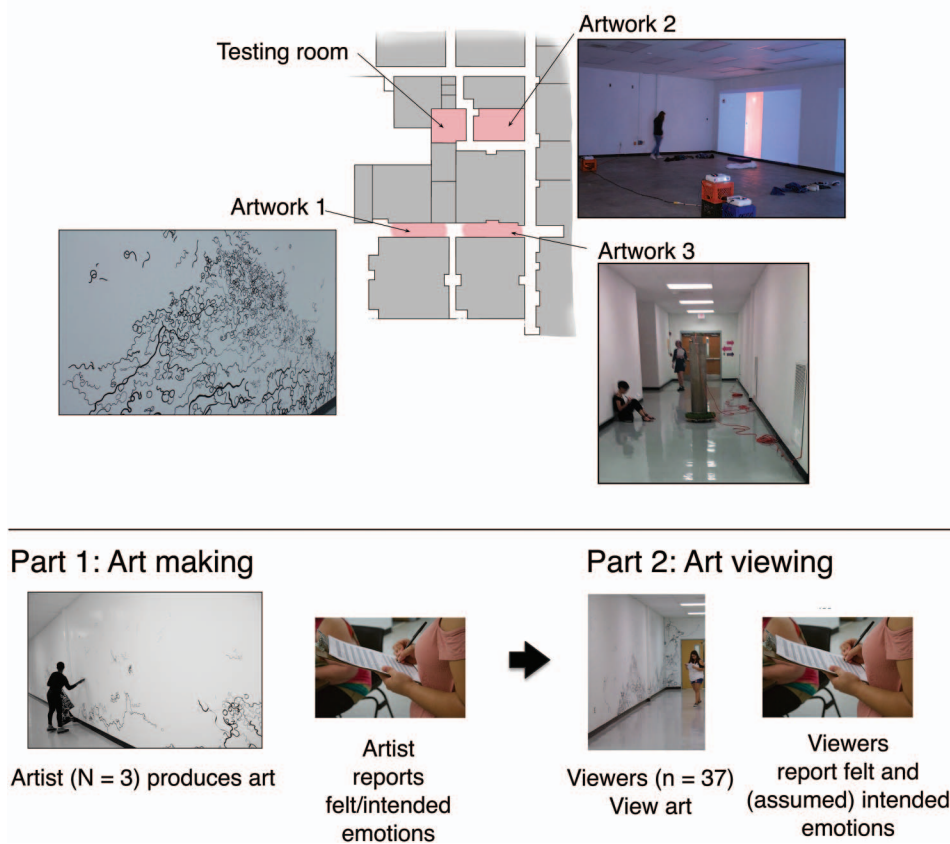


Figure 1. Study design and installation artworks examples. See the online article for the color version of this figure.

ONLINE
COLOR

AQ: 22

Gerger, & Leder, 2017) and covered generally negative, positive, and other notable (i.e., thrilled, chills, moved) reactions (see also Pelowski et al., 2018 for recent use with Installation art). Artists were instructed to think back through the week when they were making and to focus on their general feelings. However, if they had particular notable feelings—even if only fleetingly—they were also encouraged to include those as well. In addition, we also included three 8-point scales for *generally positive* or *generally negative* felt emotion, and high or low “emotional arousal” (following the circumplex model of main affect dimensions; Posner, Russell, & Peterson, 2005). The emotion scales were answered via pen with a paper booklet.

After completing the list of felt emotions, and upon turning the page, the artists were then confronted with a page with instructions to stop and go back, considering the question: “If there are any of the above feelings that you particularly wanted to make the viewer of your artwork feel, please circle the feeling word above.” This was intended to assess intentions for transmitting emotion, as well as felt emotions, while keeping these separate and without prior warning that they would be asked this second question when first reporting their felt experience. Finally, artists wrote a brief answer to the question “What did the artwork mean to you?”

Part 2: Art Viewer Emotional Experience and Assumed Artist Intentions

Participants. The viewers for the produced artworks consisted of a final sample of 37 students at Florida State University (six male, $M_{age} = 22.9$, $SD = 4.7$). All participants were volunteers with no class credit or monetary compensation given. The participants included those who were studying art making/history as well as those studying other majors, thus providing a range of art backgrounds and expertise, although this was not treated as an experimental question.

Procedure and materials. To assess the viewer experiences, the three artworks from Part 1 were displayed during a week-long exhibition. This was open to the public. However, the study participants were scheduled to visit during otherwise closed periods in order to minimize other people in the gallery spaces. Procedure followed a within-subject block design, with all artworks viewed by all viewers. Participants first were met at the entrance to the building and taken to a separate room designated as a testing space (see Figure 1).

Pretest: Art training and interest. Participants were asked to fill out an informed consent form and a brief questionnaire regarding their general demographics, level of education, and previous

art training (following previously published survey, Chatterjee, Widick, Sternschein, Smith, & Bromberger, 2010), objective art involvement (Leder, Gerger, Brieber, & Schwarz, 2014; e.g., “How often do you visit art museums?” on a 7-point scale: 1 = *less than once per year* to 7 = *once a week or more often*), and personal interest in viewing and discussing art (Pelowski, 2015; e.g., “art is important” and “I feel that I should understand a work of art” on a 7-point scale 1 = *strongly disagree* to 7 = *strongly agree*).

Viewing. Participants were then told that they would be asked to enter three galleries and view the artworks inside. They were further asked to “look at the art as you normally would, and for as long as you like.” Participants were then taken individually by a researcher to the first artwork space and asked to enter, view the art, and then return outside. The researcher did not accompany the participant inside and did not observe their interaction.

Art-viewing posttest. After finishing one artwork, participants returned to the testing area and filled out a postviewing survey. This included the same 37 emotion terms using the same scales and terminology as with the artists above, as well as the three general emotional arousal and positive/negative emotion valence questions. Upon completing the list of felt emotions, participants were then asked to stop and revisit the emotion list, again using the same format as described for artists above. In this case however, the viewers were instructed to circle any terms that they “thought the artist was specifically trying to transmit to you via the artworks, regardless of whether you [the viewer] had actually felt that emotion.”

AQ: 7 In addition, viewers were also asked to rate the artworks on 13 bipolar scales (7-point; e.g., 1 = *good*, 4 = *neither good nor bad*, 7 = *bad*). These were based on the Semantic Differential instrument of Osgood and colleagues (1957) and have been shown to provide a good means of assessing a range of appraisals and reactions to art (Pelowski, 2015). Scales were selected to cover main hedonic ratings (good-bad, beautiful-ugly, pleasant-unpleasant, meaningful-meaningless) as well as terms typically connected to stimulus potency and activity (potent-impotent, active-passive, strong-weak, etc. (Berlyne, 1974; Pelowski, 2015). Finally, participants provided answers to the questions: “Did you understand the artist’s intention?” “Did you feel a sense of the artist’s presence?” “Did you think about the way the artist must have felt when making the art?” and “Did you have a sense of what the artist was thinking when making the art?” (7-point scales: 1 = *strongly disagree* to 7 = *strongly agree*). These were intended as a check of empathic awareness and of sensitivity to the cognitive and affective aspects of the artworks. Viewers were also asked to write a short answer to the question “What did the art mean?” The entire procedure was then repeated for the two remaining artworks, with order counterbalanced between participants.

Postsurvey—trait empathy. After viewing all three artworks, participants were given the Questionnaire of Affective and Cognitive empathy (QCAE, Reniers, Corcoran, Drake, Shryane, & Völlm, 2011; see Gerger et al., 2017 for recent use with visual art viewers). This is a 31-item battery (7-point scales: 1 = *strongly disagree* to 7 = *strongly agree*) designed to assess general ability or propensity to both cognitive (knowing what others feel) and affective (feeling what others feel) empathy. It can further be divided into the five subfactors of: (a) “emotion contagion,” regarding the automatic mirroring of the feelings of others; (b)

“perspective taking,” involving intuitively putting oneself in another person’s shoes to see things from his or her perspective; (c) “online simulation,” overtly attempting to put oneself in another person’s position by imagining what that person is feeling; (d) “proximal responsivity,” a responsiveness aspect of empathic behavior or affective response when witnessing the mood of others in a close social context; and (e) “peripheral responsivity,” which refers to the tendency to instead take a more objective/detached approach to viewing emotion—that is, corresponding to an “aesthetic attitude” which may not engender high subjectively felt emotion.

The study was approved by Florida State University and complied with the ethical standards of the Helsinki convention.

Results

The above final sample of 37 viewers, from an original 43, completed all portions of the study, with answers showing no issues (e.g., monotonous answering or systematic missing answers) that would suggest lack of seriousness or engagement. Five individuals were removed due to not viewing all three artworks. Table 1 shows descriptive statistics for the viewer sample, as well as for their trait empathy scores (discussed further below). Once again, most were students at the bachelors or master level, with 48.6% studying an art-related field. The participants, regardless of major, tended to fall on the positive side of feeling themselves comfortable looking at or discussing art and feeling that they usually understood artworks. The participants also tended to like installation art.

T1, AQ: 8

Artwork Evaluations and Meaning Assessments

The evaluations of the three artworks are shown in Figure 2. In general, all three artworks were positively rated with hedonic terms—good, interesting, meaningful. This was especially true for Artworks 1 and 2. As expected, the ratings of relative pleasantness and happy/sad valence showed a range of responses, with Artwork 1 rated as the most pleasant and happy, Artwork 2 rated as more sad and unpleasant (albeit, again, assessed as good and interesting), and Artwork 3 generally at the midpoint between the other two. Artwork 1 was also viewed as the most clear and strong. Artwork 2 was viewed as the most serious, active, and intimate; Artwork 3 was, on the other hand, generally more unclear, distant, and less strong/potent. Repeated measures MANOVA for all terms, comparing between artworks, showed a significant difference, $F(28, 72) = 3.19, p < .001, \eta_p^2 = .55$, with univariate comparisons also showing differences in most of the above scales (see Figure 2).

F2

Answers to the other posttest questions regarding viewer understanding or perceived artist connections (right side, Figure 2) also showed similar differences, with Artworks 1 and 2 showing positive agreement from viewers that they could “feel a sense of the artist’s presence,” could feel “a sense of what the artist was thinking,” “thought about the way the artist must have felt when making the art,” and could “understand the artist intention.” Artwork 3 showed mean scores generally on the disagreement side of the scales for these questions.

The above differences were also echoed in participants’ written answers to “What did the art mean?” (listed in full in Table S1 of

Table 1
Study Sample Basic Descriptive Results, Personality, and Art Attitudes

Question	Mean/(number)	SD	Range
Highest level education			
1 = Below high school	(<i>n</i> = 0)		
2 = High school	(8)		
3 = Vocational school	(1)		
4 = University (Bachelor)	(23)		
5 = Masters/PhD	(3)		
No answer	(2)		
Studied fine art/art history?			
1 = No	(18)		
2 = Yes, Bachelor's degree	(12)		
3 = Yes, Masters	(6)		
No answer	(1)		
I consider myself knowledgeable about art	4.51 ^a	1.84	[1–7]
I am comfortable looking at/discussing art	6.11	1.35	[2–7]
I usually understand art	5.31	1.43	[1–7]
I like installation art	5.89	1.25	[1–7]
Art should make you feel	6.14	.88	[4–7]
Art should make you think	6.00	1.24	[1–7]
Trait empathy cognitive composite	5.28	.73	[3.40–6.68]
Trait empathy affective composite	4.63	.68	[2.75–5.58]
Trait empathy subscale: Perspective taking	5.42	.92	[3.20–7]
Trait empathy subscale: Emotion contagion	5.10	1.24	[2.50–7]
Trait empathy subscale: Online simulation	5.09	.74	[3.56–6.33]
Trait empathy subscale: Peripheral responsivity	3.45	.88	[2.50–6.25]
Trait empathy subscale: Proximal responsivity	5.36	1.18	[1.50–7]

Note. Results based on final sample of 37 students from Florida State University (six male, $M_{age} = 22.9$, $SD = 4.7$).

^a Results based on 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*).

online supplementary materials). Answers were independently scored by two raters for a general match/mismatch to the artists' written statement regarding intended meaning. For Artwork 1, 51.4% noted meanings also given by the artist (e.g., "Commented

on a cultural aspect of blackness, caused me to think about womanhood and the intersectionality of race and gender;" some also mentioned emotional ideas of chaos or being untamed/free). On the other hand, 10.8% gave differing meanings (e.g., having hair

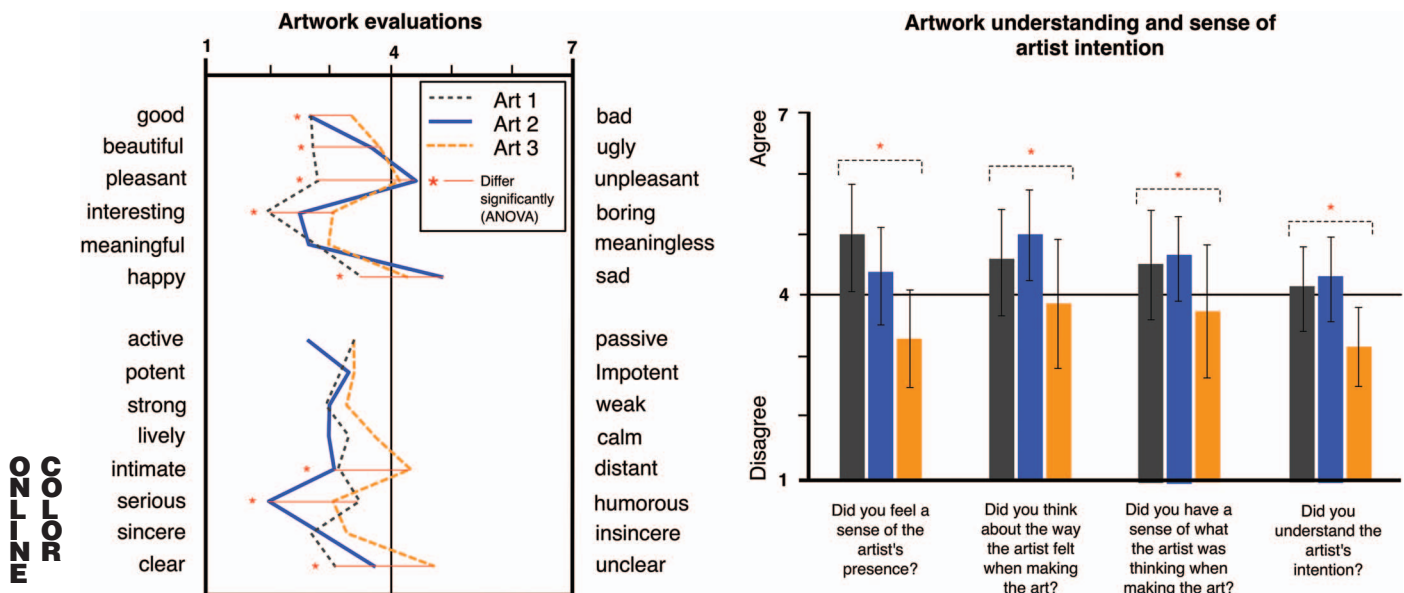


Figure 2. Artwork evaluations and general reported understanding and awareness of artist thoughts and feelings. *Indicates significant difference ($p < .002$) for post hoc pairwise comparisons. See the online article for the color version of this figure.

DO YOU FEEL LIKE I DO?

9

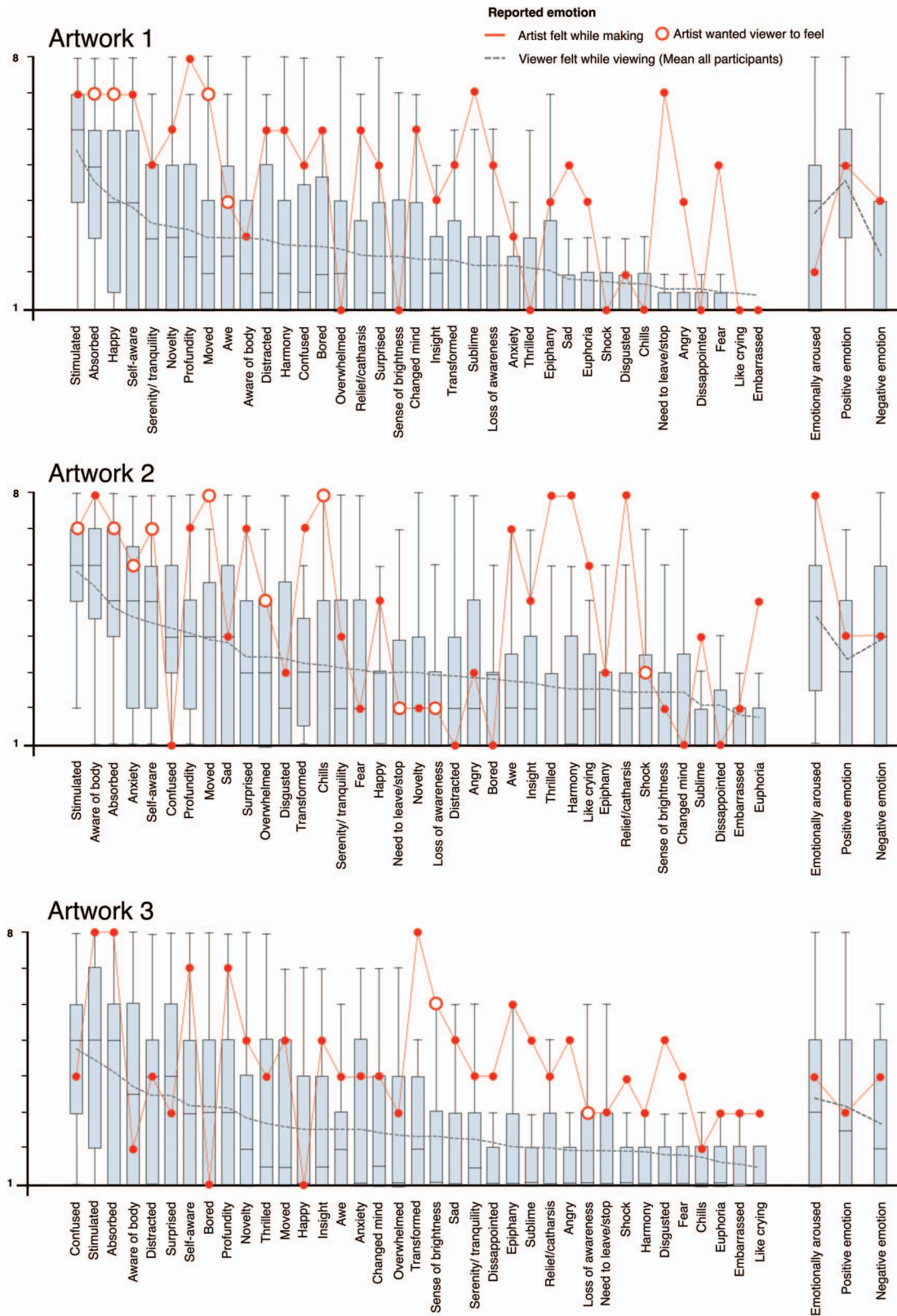


Figure 3. Felt emotions as reported by artists when making and viewers when perceiving three installation artworks (Raw totals). Dotted line represents group Mean, horizontal lines represent Median. See the online article for the color version of this figure.

vs. baldness); 37.8% said they could not understand. For Artwork 2, 35.1% gave matching meanings, picking up on the ideas of sexual assault or more generally walking into dangerous situations and the emotions of anxiety or threat. Whereas, in Artwork 3, only 13.5% noted the artist's ideas of media bombardment or social familiarity. Most viewers (45.9%) instead gave a range of answers regarding topics such as political issues or commentary on the 1990s (due presumably to the TV programs), global warming/ environmental issues (presumably because of the grass); 40.5% stated that they could not understand. Chi-square for meaning Match/Mismatch $\times$ Artwork, $\chi^2(2, N = 111) = 10.61, p = .005$.

Reported Felt Emotions in Artist-Making and Participant-Viewing Experience

Figure 3 displays the mean scores and boxplots for reported felt emotions of all viewers, ordered from those emotions with the highest reported magnitude to the lowest. This also displays the reported felt emotions from the artists when making the art (red circles). In general, Artwork 2 tended to generate the highest emotional arousal for both the artist and the viewers as well as relatively higher negative affect. Looking at the individual emotions, the highest reported terms for all three artworks included a sense of being stimulated, absorbed, surprise, and self-awareness. Artwork 1 also showed generally higher reported happiness, serenity, novelty, profundity, and sense of being moved. Artwork 2 as well showed a generally high sense of profundity and being moved, while also showing relatively higher sadness, anxiety, confusion, disgust, and sense of being overwhelmed. Artwork 3 had, as its highest rated emotion, a sense of confusion, as well as being distracted, aware of one's body, bored, but also having a sense of novelty and thrill. Similar results can also be seen in the general positive, negative affect and sense of arousal questions (Figure 3, far right). Note also that all three artists tended to indicate generally higher magnitude of emotions than almost all of the viewers (as can be seen from comparison with the boxplots). It is also notable that emotions designated by the artists as those they wanted to transmit to viewers (red open circles, Figure 3) were also generally rated as among the most felt by the artist personally (a "7" or "8" on the 8-point scale).

Did Viewers Feel the Same When Viewing as the Artist Felt When Making Art?

F4 In order to assess whether the viewers' reported emotions generally matched those reported by artists when they were themselves making the art, again regardless of intentions, we assessed correlations between the artists' scores and the mean scores for all viewers, across all emotions (Pearson product-moment, two-tailed, see also Figure 4). This resulted in a significant positive correlation for Artwork 1, $r = .574, p < .001$; and Artwork 2, $r = .378, p = .02$; and smaller, nonsignificant correlation for Artwork 3, $r = .300, p = .07$. We also conducted similar correlations independently between the artist scores and each viewer (Figure 4 displays this as histograms). For Artwork 1, this revealed a range of $r = .606$ to $.017$, with a mean of $r = .299$ ($SD = .165$) and 17 individuals (45.9%) showing a correlation score above an alpha threshold of .05; for Artwork 2, a range of $r = .565$ to $-.243$ ($M_r = .176, SD = .206; n = 9, 24.3\%$, above significance

threshold); and for Artwork 3, a range of $r = .454$ to $-.318$ ($M_r = .101, SD = .215; n = 5, 13.5\%$, significant).

Did Viewers Experience Specific Emotions Intended by the Artist?

We then moved to the question of whether viewers experienced emotions that were intended by the artist. Those emotions designated by the artists as targets to transmit to a viewer (Figure 3, open red circles) were as follows: Artwork 1's artist suggested that they wanted others to feel happy, absorbed, moved, and awe. Artwork 2's artist noted stimulated, absorbed, moved, as well as anxiety, self-awareness, sense of being overwhelmed, chills, need to leave, loss of awareness, and shock. Artwork 3's artist noted a desired loss of awareness and a sense of brightness.

To test this, we compared the magnitude of viewer-reported felt emotions that the artist had also designated (combined mean) against the mean magnitude of emotions that the artist did not designate. As shown in Figure 5, for both Artwork 1 and 2, viewers F5 reported feeling the emotions that were intended by the artist at a higher magnitude than the emotions not intended. Whereas, for Artwork 3, we found an opposite pattern. A two-way ANOVA with *artwork* and emotion *intention* (intended/not intended) as within-participant factors revealed a main effect of artwork, $F(2, 66) = 19.9, p < .001, \eta_p^2 = .38$; and of intention, $F(1, 33) = 24.5, p < .001, \eta_p^2 = .42$, as well as an interaction between the two, $F(2, 66) = 33.3, p < .001, \eta_p^2 = .50$. At the individual participant level as well, Artwork 1 had 75.7% ($n = 28$) of participants reporting a positive difference favoring artist intended emotions; Artwork 2 had 86.5% of participants ($n = 32$); Artwork 3 had only 16.2% ($n = 6$).

Were Viewers Able to Guess the Intended Emotions for an Artwork?

We then considered whether viewers were able to identify the emotion-transmission intentions of the artist. To assess this question, we treated each individual emotion as a separate binary "yes/no" question and divided the viewers' answers among "hits" (cases in which the viewer had circled, on the survey, an emotion that the artist had also indicated), "false alarms" (cases where the viewer had indicated that an emotion was intended but which had not actually been specified by the artist), "misses" (cases where the viewer had not circled an emotion which the artist had circled), and "correct rejections" (both artist and viewer had not circled an emotion). The breakdown of answers for each artwork is shown in Table 2.

T2 In Artwork 1, 24.3% ($n = 9$) correctly identified at least one emotion, four individuals identified at least 50% (of the four emotions circled by the artist). The total correct guess rate, combining yes and no answers, was 89.4%, and the success rate when making a positive "yes" guess was 28.9% (vs. 10.8% chance given four of 37 possible artist-identified emotions). With Artwork 2, 64.9% of viewers ($n = 24$) had at least one correct hit, two viewers correctly identifying over half of the (10) artist-selected emotions, the total rate of correct (yes + no) answers was 73.5%, and the success rate when making a "yes" guess was 62.9% (vs. 27% chance). With Artwork 3, the correct yes/no guess rate was 89.6%, however no participants (0%) correctly made any "yes" identifi-

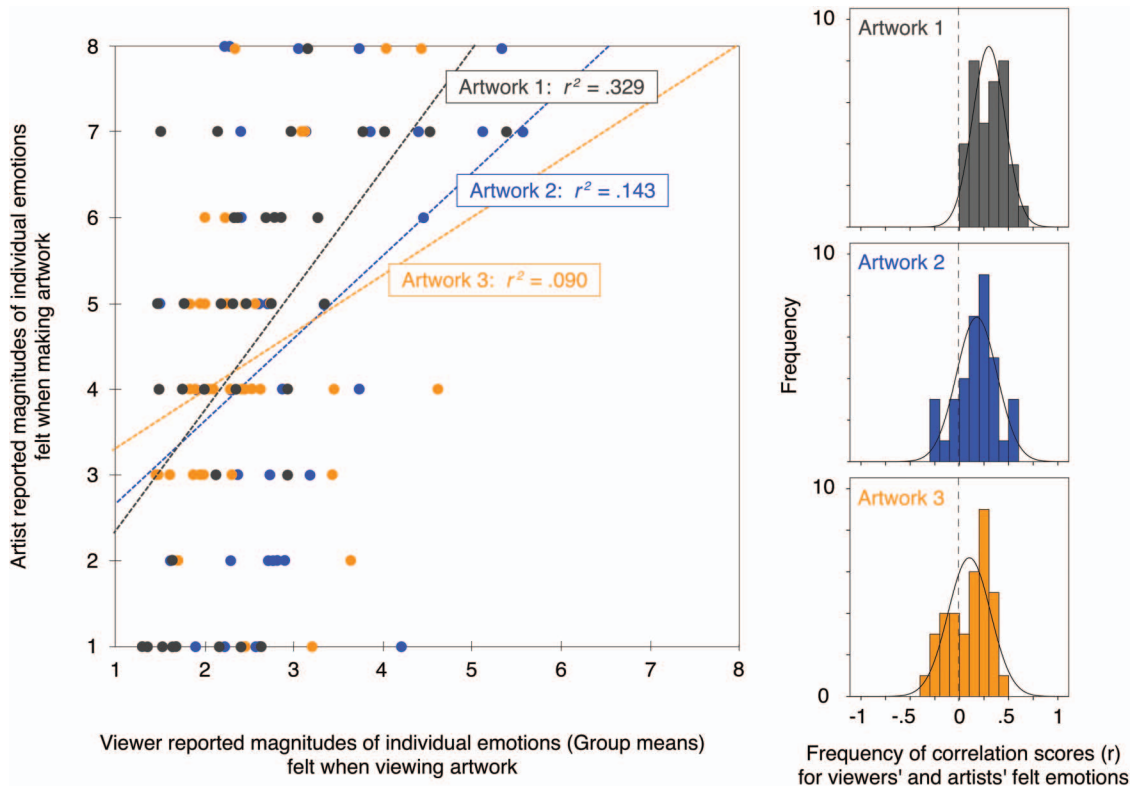


Figure 4. Right: correlation of magnitudes of individual reported emotions as felt by the artist when making and the viewer when viewing artworks (each dot represents an individual emotion term from the 37-item self-report survey; viewer scores are based on the group mean). Left: Histograms of individual viewer's correlation scores for each artwork. See the online article for the color version of this figure.

cations of emotions. A chi-square test, conducted separately for each artwork over the Artist/Viewer $\times$ Yes/No Combinations was significant for Artwork 1, $\chi^2(1, N = 1,369) = 14.55, p < .001$; and for Artwork 2, $\chi^2(1, N = 1,369) = 36.78, p < .001$, but not for Artwork 3, $\chi^2(1, N = 1,332) = 2.93, p = .09$, Yates corrected for cell with zero items.

To statistically assess the above answering patterns, we calculated a sensitivity index (d') independently for each room and for each viewer. This measure (e.g., following Dubal et al., 2014 discussed in the introduction) accounts for participant differences in propensity to use “yes/no” answers as well as for differences in the distribution of possible correct artist-identified emotions between artworks. Cases where the ratio of false alarms or correct hits was “0” such as with all viewers with Artwork 3, were adjusted following Hautus (1995; see also Stanislaw & Todorov, 1999). The group means are reported for each artwork in the far-right columns of Table 2. All artworks had a generally positive sensitivity among the viewers, suggesting some ability at signal detection. Note that a d' of “0” would signify no sensitivity or essentially random answering with yes and no; a negative score would suggest a detection of differences between emotion groups but answering in the wrong direction. Artwork 1 had a $M_{d'} = .78$ ($SD = .40$) with all but two (35) individuals answering >0 ; Artwork 2: $M_{d'} = .62$, $SD = .47$, $n = 31 > 0$; Artwork 3: $M_{d'} = .63$, $SD = .55m$, $n = 29 > 0$). One-sample t tests against “0” using the group means were significant for all

three artworks (Artwork 1: $t(36) = 11.73, p < .001$; Artwork 2: $t(35) = 7.87, p < .001$; Artwork 3: $t(36) = 7.04, p < .001$, Bonferroni corrected). A repeated measures ANOVA comparing sensitivity scores within participants between each artwork, on the other hand, was not significant, $F(2, 70) = 1.19, p = .31$.

Note also that the decision criterion statistic (C , see Table 2), which measures an individual's likelihood of using “no” versus “yes” responses, suggested that participants tended to take a rather strict approach to providing “yes” answers, albeit, importantly, with equivalence in approach between artworks (repeated measures ANOVA ns).

Emotion Sharing/Understanding Consistency Across Artworks; Within-Viewer Correlation?

Because the above artwork cases were treated separately, we were also interested in assessing whether individual viewers showed similar responses across the artwork/artist interactions. Cronbach's alphas, considering agreement for each artwork per viewer, showed relatively high internal consistency for general magnitude of emotion arousal ($\alpha = .731$), as well as for correlation scores measuring artist-viewer felt emotion sharing ($\alpha = .626$). On the other hand, lower consistency was found for the difference scores between felt emotions that had and had not been intended by the artist ($\alpha = .488$) and for the sensitivity index when guessing artists' emotion intentions ($\alpha = .319$). We also assessed whether

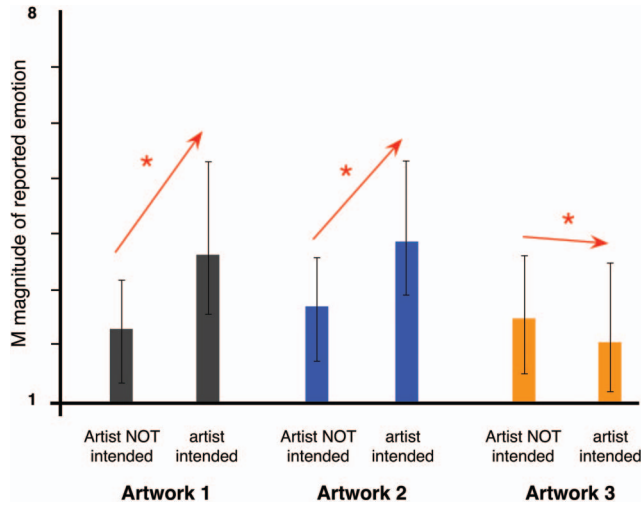


Figure 5. Difference between viewers' reported felt emotions as intended and not intended by artist. (Two-way ANOVA with artwork and intended/not intended emotions as within-participant factors revealed a main effect of artwork and of intention as well as an interaction between the two factors. Post hoc pairwise analysis of intended versus nonintended emotion means also significant within all three artworks, all $ps < .009$). See the online article for the color version of this figure.

the above factors themselves correlated within participants. Correlation coefficients (Pearson product-moment, two-tailed, calculated on mean scores from all three artworks) are shown in Table 3. Overall, results did not suggest a strong relationship between the separate factors. The only significant correlation between the main emotion/intention responses involved understanding artists' intentions in general and the correlation scores for sharing the same pattern of emotions with the artists.

Trait Empathy and Felt Emotion or Emotion Sharing and Understanding

Moving to our subquestions, the descriptive results of the trait empathy questionnaire are shown in Table 1. Overall, this showed a good distribution ranging from the high-end to just below the midpoint of the various scales. First, we considered whether trait empathy could meaningfully predict differences in general magnitude of emotional response. This was assessed via linear regression with the scores from the trait empathy questionnaire as

predictors and viewers' answers to the general question of relative emotional arousal felt while viewing the art, averaged across the three artworks, as the dependent variable. In a first step, using the two "cognitive" and "affective" empathy components, we found very little evidence for a reliable relation, with the model only explaining 2.3% of variance, $F(2, 28) = 0.33, p = .72$. An analysis conducted with the five empathy subcomponents as predictors (online simulation, perspective taking, peripheral responsivity, proximal responsivity, emotion contagion) returned a marginally better prediction. These results are shown in Table 4. The model explained 23% of variance, with the cognitive empathy components "online simulation" showing a significant predictive ability.

A similar regression utilizing the correlation scores between viewers' and artists' self-reported emotional experiences as the dependent variable also showed that the two general cognitive and affective empathy components had very low predictive power. Therefore, we again focused on the five subcomponents (see Table 4). Unlike the regression above, however, we found little evidence for a predictive relationship. "Online simulation" was again generally the strongest predictor—albeit with quite small effects. Regressions conducted for both differences in reported magnitude of felt emotions intended or not intended by the artists as well as with sensitivity (d') when discerning artist intended emotions also did not reveal statistically reliable results and with no subcomponents approaching significance.

Emotion Sharing/Understanding and Positively Rating the Art?

Finally, we assessed whether the above emotion arousal, emotion understanding, and/or emotion sharing factors coincided with positively rating the art. To assess this, we conducted multiple regressions, with answers to the good–bad rating scale as the dependent variable and the emotion sharing and general artwork understanding factors discussed above as predictors. These were again conducted separately for each artwork due to the above differences in ratings and emotion/understanding responses. To consider best model fit these were assessed with a stepwise approach. Results are shown in Table 5 (note, this displays both B and standardized β due to the differences in underlying scales; all predictors showed variance inflation factor scores of <1.5 suggesting low multicollinearity). The results suggested slightly different patterns, however, with sharing the felt emotions of the artist a significant predictor for both Artworks 2 and 3. Artwork 1 showed significance for both feeling more those emotions intended

Table 2

Distribution of Viewer Guesses Regarding Artist-Intended and Nonintended Evoked-Emotions and Signal Detection Statistics

Artwork	Number of artist identified emotions	Hits (artist yes, viewer yes)		Correct rejections (artist no, viewer no)		False alarms (artist no, viewer yes)		Misses (artist yes, viewer no)		Correct guess rate (yes + no)	Mean d' (SD)	Mean C (SD)
		Total	% of possible	Total	% of possible	Total	% of possible	Total	% of possible			
Artwork 1	4	14	9.5%	1,183	96.9%	38	3.1%	134	90.5%	89.4%	.78 (.40)	1.48 (.40)
Artwork 2	10	51	13.8%	955	95.6%	44	4.4%	319	86.2%	73.4%	.62 (.47)	1.40 (.43)
Artwork 3	2	0	.0%	1,194	94.8%	66	5.2%	72	100%	89.6%	.63 (.55)	1.35 (.35)

Note. Total and percentage results for hits, correct rejections, false alarms, and misses based on aggregate score of 37 participant art viewers. Mean d' and C statistics based on averaged scores across all participants for each artwork.

DO YOU FEEL LIKE I DO?

13

Table 3

Correlations Between Main Artwork Understanding, Emotion Guessing, and Felt Emotion Sharing Measures

Measures	General emotion arousal	Felt emotion intended versus not intended by artist (dif. score)	Shared artist emotion (viewer/artist correlation)	Guessing artist emotion intention (d')	Understood artist intention
General emotion arousal	—				
Felt emotion intended versus not intended by artist (dif. score)	.395*	—			
Shared felt emotion (viewer/artist correlation)	.566**	.299	—		
Guessing artist emotion intention (Y/N sensitivity index d')	-.200	-.036	-.047	—	
Understood artist intention	.499**	.308	.462**	-.271	—

Note. All measures based on mean scores from three artworks. Correlations coefficients refer to Pearson product moment, two-tailed.

* Significant at $p < .05$. ** Significant at $p < .01$, uncorrected for multiple comparisons.

by the artist and general emotional arousal. Artwork 3, in addition to sharing felt emotions with the artist, also included generally understanding the artist intentions, potentially due to the comparatively higher lack of understanding and reported confusion (as noted

above). Note also that correctly guessing emotion-transmission intentions, as opposed to feeling emotions whether shared or intended by the artist, did not significantly predict liking in any case.

Discussion and Conclusion

This study considered the topic of emotion sharing and understanding through works of visual art, utilizing a unique program involving both viewers and artists and recording their (explained or assumed) emotion-transmission intentions, artwork understanding, and actual felt emotional experiences. We employed an artwork type, installation art, which was argued to be specifically suited to this topic due to its focus on creating sensory or affective experiences without using clearly defined human bodies or mimetic emotion-referents, and thus offering a fascinating look not only into our art engagement but into human empathy or emotion transmission and contagion itself.

Looking to the results, and stressing again their exploratory nature, we do find support for all three main research questions—*Can viewers feel and report specific emotions that were intended by an artist? Can viewers objectively guess what emotions artists were intending? Can artists and viewers share emotions when looking at or making art, even if these may not have been intended?*—albeit coinciding with artwork/artist-based differences. With two of our three artworks (1 and 2), viewers did report significantly higher magnitudes of felt emotions that were intended by the artist, compared with other nonintended emotions. With all three artwork cases, viewers also showed a significant ability to guess what emotions an artist was intending or, more often, not intending to transmit above chance levels (as assessed by sensitivity indexes, d'). With Artworks 1 and 2, we also found significant positive correlations between the patterns of felt emotions reported by viewers and by the artist when originally making the art. These results support the argument that emotion sharing and guessing may be present in at least some art interactions, with our population of viewers able to show a pattern of responses above a threshold of chance or statistical significance. This occurred (again especially with Artworks 1 and 2) across artwork trials that themselves differed greatly in theme, intended emotions, and basic positive/negative valence, and in the absence of any priming about the nature of the tasks with both artists and viewers.

Regarding the feelings and guessing of emotions as intended by an artist, results are also in line with past research that has shown that individuals may often generally agree on the emotional va-

Table 4

Regression Analyses of Affective and Cognitive Empathy Components as Predictors of General Emotion Arousal, Felt Emotion Correlations, or Intended Emotion Feeling and Guessing in Three Installation Artworks

Predictor	B	t	p
General emotional arousal $R^2 = .227$, $F(5, 25) = 1.47$, $p = .24$			
Intercept	5.638	1.851	.07
Perspective taking (cognitive)	-.071	-1.718	.10
Online simulation (cognitive)	.121	2.229	.04*
Emotion contagion (affective)	-.009	-.125	.90
Peripheral responsivity (affective)	-.105	-1.142	.26
Proximal responsivity (affective)	-.075	-.825	.42
Artist-viewer felt emotions correlation $R^2 = .088$, $F(5, 25) = .49$, $p = .78$			
Intercept	.306	1.266	.22
Perspective taking (Cognitive)	-.003	-1.001	.33
Online simulation (Cognitive)	.004	1.041	.31
Emotion contagion (affective)	-.004	-.141	.50
Peripheral responsivity (affective)	.000	.008	.99
Proximal responsivity (affective)	-.002	-.345	.73
Viewer felt emotions, artist intended versus not intended $R^2 = .034$, $F(5, 25) = .17$, $p = .97$			
intercept	1.470	.950	.35
Perspective taking (cognitive)	-.003	-.131	.90
Online simulation (cognitive)	-.007	-.238	.81
Emotion contagion (affective)	.018	.477	.64
Peripheral responsivity (affective)	-.035	-.748	.46
Proximal responsivity (affective)	-.010	-.205	.84
Artist intended emotions guessing (d') $R^2 = .157$, $F(5, 25) = .75$, $p = .60$			
Intercept	1.880	1.746	.09
Perspective taking (cognitive)	.001	.206	.84
Online simulation (cognitive)	-.010	-1.070	.30
Emotion contagion (affective)	.005	.437	.67
Peripheral responsivity (affective)	.017	1.102	.28
Proximal responsivity (affective)	-.007	-.458	.65

Note. All measures based on mean scores from three artworks. Empathy components based on the QCAE, Reniers, Corcoran, Drake, Shryane, and Völlm (2011).

* Significant at $p < .05$.

Table 5
Regression Analyses of Felt, Shared, or Guessed Emotion Factors and Artwork Understanding as Predictors of “Good–Bad” Appraisals of Three Installation Artworks

Predictor	<i>B</i>	β	<i>t</i>	<i>p</i>
Artwork 1: Model: Stepwise, 2 iterations: $R^2 = .356$, $F(2, 33) = 8.56$, $p = .001$				
Intercept	4.132		9.684	<.001***
Felt artist intended versus not intended emotion difference	.411	.363	2.329	.027*
General emotion arousal	.248	.356	2.287	.029*
Excluded predictors				
Understood artists' intentions		-.215 ^a	.226	.682
Shared felt artist-viewer emotion correlation		-.554	.109	.877
Guessed artist emotion intention (d')		-.069	.326	.886
Artwork 2: Model: Stepwise, 1 iteration: $R^2 = .186$, $F(1, 33) = 7.52$, $p = .01$				
Intercept	3.248		11.417	<.001***
Shared felt artist-viewer emotion correlation	2.883	.431	2.743	.010*
Excluded predictors				
Understood artists' intentions		.226	1.260	.217
Felt artist intended versus not intended emotion difference		.102	.580	.566
General emotion arousal		.070	.394	.697
Guessed artist emotion intention (d')		-.007	.046	.964
Artwork 3: Model: Stepwise, 2 iterations: $R^2 = .480$, $F(2, 33) = 14.781$, $p = .001$				
Intercept	4.381		14.720	<.001***
Shared felt artist-viewer emotion correlation	3.088	.497	3.704	.001**
Understood artists' intentions	.213	.353	2.631	.013*
Excluded predictors				
Guessed artist emotion intention (d')		.162	1.158	.256
General emotion arousal		.114	.600	.553
Felt artist intended versus not intended emotion difference		-.019	.143	.887

Note. Results rotated so that positive signs suggest higher relative “good” ratings.

^a Refers to Beta score if included in model.

* Significant at $p < .05$. ** Significant at $p < .01$. *** Significant at $p < .001$, uncorrected.

lence or the specific emotional flavor of music, patterns, colors, or even abstract paintings, presumably which might also be employed by an artist. The signal detection findings are also similar to the previous studies of Takahashi (1995) with abstract paintings and especially of Dubal et al. (2014) who used a similar methodology with line drawings. Here, we expand on these findings, providing positive indication that individuals can detect specifically intended emotions from a much larger list of possibilities and thus perhaps aligning to a relatively more “natural” art making and corresponding art-viewing circumstance.

Perhaps even more intriguing is the suggestion of some spontaneous emotion sharing or contagion, with viewers reporting the same patterns of felt emotions when viewing as the artists reported when they were producing the art, *regardless of whether or not this was intended for transmission or as part of the artwork design*. To our knowledge, this is the first study that has detected such spontaneous connections, perhaps as a result of a shared embodied experience or similar response to the artworks or their elements. This supports the more general argument that mirroring the emotional or physiological state or actions of another may be a basic aspect to human engagement with the environment and to the interaction with or perception of art (Freedberg & Gallese, 2007; Umiltà et al., 2012), further possible even without human faces or bodies but only action traces. Interestingly, this aspect also showed the highest internal consistency between artworks for individual viewers, suggesting that those who could share artist emotions via one artwork could do so for all three, in each case with a different artist. Whereas feeling or guessing the intended emotions showed

lower much consistency. This would mark a fascinating topic for future psychological, aesthetic, or even neuroscientific approaches.

These findings also raise other implications for several aspects of interaction with art. For art making or artistic communication and art enjoyment this study provides evidence that artists can sometimes design art to evoke certain responses or may in effect “call their shots” or anticipate how their artworks will actually impact a viewer. They may also feel, while making, the same as the eventual viewer. However, our finding of consistent differences between Artworks 1 and 2 versus 3 also suggests that these connections may have differing levels of relative success. Importantly, the lower emotion sharing and feeling of intended emotions scores with Artwork 3 also coincided with a lower reported felt “presence of the artist,” lower reported awareness of what the artist was thinking and feeling, as well as lower reported understanding of the general intentions of the artist, lower ability to explain the artworks’ meaning in a manner coinciding with the actual explanation of the artists, and lower liking ratings for the art. This would support the argument that such emotion understanding and connections are important for enjoying art and for making “successful” artworks (e.g., Freedberg & Gallese, 2007; Gombrich, 1963; Kozbelt, 2006). The benefit of the within-participant design also suggests that such differences between artworks cannot be attributed to individual personality or taste differences.

This also obviously raises a question of what viewer or artwork factors or design decisions might be driving the artwork differences. Although we will not delve into this question, this could involve any number of issues—lower understanding of meaning or

intentions reported above, potential misreading of artwork elements, as well as differences involving the artists' own selection of the emotions they desired to evoke. It could also be that the particular artwork did not resonate with our selected sample of viewers, whereas a different viewer group with different training or background may find a different engagement. This is a fascinating topic for future study that might use the present paradigm as a basis, shifting the question from does art satisfy an emotional contagion function—here tentatively answered 'yes'—to what circumstances or what kinds of art succeed in this regard.

Interestingly, in response to our subresearch question of which specific factors may be most important, we might also provide one suggestion. Although again when looking between artworks all main factors appeared to coincide with more positive appraisals, when weighing the various factors together for specific artworks, spontaneous emotion sharing was a significant predictor of liking with two of the three (Artworks 2 and 3); feeling more those emotions intended by the artist and general emotional arousal were also predictors for Artwork 1. No artworks had guessing the intended emotions as a significant predictor. This may support the longstanding argument (e.g., Freedberg & Gallese, 2007; Swanger, 1993) that "feeling into" the emotions of art and the artist, in a felt as opposed to simply intuited fashion, is key. Artists may also need to create the grounds for such connections.

It should also be noted that the individual assessments or reactions—that is, guessing emotion transmission intentions, feeling intended emotions, sharing emotions with the artist—tended to not show correlations within individual viewers. Although all tended to be higher for certain artworks and thus may generally coincide together, this also leaves open the possibility for future study to consider what factors of artwork, context, or viewer may drive one or the other response.

Finally, we found some evidence that art emotion sharing may tie to interpersonal viewer differences. Especially with felt emotion and emotion sharing, viewers tended to show high agreement across all three artworks. This suggests that such connections and understanding may be either an innate or potentially learned trait for art engagement. Trait empathy also coincided with higher emotion arousal and, to a much lesser, nonsignificant, extent with sharing the felt emotions of the artist. This involved the cognitive factor of "online simulation," overtly attempting to put oneself in another person's position by imagining what that person is feeling. This finding follows similar results with music which also tie trait empathy to basic amount of emotion aroused (Dawson et al., 2000). The latter finding is also in line with previous studies that have shown higher trait empathy to predict better emotion contagion and discernment with music and with other people (Baltes & Miu, 2014; Egermann & McAdams, 2013; Wöllner, 2012). However, no connections were again found for emotion guessing. Overall, the tie to trait empathy was also lower than was expected. Again, future study should follow up on this or consider other personality or demographic items that might predict the above varieties of engagement.

AQ: 9

This study also of course comes with caveats and much need for future research. Although representing a rare opportunity to work directly with artists, the study involved only three artworks of one art variety, with results that were very exploratory. This should be treated primarily as a proof of concept, and researchers should be careful when making inferences and should begin to address each

of the research questions in a confirmatory set-up as well as extend and replicate our findings with more art examples. Especially, this might involve more trials with a range of emotional valence. The use of installation art, although chosen especially due to its emotion- or sensory-experience-centered focus and lack of overt human bodies or other emotional imagery, may also provide results that could differ from other representational or mimetic artforms. Future study might test the findings with painting or also compare representational with abstract art. It is also possible that our artists, although having passed a competitive entrance examination, may not show the same emotion-transferring ability as more established professionals. One might, of course, also argue just the opposite! As artists become more and more professional, or make works with a specific 'brand', they may in fact produce art that is more esoteric or less mindful of a general audience. Future studies might consider such artist-differences.

It should also be stressed that, even though we do find evidence for emotion sharing and guessing above levels that meet thresholds for significance, results are also rather modest. To take one example, emotion sharing was connected to correlation coefficients in the midrange of those typically considered to be robust in the social sciences ($r = .35-.55$), and certainly cannot be argued to mean that viewers perfectly matched or felt exactly the same emotions as an artist. There is also the question of what this overlap may mean in practical terms of emotion experiencing or whether there are between-individual differences. In the case of signal detection, although combined successful emotion selections with successful rejections showed that the participant groups could answer significantly above chance levels, when looking to actual selection of emotions specifically chosen by the artist, again success in the best cases was only about 25%–50%; Artwork 3 showed that no participants could successfully guess an artist selected emotion (although they were again quite good at choosing what an artist did not intend). Here as well, the results do not obviously suggest that viewers could perfectly tune themselves to the artist. However, we would also stress that in Artwork 1 and 2, even the modest hit rate was still about double what would be expected by chance. That said, this study does provide evidence that emotion sharing can be a component of viewing art in an ecologically valid setting, involving both feeling similarly to the artist and understanding—as well as feeling—the result of their aesthetic decisions. When this connection is established—as presumed from the statistical analyses—it also reveals a rather clear pattern of better artwork understanding and enjoyment. This raises compelling implications for future studies employing a potentially wide range of human-derived media where transmitted emotion might be a key aspect of the experience.

References

- Aversano, S., & Garefino, A. (Producers), & Parker, T. (Director). (2004). Team America World Police [Motion Picture]. United States: Paramount Pictures.
- Bachelard, G. (1958). *La poétique de l'espace* [The poetics of space]. Paris: Presses Universitaires de France.
- Bal, P. M., & Veltkamp, M. (2013). How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation. *PLoS ONE*, 8, e55341. <http://dx.doi.org/10.1371/journal.pone.0055341>

- Balconi, M., & Bortolotti, A. (2012). Detection of the facial expression of emotion and self-report measures in empathic situations are influenced by sensorimotor circuit inhibition by low-frequency rTMS. *Brain Stimulation*, 5, 330–336. <http://dx.doi.org/10.1016/j.brs.2011.05.004>
- Baltes, F. R., & Miu, A. C. (2014). Emotions during live music performance: Links with individual differences in empathy, visual imagery, and mood. *Psychomusicology: Music, Mind, and Brain*, 24, 58–65. <http://dx.doi.org/10.1037/pmu0000030>
- Barbiere, M. J., Vidal, A., & Zellner, D. A. (2007). The color of music: Correspondence through emotion. *Empirical Studies of the Arts*, 25, 193–208. <http://dx.doi.org/10.2190/A704-5647-5245-R47P>
- Barwell, I. (1986). How does art express emotion? *The Journal of Aesthetics and Art Criticism*, 45, 175–181. <http://dx.doi.org/10.2307/430558>
- Benjamin, W. (1969). *Illuminations*. New York, NY: Schocken.
- Berger, J. (1972). *Ways of seeing*. London, UK: Penguin.
- Berlyne, D. E. (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation*. Oxford, UK: Hemisphere.
- Bishop, C. (2005). *Installation art: A critical history*. London, UK: Tate Publishing.
- Carstensen, L. L., Pasupathi, M., Mayr, U., & Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology*, 79, 644–655. <http://dx.doi.org/10.1037/0022-3514.79.4.644>
- 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. <http://dx.doi.org/10.2190/EM.28.2.f>
- Clarke, E., DeNora, T., & Vuoskoski, J. (2015). Music, empathy and cultural understanding. *Physics of Life Reviews*, 15, 61–88. <http://dx.doi.org/10.1016/j.plrev.2015.09.001>
- Collier, G. L. (2007). Beyond valence and activity in the emotional connotations of music. *Psychology of Music*, 35, 110–131. <http://dx.doi.org/10.1177/0305735607068890>
- Cupchik, G. C., & Laszlo, J. (1992). *Emerging visions of the aesthetic process*. New York, NY: Cambridge University Press.
- Currie, G. (2011). Empathy for objects. In A. Coplan & P. Goldi (Eds.), *Empathy: Philosophical and psychological perspectives* (pp. 82–98). UK: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199539956.003.0007>
- Dawson, M. E., Schell, M. S., & Fillion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 200–222). UK: Cambridge University Press.
- de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, 10, 435–441. <http://dx.doi.org/10.1016/j.tics.2006.08.008>
- Dewey, J. (1980). *Art as experience*. New York, NY: GP Putnam's Sons.
- Dimberg, U., Andréasson, P., & Thunberg, M. (2011). Emotional empathy and facial reactions to facial expressions. *Journal of Psychophysiology*, 42, s50.
- Dissanayake, E. (2008). The arts after Darwin: Does art have an origin and adaptive function? In K. Zijlman & W. van Damme (Eds.), *World art studies: Exploring concepts and approaches* (pp. 241–263). Amsterdam: Valiz.
- Dubal, S., Lerebours, A. E., Taffou, M., Pelletier, J., Escande, Y., & Knoblauch, K. (2014). A Psychophysical exploration of the perception of emotion from abstract art. *Empirical Studies of the Arts*, 32, 27–41. <http://dx.doi.org/10.2190/EM.32.1.EOV.4>
- Eerola, T., Vuoskoski, J. K., & Kautiainen, H. (2016). Being moved by unfamiliar sad music is associated with high empathy. *Frontiers in Psychology*, 7, 1176. <http://dx.doi.org/10.3389/fpsyg.2016.01176>
- Eerola, T., Vuoskoski, J. K., Peltola, H., Putkinen, V., & Schäfer, K. (2018). An Integrative Review of the Enjoyment of Sadness Associated with Music. *Physics of Life Reviews*, 25, 100–121. <http://dx.doi.org/10.1016/j.plrev.2017.11.016>
- Egermann, H., & McAdams, S. (2013). Empathy and emotional contagion as a link between recognized and felt emotions in music listening. *Music Perception*, 31, 139–156. <http://dx.doi.org/10.1525/mp.2013.31.2.139>
- Eisenberg, N., & Fabes, R. A. (1990). Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion*, 14, 131–149. <http://dx.doi.org/10.1007/BF00991640>
- Ekman, P., Friesen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology*, 39, 1125–1134. <http://dx.doi.org/10.1037/h0077722>
- Elkins, J. (1999). *What painting is: How to think about oil painting, using the language of alchemy*. New York, NY: Routledge.
- Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, 11, 197–203. <http://dx.doi.org/10.1016/j.tics.2007.02.003>
- Frith, U., & Frith, C. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 365, 165–176. <http://dx.doi.org/10.1098/rstb.2009.0160>
- Gage, J. (1999). *Color and meaning: Art, science and symbolism*. Berkeley: University of California Press.
- Gell, A. (1998). *Art and agency, an anthropological theory*. Oxford, UK: Clarendon Press.
- Gerger, G., Leder, H., & Kremer, A. (2014). Context effects on emotional and aesthetic evaluations of artworks and IAPS pictures. *Acta Psychologica*, 151, 174–183. <http://dx.doi.org/10.1016/j.actpsy.2014.06.008>
- Gerger, G., Pelowski, M., & Leder, H. (2017). Empathy, Einfühlung, and aesthetic experience: The effect of emotion contagion on appreciation of representational and abstract art using fEMG and SCR. *Cognitive Processing*, 19, 167.
- Goldin, P. R., McRae, K., Ramel, W., & Gross, J. J. (2008). The neural bases of emotion regulation: Reappraisal and suppression of negative emotion. *Biological Psychiatry*, 63, 577–586. <http://dx.doi.org/10.1016/j.biopsych.2007.05.031>
- Goldstein, T. R., & Winner, E. (2012). Enhancing empathy and theory of mind. *Journal of Cognition and Development*, 13, 19–37. <http://dx.doi.org/10.1080/15248372.2011.573514>
- Goldwater, R., & Treves, M. (1958). *Artists on art from the XIV to the XX century* (3rd ed.). New York, NY: Pantheon.
- Gombrich, E. H. (1963). *Meditations on a hobby horse*. London, UK: Phaidon.
- Gombrich, E. H., & Saw, R. (1962). Symposium: Art and the language of the emotions. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 36, 215–246.
- Goodman, N. (1968). *Languages of art: An approach to a theory of symbols*. Indianapolis, IN: Hackett publishing.
- Harris, D. A. (2007). Dance/movement therapy approaches to fostering resilience and recovery among African adolescent torture survivors. *Torture*, 17, 134–155.
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, 27, 46–51. <http://dx.doi.org/10.3758/BF03203619>
- Herman, D. (2007). Storytelling and the sciences of mind: Cognitive narratology, discursive psychology, and narratives in face-to-face interaction. *Narrative*, 15, 306–334. <http://dx.doi.org/10.1353/nar.2007.0023>
- Hoenen, M., Lübke, K. T., & Pause, B. M. (2017). Sensitivity of the human mirror neuron system for abstract traces of actions: An EEG-study. *Biological Psychology*, 124, 57–64. <http://dx.doi.org/10.1016/j.biopsycho.2017.01.010>
- Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31, 559–575. <http://dx.doi.org/10.1017/S0140525X08005293>

- Kandinsky, W. (1947). *Concerning the spiritual in art*. New York, NY: Wittenborn, Schutz, Inc.
- Korb, S., With, S., Niedenthal, P., Kaiser, S., & Grandjean, D. (2014). The perception and mimicry of facial movements predict judgments of smile authenticity. *PLoS ONE*, 9, e99194. <http://dx.doi.org/10.1371/journal.pone.0099194>
- Kozbelt, A. (2006). Dynamic evaluation of Matisse's 1935 large reclining nude. *Empirical Studies of the Arts*, 24, 119–137. <http://dx.doi.org/10.2190/A2VY-TEBW-VH45-285E>
- Kreitler, H., & Kreitler, S. (1972). *Psychology of the arts*. Durham, NC: Duke University Press.
- Künecke, J., Hildebrandt, A., Recio, G., Sommer, W., & Wilhelm, O. (2014). Facial EMG responses to emotional expressions are related to emotion perception ability. *PLoS ONE*, 9, e84053. <http://dx.doi.org/10.1371/journal.pone.0084053>
- Lamm, C., Porges, E. C., Cacioppo, J. T., & Decety, J. (2008). Perspective taking is associated with specific facial responses during empathy for pain. *Brain Research*, 1227, 153–161. <http://dx.doi.org/10.1016/j.brainres.2008.06.066>
- Lanzoni, S. (2009). Practicing psychology in the art gallery: Vernon Lee's aesthetics of empathy. *Journal of the History of the Behavioral Sciences*, 45, 330–354. <http://dx.doi.org/10.1002/jhbs.20395>
- Leder, H., Bär, S., & Topolinski, S. (2012). Covert painting simulations influence aesthetic appreciation of artworks. *Psychological Science*, 23, 1479–1481. <http://dx.doi.org/10.1177/0956797612452866>
- 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. <http://dx.doi.org/10.1080/02699931.2013.870132>
- Leder, H., Gerger, G., Dressler, S. G., & Schabmann, A. (2012). How art is appreciated. *Psychology of Aesthetics, Creativity, and the Arts*, 6, 2–10. <http://dx.doi.org/10.1037/a0026396>
- AQ: 14 Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, 63, 234–246. <http://dx.doi.org/10.1037/0022-3514.63.2.234>
- Levinson, J. (1990). *Music, art, and metaphysics*. UK: Oxford University Press.
- Lipps, T. (1903). Einfühlung, innere Nachahmung und Organempfin [Empathy, Inner Imitation and Sense-Feelings]. *den Archiv für die Gesamte Psychologie*, 1, 185–204.
- AQ: 16 Marks, L. E. (1983). Categories of perceptual experience: A psychophysicist peruses synesthetic metaphors. *Advances in Psychology*, 11, 336–353. [http://dx.doi.org/10.1016/S0166-4115\(08\)62070-1](http://dx.doi.org/10.1016/S0166-4115(08)62070-1)
- Mayer, J. D., DiPaolo, M., & Salovey, P. (1990). Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence. *Journal of Personality Assessment*, 54, 772–781. <http://dx.doi.org/10.1080/00223891.1990.9674037>
- Melcher, D., & Bacci, F. (2013). Perception of emotion in abstract artworks: A multidisciplinary approach. *Progress in Brain Research*, 204, 191–216. <http://dx.doi.org/10.1016/B978-0-444-63287-6.00010-5>
- Menninghaus, W. (2003). *Disgust. Theory and history of a strong sensation*. Albany, NY: SUNY Press.
- Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., & Jacobsen, T. (2015). Towards a psychological construct of being moved. *PLoS ONE*, 10, e0128451. <http://dx.doi.org/10.1371/journal.pone.0128451>
- Miu, A. C., & Baltes, F. R. (2012). Empathy manipulation impacts music-induced emotions: A psychophysiological study on opera. *PLoS ONE*, 7, e30618. <http://dx.doi.org/10.1371/journal.pone.0030618>
- Omigie, D. (2015). Music and literature: Are there shared empathy and predictive mechanisms underlying their affective impact? *Frontiers in Psychology*, 6, 1250. <http://dx.doi.org/10.3389/fpsyg.2015.01250>
- Osborne, P. (2002). *Conceptual art: Themes and movements*. London, UK: Phaidon.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Pelowski, M. (2015). Tears and transformation: Feeling like crying as an indicator of insightful or “aesthetic” experience with art. *Frontiers in Psychology*, 6, 1006. <http://dx.doi.org/10.3389/fpsyg.2015.01006>
- Pelowski, M., Forster, M., Tinio, P. P., Scholl, M., & Leder, H. (2017). Beyond the lab: An examination of key factors influencing interaction with ‘real’ and museum-based art. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 245–264. <http://dx.doi.org/10.1037/aca0000141>
- Pelowski, M., Leder, H., Mitschke, V., Specker, E., Gerger, G., Tinio, P. P. L., . . . Husslein-Arco, A. (2018). Capturing aesthetic experiences with installation art: An empirical assessment of emotion, evaluations, and mobile eye tracking in Olafur Eliasson's ‘Baroque, Baroque!’ *Frontiers in Psychology*, 9, 1255. <http://dx.doi.org/10.3389/fpsyg.2018.01255>
- 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. <http://dx.doi.org/10.1016/j.plrev.2017.02.003>
- AQ: 19 Pelowski, M., Markey, P. S., Luring, 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. <http://dx.doi.org/10.3389/fnhum.2016.00160>
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17, 715–734. <http://dx.doi.org/10.1017/S0954579405050340>
- Ramachandran, V. S., & Hirstein, W. (1999). The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies*, 6, 15–51.
- Reiss, J. H. (2001). *From margin to center: The spaces of installation art*. Cambridge, MA: MIT Press.
- Reniers, R. L., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, 93, 84–95. <http://dx.doi.org/10.1080/00223891.2010.528484>
- Sato, W., Fujimura, T., Kochiyama, T., & Suzuki, N. (2013). Relationships among facial mimicry, emotional experience, and emotion recognition. *PLoS ONE*, 8, e57889. <http://dx.doi.org/10.1371/journal.pone.0057889>
- Sbriscia-Fioretti, B., Berchio, C., Freedberg, D., Gallese, V., & Umiltà, M. A. (2013). ERP modulation during observation of abstract paintings by Franz Kline. *PLoS ONE*, 8, e75241. <http://dx.doi.org/10.1371/journal.pone.0075241>
- Schaefer, A., Collette, F., Philippot, P., van der Linden, M., Laureys, S., Delfiore, G., . . . Salmon, E. (2003). Neural correlates of “hot” and “cold” emotional processing: A multilevel approach to the functional anatomy of emotion. *NeuroImage*, 18, 938–949. [http://dx.doi.org/10.1016/S1053-8119\(03\)00009-0](http://dx.doi.org/10.1016/S1053-8119(03)00009-0)
- Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. *Psychological Inquiry*, 14, 296–303. <http://dx.doi.org/10.1080/1047840X.2003.9682896>
- Sherman, A., & Morrissey, C. (2017). What is art good for? The socio-epistemic value of art. *Frontiers in Human Neuroscience*, 11, 411. <http://dx.doi.org/10.3389/fnhum.2017.00411>
- Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156, 81–96. <http://dx.doi.org/10.1111/j.1749-6632.2009.04418.x>
- Sloboda, J. (1992). Empirical studies of emotional response to music. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 33–46). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/10104-003>

- Sonnby-Borgstrom, M. (2002). Automatic mimicry reactions as related to differences in emotional empathy. *Scandinavian Journal of Psychology*, 43, 433–443. <http://dx.doi.org/10.1111/1467-9450.00312>
- Sonnby-Borgstrom, M., Jönsson, P., & Svensson, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behavior*, 27, 3–23. <http://dx.doi.org/10.1023/A:1023608506243>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31, 137–149. <http://dx.doi.org/10.3758/BF03207704>
- Swanger, D. (1993). The arts, empathy, and Aristotle. *Journal of Aesthetic Education*, 27, 41–49. <http://dx.doi.org/10.2307/3333340>
- Takahashi, S. (1995). Aesthetic properties of pictorial perception. *Psychological Review*, 102, 671–683. <http://dx.doi.org/10.1037/0033-295X.102.4.671>
- Tinio, P. P. (2013). From artistic creation to aesthetic reception: The mirror model of art. *Psychology of Aesthetics, Creativity, and the Arts*, 7, 265–275. <http://dx.doi.org/10.1037/a0030872>
- Tolstoy, L. (1896). *What is art?* (A. Maude, Trans.). Indianapolis, IN: The Liberal Arts Press.
- Trainor, L. J., & Trehub, S. E. (1992). The development of referential meaning in music. *Music Perception*, 9, 455–470. <http://dx.doi.org/10.2307/40285565>
- Umla, M. A., Berchio, C., Sestito, M., Freedberg, D., & Gallese, V. (2012). Abstract art and cortical motor activation: An EEG study. *Frontiers in Human Neuroscience*, 6, 311. <http://dx.doi.org/10.3389/fnhum.2012.00311>
- van Leeuwen, C., Verstijnen, I., & Hekkert, P. (1999). Common unconscious dynamics underlie uncommon conscious effects: A case study in the iterative nature of perception and creation. In J. S. Jordan (Ed.), *Modeling consciousness across the disciplines* (pp. 179–218). Lanham, MD: University of America 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, e0134241. <http://dx.doi.org/10.1371/journal.pone.0134241>
- Vischer, R. (1873). On the optical sense of form: A contribution to aesthetics. In H. F. Mallgrave (Ed.), *Empathy, form, and space: Problems in German aesthetics 1873–1893*. IL: University of Chicago Press.
- Wagner, V., Menninghaus, W., Hanich, J., & Jacobsen, T. (2014). Art schema effects on affective experience: The case of disgusting images. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 120–129. <http://dx.doi.org/10.1037/a0036126>
- Wedin, L. (1972). A multidimensional study of perceptual-emotional qualities in music. *Scandinavian Journal of Psychology*, 13, 241–257. <http://dx.doi.org/10.1111/j.1467-9450.1972.tb00072.x>
- Weingarden, L. S. (2014). Installation art and experiential aesthetics: Mediating art history and neuroscience research. In *Twenty-third Biennial Congress of the International Association of Empirical Aesthetics*. New York, NY: International Association of Empirical Aesthetics. Retrieved from <http://www.science-of-aesthetics.org/data/proceedings/IAEACongressProceedings2014.pdf>
- Weingarden, L. S. (2015). Framing an experiential aesthetics: The affective turn in contemporary installation art. In V. A. Journeau & C. V. Kayser (Eds.), *Notions esthétiques (II): la perception sensible organisée* [Aesthetic notions (II): organized sensory perception] (pp. 197–210). Paris, France: Editions L'Harmattan, collection L'univers esthétique. **AQ: 20**
- Wöllner, C. (2012). Is empathy related to the perception of emotional expression in music? A multimodal time-series analysis. *Psychology of Aesthetics, Creativity, and the Arts*, 6, 214–223. <http://dx.doi.org/10.1037/a0027392>
- Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8, 494–521. <http://dx.doi.org/10.1037/1528-3542.8.4.494>
- Zentner, M., Meylan, S., & Scherer, K. (2000, August). *Exploring musical emotions across five genres of music*. Paper presented at the Sixth Conference of the International Society for Music Perception and Cognition (IMPC), Keele, UK.
- Zickfeld, J. H., Schubert, T. W., Seibt, B., & Fiske, A. P. (2017). Empathic Concern is part of a more general communal emotion. *Frontiers in Psychology*, 8, 723. <http://dx.doi.org/10.3389/fpsyg.2017.00723>

Received February 6, 2018

Revision received July 31, 2018

Accepted August 19, 2018 ■

Supplementary Materials (Integral)

Table S1

Participant answers to “What did the art mean?”

Artwork 1 (participant)	Artwork 2	Artwork 3
Artist intended meaning (P1): Commented on a cultural aspect of blackness, caused me to think about womanhood and the intersectionality of race and gender. (P5): Expression of relationship to culture and hair dynamics. (P6): Well I think it's about sharing life experience but I could be totally wrong. (P8): Identity, placement. (P9): Hair, identity. (P11): Perhaps it was a memory about brushing hair or about beauty standards. (P14): It was trying to convey a black woman letting down her hair and showing her natural curls. She's breaking society's standards by going natural. (P16): It's about her hair, the identity she finds in it, not necessary to contain it. (P18): They have empowered the viewer with kinky hair to face the stigma they have. (P19): Exploring the class and energy inherent in black culture. (P20): The segments of hair represent a source of identity for the artist. (P22): The meaning of black hair in US society. (P23): Somethings about female representation in media. (P24): Nappy hair? Chaos? (P25): It seems to be obvious hair and hair supplies, the anxiety from trying to tame chaos, or not chaos but differences. Let it be free to be wild and not conform to the social "order" or "norm." (P30): Hair - particularly natural curly hair. The efforts to tame it under the pressure of society. (P31): I think the art has something to do with African American intimacy being on display - hair care becoming a landscape to be enjoyed and not hidden away. (P34): It's likely the artist's hair being represented and she is defining herself as hair, which kind of strips her of her identity. And women tend to be judged by the quality of their hair. Other meaning (P2): How bits of hair give spaces a "lived in" feeling and remind us about people. (P7): It reminds me of natural hair. (P32): HAIR. (P33): hair pervading environment - except for bald people. (P37): identity of being black women in hallways as janitors. Their thoughts roaming and leaving a remnant in the space. No meaning/No answer (---) (P3): ? (P12): Not sure.	Artist intended meaning (P2): The difficult associations with doors. I think it spoke to how doors can lead to bad places. (P5): Something dark/not sure - maybe threatening of some kind. (P8): Trauma. (P7): The art was a literal representation of where the artist had been assaulted. (P10): It represents the idea that when people are raped they should have known it was going to happen (what they were walking into). (P11): When walking through a door, you don't really know what is on the other side. You really don't know what you're getting into, but it's also not something you really think about too hard because you don't think anything will happen. (P12): Date Rape. (P14): Rapes that happened in 3 different rooms. (P15): Bad things happened behind these doors. (P22): Emotional/ sexual abuse. (P30): sexual assault/rape - how potent and ignored it is. (P26): Put you in the space of walking into something bad. (P31): The art was about some sort of intimate trauma. --> the choice of walking through a door and not being able to come back. Other meaning (P9): Choices, sex, outcomes, hesitancy about intimacy, power, disempowerment, class. (P16): I think it was an introduction to the next two rooms. (P18): Door provoke mystery. (P19): The information given to us through television (especially political) is inorganic and supports itself-industrially. (P20): A sense of a transforming journey with multiple stages. (P24): Different doorways of life? (P25): 3 different places with discontent between them. The emotions discarded at the doorsteps. (P27): It could be cathartic to those who have maybe experienced something agonistic and subliminal. (P32): Doorways led to shame/nudity/fear. (P34): Different pathways, different people, different lifestyles. I wonder if the flickering projection was meant to point out the imperfections that outsiders tend to ignore when looking at the unknown. (P37): Leaving and coming back home, from work, starting the day getting dressed, undressed after work. Home porch entry? Space/private/intimacy. No meaning/No answer (---) (P1): --- (P3): --- (P4): --- (P6): --- (P13): --- (P17): ---	Artist intended meaning (P8): How we receive information, technology. (P14): Truly don't know. Best guess is TV fueling society and politics controlling us. (P22): How do we process and receive information. (P25): Maybe moments with family on field. Maybe capitalism and all the infrastructure destroying nature. So possibly political. (P33): Media is a pillar of daily existence; constant news of the past is particularly nostalgic today, considering climate today. Other meaning (P2): How all these unnatural man-made stuff (TVs, shows, elections, etc.) are birthed from a natural source. (P5): Political commentary on America. (P7): Definitely a political statement about the most recent presidency and their thoughts/actions against the environment. (P9): Conflict, politics, entertainment, environment; hypocrisy; political and cultural machines, history of bullshit. (P10): That the words from the politicians are powering the change in the environment? But they didn't happen so it's not growing well? Idk. (P11): I think it was something about how our political system and the entertainment industry is destroying our environment. (P12): Politics and the 1990s. (P19): Something about [?] (conceptually) and movement and the self. (P20): The art raised questions about climate change, the selfish nature of human beings, and materialism. (P23): [?] into an unmown situation, bare. (P24): Alternative energy source? (P27): I think this piece was a comment on sport and hyper-masculinity (re. phallic covered in sod + women on TV). (P30): environmental issues - political noise. (P31): The art seemed to be about [?] Power - the lack of distinction b/w dramatized news and political dramas. (P32): Green technology. (P34): environmental and political themes the connection between the two -- one feeling the other. (P37): America, Capitalism, Production, Economy, Politics. No meaning/No answer (---) (P1): I have no understanding of the work. (P6): I really did not understand. (P16): I have no idea. (P18): I have no idea. (P26): I have no idea. (P3): ---

(P26): Nothing.
(P4): ---
(P10): ---
(P13): ---
(P15): ---
(P17): ---
(P21): ---
(P27): ---
(P28): ---
(P29): ---
(P35): ---
(P36): ---

(P21): ---
(P23): ---
(P28): ---
(P29): ---
(P33): ---
(P35): ---
(P36): ---

(P4): ---
(P13): ---
(P15): ---
(P17): ---
(P21): ---
(P28): ---
(P29): ---
(P35): ---
(P36): ---

Note. Answers coded by two independent scorers for general match or mismatch to artists' written statement concerning intended meaning.

2.3.3

Grüner, S., **Specker, E.**, & Leder, H. (2019). Effects of Context and Genuineness in the Experience of Art. *Empirical Studies of the Arts*, <https://doi.org/10.1177/0276237418822896>

My Contribution:

In this case, the concept and idea underlying the manuscript was developed by GS and HL. GS and HL were responsible for the study design. This paper is the result of a master thesis supervised by HL and co-supervised by me. I assisted GS in the preparation of the study and data collection. In addition, I assisted GS throughout the analysis process. Though GS is the main author of the manuscript, we closely collaborated and wrote it together. Once a more refined version of the manuscript was established, we shared it HL who provided feedback on consecutive versions. I then assisted GS in the submission and review process.

Effects of Context and Genuineness in the Experience of Art

Empirical Studies of the Arts

0(0) 1–15

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0276237418822896

journals.sagepub.com/home/art



Susanne Grüner^{1,2}, Eva Specker^{2,3},
and Helmut Leder¹

Abstract

Studies found that genuine artworks viewed in a museum receive higher appreciation ratings than reproductions in the laboratory. Due to the mutual variation of context and genuineness, these studies were not able to disentangle these factors. A study designed by Brieber, Leder, and Nadal to systematically differentiate between these two variables did not find an effect of context or genuineness. To substantiate these results, we setup a conceptual replication by using the same 2 (museum/laboratory) by 2 (genuine/reproduction) between-subjects design with improved manipulations of context and genuineness. We found an effect of context, as artworks presented in a museum were liked more and rated more interesting than in the laboratory. We did not find effects of genuineness. Exploratively, we found that art style had a big impact on how artworks were rated regardless of context and genuineness, indicating that this may be a more important factor for aesthetic experience.

Keywords

museum study, style, original, reproduction, painting, art appreciation, art evaluation

¹Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, Austria

²Department of Traffic and Engineering Psychology, Faculty of Psychology, Technische Universität Braunschweig, Germany

³Department of Art History, Faculty of Historical and Cultural Studies, University of Vienna, Austria

Corresponding Author:

Susanne Grüner, Department of Traffic and Engineering Psychology, Faculty of Psychology, Technische Universität Braunschweig, Braunschweig, Germany.

Email: susanne.gruener@tu-braunschweig.de

When thinking about the fascination people have with art, famous museums and highly priced single pieces come to mind. But, explaining why art really touches us seems difficult. Researchers suggest that other factors, besides the artwork itself, influence our art experience. Two such factors are context and genuineness, the location in which the artwork is presented, and the fact that in a museum we see the real, unique artwork. However, in empirical studies, up until now, these effects have often been confounded. Typically, genuine artworks are viewed in a museum context, while in the context of comparison, the laboratory, reproductions of artworks are presented. Therefore, it is one of the greatest challenges for research in this field to disentangle these two factors in order to examine the impact of them separately as well as combined.

This article aims to address this issue and investigates the effects of context and genuineness in a way that differentiates between these two variables. We first give a short review of the literature on context and genuineness and consequently discuss this study.

Context

Previous research has established that visiting an exhibition in a museum is generally pleasurable (Smith & Wolf, 1996). Museums are places where people take time to reflect and enjoy cultural heritage (Smith, 2014). Most research to date has investigated art experience of the different stimuli in only one type of context (Locher & Dolese, 2004; Locher, Smith, & Smith, 1999, 2001). This consequently does not allow for a direct assessment of the influence of context, for this one would need at least two different types of context. Some research has investigated the influence of museum context directly, finding enhanced aesthetic experience and higher ratings in aesthetic value, liking, and interest in museums (Brieber, Nadal, & Leder, 2015; Specker, Tinio, & van Elk, 2017) as compared with laboratories. In addition, people look longer at artworks in a museum context than in a laboratory context (Brieber, Nadal, Leder & Rosenberg, 2014). Artworks, as well as added information concerning them, were remembered better and longer when they were viewed in a museum context (Brieber, Nadal, et al., 2015; Specker et al., 2017). Together, the presented studies draw a consistent picture of the positive effect of museum contexts on art experience.

Genuineness

Different theories exist in the literature on how genuineness can impact art experience. The majority of the literature on genuineness pays particular attention to *ideal aspects*, for example, Benjamin (2010), when discussing the *aura* of an artwork as a standalone value. Seeing it from a different angle, Newman and Bloom (2012) explained the impact of genuineness through two aspects: An artwork is understood first as a result of a performance and second as a

source of physical contact with the artist. Despite the theoretical importance of genuineness, empirical work is scarce, probably due to practical concerns such as limited access to genuine artworks, which leads to difficulties in running scientific experiments. Coping successfully with this challenge, Locher et al. (1999, 2001) showed that only a quarter of their stimuli set of genuine artworks was rated as more interesting and more pleasant than replications (slides and computer; Locher et al., 1999), and that only art-trained people rated genuine artworks differently from reproductions (Locher et al., 2001). Art-untrained people did not rate genuine artworks differently from their reproductions.

In reaction to these unexpected findings, Locher et al. (2001) proposed the process of facsimile accommodation as an explanation: If people are aware that the object they are evaluating is a reproduction and if this reproduction is of high quality, they will automatically adjust their evaluation of an artwork as if they were looking at the original artwork. This hypothesis may be an explanation for the lack of a genuineness effect in previous studies. Locher et al. (2001) attributed importance to the special role of reproduction quality because reproductions are typically of a lower quality than the original and have a reduced level of artistic finesse. Size, relief structures, deepness of color, and so on are just some artistic aspects that can for the most part not be reproduced with the media used. Therefore, it is an open question whether and how the quality of reproductions plays a role in the effects of genuineness.

To summarize, context and genuineness usually are highly confounded variables in empirical research. Therefore, none of the discussed studies can clearly determine whether the found effects are due to context or to genuineness. The only study up to now that aimed to address this issue and disentangle these two factors is a study by Brieber, Leder, and Nadal (2015). They used a 2 (laboratory vs. gallery) by 2 (genuine vs. reproduction) between-subjects design. The study was conducted in two rooms of a gallery (gallery condition) and a laboratory room (laboratory condition). Contemporary photographs were used as stimuli, which were either hanged on the wall (genuine condition) or shown in a computer presentation (reproduction condition). Participants reported their experience in terms of liking, interest, arousal, valence, and understanding. In contrast to Brieber, Leder, et al.'s (2015) expectations, no main effects (either of context or genuineness) were found, nor did they find interaction effects. These findings are in strong contrast to the existing literature. This contrast creates a necessity of further research to gain a better understanding of the (lack of) effects of context and genuineness reported by Brieber, Leder, et al. (2015). This study aims to fill this need.

The present study

This study aims to assess the effect of context and the effect of genuineness further. To do this, we carried out a conceptual replication of Brieber, Leder,

et al. (2015). We used the same 2×2 design to allow for comparability of the findings but improved the manipulation of context and genuineness, as we are critically questioning the effectiveness of their manipulation. In addition, we exploratively investigated the influence of art style on art experience. Next, we shortly discuss the improvements made and our motivation for them, first regarding context, then regarding genuineness.

Context

Brieber, Leder, et al. (2015) addressed the potential of an inverse white cube effect. The white cube effect refers to the fact that a museum context can transform any kind of ordinary object into an art object or artwork. However, the contrary may also occur: Not only can a museum context turn things into art but also artworks hanging in a random physical context can turn this context into a museum. Brieber, Leder, et al. (2015) later supposed that the photographs could have turned the laboratory into a gallery. So do we, and critically question the effectiveness of the context manipulation as well. Participants might have felt as if they were in a gallery context rather than in a laboratory. In this study, we used separated places that represented real museum or real laboratory contexts. As Brieber, Leder, et al. (2015) did, we made sure that both contexts were located in different buildings. In contrast to Brieber, Leder, et al. (2015), we additionally used a museum rather than a commercial gallery context. As a museum of art is an even more stereotypical art exhibition context than a commercial gallery, especially for laypeople, we aimed to enhance the manipulation of context. In addition, as a commercial gallery space is not only a place where art is exhibited but also a place where it is sold, it could be that these monetary aspects of the gallery had an influence on the effect. Using a museum space controlled for these potential confounds. Especially because the museum in which we conducted the study is a city museum, which has a nonprofit educational task and free entrance for all visitors. Furthermore, we kept room size, lighting, as well as distance to the artworks in both contexts constant, as these are well-known confounding variables (Griswold, Mangione, & McDonnell, 2013, p. 352; Pelowski, Forster, Tinio, Scholl, & Leder, 2017).

This leads to improved control of confounding variables like physical characteristics or expectancy of the participants (Cupchik, 2002; Locher, Overbeeke & Wensveen, 2010; Scherer, 2005). The participants' expectations were and were not fulfilled in either of the two conditions, as people expect genuine artworks in a museum and reproductions in a laboratory but not vice versa.

Genuineness

Brieber, Leder, et al. (2015) discussed weaknesses of the used medium (photography). Due to the widespread digital presentation of photographs, digital

reproductions may have a similar level of assumed genuineness as photographic prints, especially by lay viewers. Therefore, in this study, we only used artworks painted on canvas in the genuine condition and high-quality pictures in the computer presentation for the reproduction condition. We postulated that, especially for a lay sample, paintings would be seen as more genuine than photographs (Dutton, 2009) and would crucially have a higher level of assumed genuineness than their digital reproductions.

For a satisfying comparison, various features have to be matched. As a central point, reproductions were shown in exactly the same size as the genuine artworks to exclude size as a confound (Pelowski et al., 2017). In keeping this feature constant between conditions, this study seems to be a first fair test of art in different contexts.

In addition, we did not provide an information about any of the artworks. In the study by Brieber, Leder, et al. (2015), additional information was given for some of the artworks. This information was mostly technical data about the photographs. This led Brieber, Leder et al. (2015) to suggest that perhaps the participants did not *feel* the artwork. They did not focus on the content of the artwork but more on the explained abstract aspects of the medium photography. As information about artworks is typically offered in a museum context, omitting information may seem rather unnatural in this context. However, we feel that the added gains of comparability and exclusion of potential effects of information outweighed the potential drawbacks.

As an explorative extension, this study investigates the impact of art style on art experience in direct comparison with context and genuineness. Effects of art style have been consistently found with reproductions in the laboratory (Leder, Gerger, Dressler, & Schabmann, 2012) as well as with genuine artworks in the museum (Mastandrea et al., 2018; Tröndle & Tschacher, 2012). But evidence is lacking that the effect of art style is not influenced by context or genuineness themselves. To analyze the impact of art style in this study, we showed five artworks in abstract style and five in figurative style.

In sum, these improvements to the experimental design aim to create the optimal conditions for finding the effects of context and genuineness. As Brieber, Leder, et al. (2015), we hypothesized that both context and genuineness positively affect art experiences. Specifically, we hypothesize that art experience is enhanced in a museum context (vs. a laboratory context) as well as when looking at a genuine artwork (vs. a reproduction).

Method

Participants

Our total sample consisted of 155 participants, aged between 18 and 49 years (mean age 24 years, 108 women). Psychology students from the University of

Vienna were rewarded with course credits for their participation. Six participants had an educational background in art. In view of age, sex, and art education, the participants in this study and the one of Brieber, Leder, et al. (2015) were highly similar.

Materials

Context. There were two different types of context: A museum context (the MUSA: Museum Start Gallery Artothek, Vienna) and a laboratory context (a laboratory room of the University of Vienna). The MUSA was established by the City of Vienna's Department of Culture and exhibits the contemporary art collection of the City of Vienna. The laboratory room is a typical laboratory space that is used for empirical experiments and as computer room on daily basis.

Genuineness. People either saw real artworks (paintings) hanging on the walls (genuine condition) or reproductions of these artworks in a power-point presentation on high-resolution monitors (reproduction condition).

Stimuli. Ten preperated artworks (five oil painting, three mixed technique, one tempera, and one gouache paint) were used; nine from the collection of MUSA and one from the private collection of the third author. To select paintings, we ran an online prestudy in which 20 paintings were rated regarding arousal, complexity, and valence. Based on this, we selected 10 artworks (5 figuratives and 5 abstracts). Furthermore, the 10 paintings were made by nine different artists to keep the influence of individual painting style and specific character of the artist low. In the reproduction condition, artworks were shown in their original size in order to control for a potential confounding effect of artwork size between the two different conditions. In both contexts, the artworks hung at the same height (upper image border—182 cm) and with the same distance to each other (center of the artwork—89 cm).

Art experience. To assess the experience of artworks, we measured arousal, valence, liking, interest, and understanding. Using these scales enables the measurement of affective (valence and arousal) and cognitive (liking, interest, and understanding) aspects of art experience and further enables the comparison with previous studies (Brieber, Leder, et al., 2015; Brieber, Nadal, et al., 2015; Brieber et al., 2014). The operationalization was as follows: Arousal (*How aroused do you feel when looking at this artwork?*), valence (*How does this artwork make you feel?*), liking (*How much do you like this artwork?*), interest (*How interesting do you find this artwork?*), and understanding (*How much do you have a sense of understanding this artwork?*). All were rated on 7-point Likert-type

scales from 1 (*very little*) to 7 (*a lot*), except for valence (1 = *very negative* to 7 = *very positive*) and arousal (1 = *very calm* to 7 = *very excited*).

Art interest and art knowledge. As in Brieber, Leder, et al. (2015), we measured individual differences in art interest and art knowledge. These individual differences are variables that can play a measurable role in art experience (Brieber, Nadal, et al., 2015; Leder, Belke, Oeberst, & Augustin, 2004; Locher et al., 2001). We used the validated Vienna Art Interest Art Knowledge Questionnaire (Specker et al., 2018).

Procedure

As in the study of Brieber, Leder, et al. (2015), participants were asked to freely walk around to view the artworks hanging on the walls (genuine condition) or to take place to view the high-quality digital reproductions on the screen (reproduction condition). Due to the analog way of presenting artworks in the genuine condition, it was not possible to fully randomize the presentation order of the artworks. Therefore, we created three different artwork orders. Participants were assigned to one of the three orders based on the day of participation (order 1: 50 participants, order 2: 57 participants, and order 3: 48 participants). Participants viewed the artworks either in the MUSA (museum condition) or in the laboratory (laboratory condition). Figure 1 provides an overview of the experimental design. There were never more than two participants in the same room. Following the viewing period, participants were given written instructions and were asked to rate the artworks, and the scale order was randomized. The artworks were rated individually; participants of every condition were urged to return to the artwork while rating it in the various scales (by selecting the artwork in the presentation [genuine condition] or physically moving back to the artwork [museum condition]). Afterward, people filled out questionnaires assessing art interest and art knowledge as well as demographic questions.

Results

Participants did not differ in art interest or art knowledge between conditions, art interest: $F(3,151) = .07, p = .974$; art knowledge: $F(3,151) = .95, p = .416$. In general, our sample had a moderate interest in art ($M = 43.92$ on the scale, with a maximum score of 77, standard deviation [SD] = 14.64) and low knowledge in art ($M = 12.51$ on the scale, with a maximum score of 26, $SD = 5.76$).

To analyze the effects of context and genuineness, we ran a multivariate analysis of covariance (MANCOVA) with context (museum/laboratory) and genuineness (genuine/reproduction) as independent variables, and the rating scales (measured arousal, valence, liking, interest, and understanding) as dependent variables. To ensure comparability with the findings of Brieber,

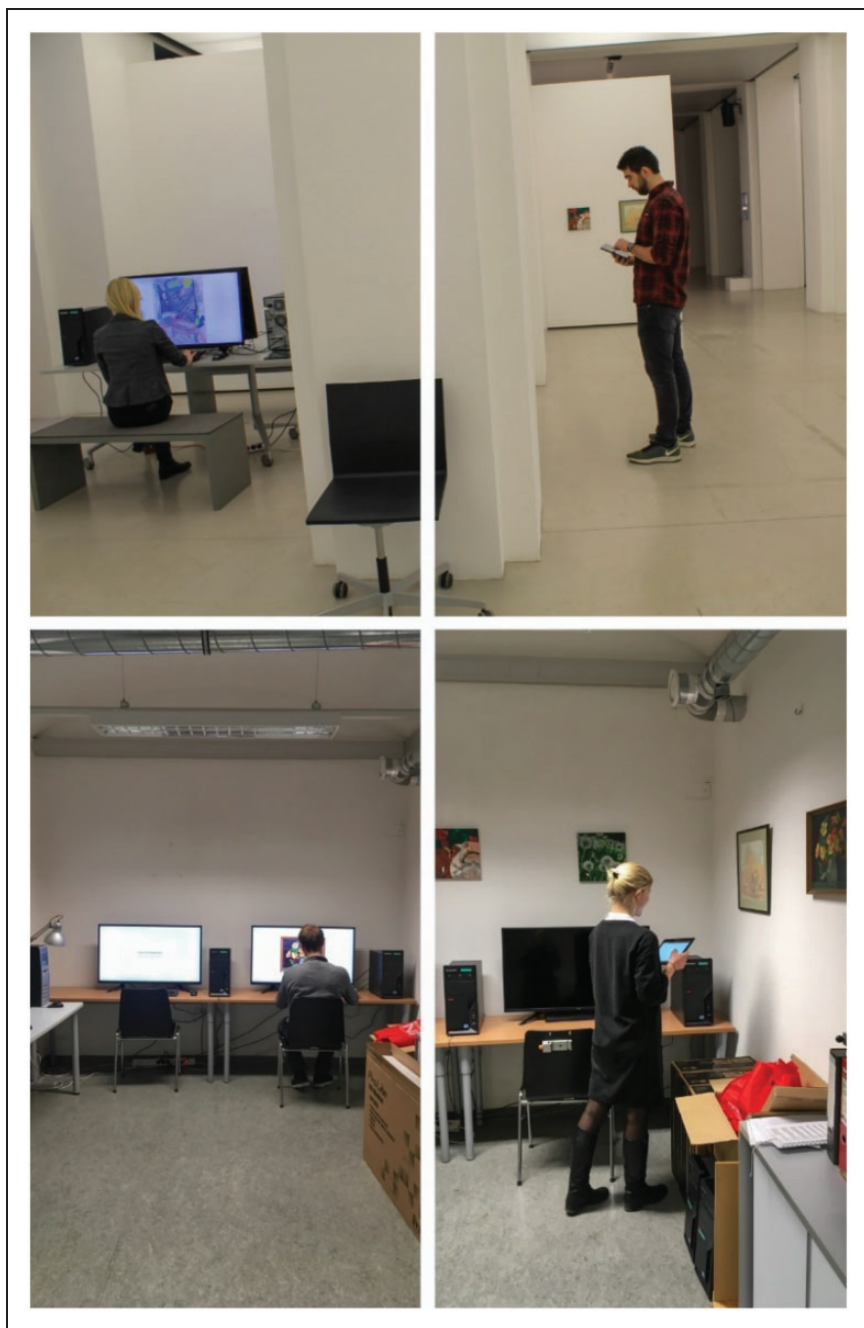


Figure 1. Photographs of the experimental design (top-left: museum/reproduction, top-right: museum/genuine, bottom-left: laboratory/reproduction, and bottom-right: laboratory/genuine).

Table 1. Mean (and Standard Deviations) in Art Experience Divided by Conditions.

	Museum/G enuine (<i>n</i> = 39)	Museum/R eproduction (<i>n</i> = 36)	Laboratory/G enuine (<i>n</i> = 41)	Laboratory/R eproduction (<i>n</i> = 39)
Liking	4.31 (.59)	4.59 (.67)	4.21 (.71)	4.20 (.63)
Interest	4.50 (.64)	4.87 (.55)	4.31 (.70)	4.24 (.71)
Arousal	4.15 (.63)	4.19 (.50)	4.11 (.49)	4.07 (.50)
Valence	4.28 (.46)	4.50 (.53)	4.28 (.51)	4.39 (.44)
Understanding	4.08 (.83)	4.03 (.77)	4.19 (.83)	4.01 (.76)

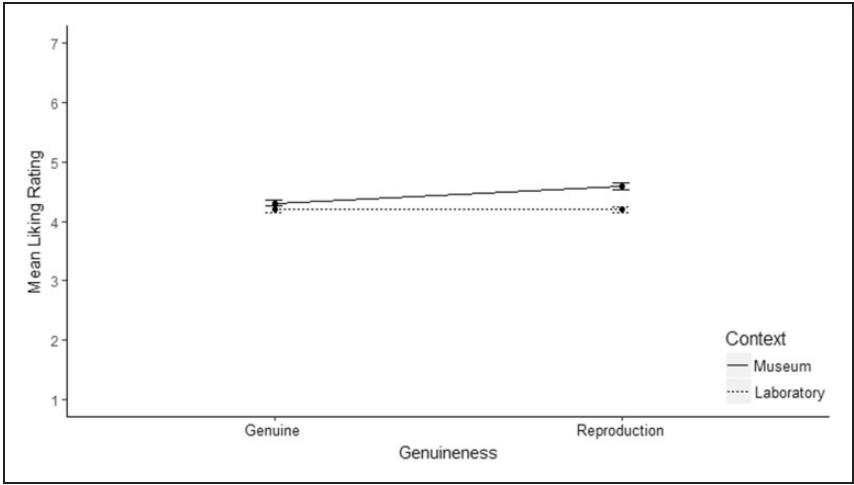


Figure 2. Mean liking rating shown for context and genuineness, with error bars.

Leder, et al. (2015), we included art interest as a covariate. Means and *SDs* divided into condition are given in Table 1.

There was a main effect of context, explaining 11% of variance in the model, Wilks' $\Lambda = .89$, $F(5, 146) = 3.56$, $p = .005$, $\eta_p^2 = .11$. Follow-up analyses showed that context influences liking, $F(1, 150) = 5.27$, $p = .023$, $\eta_p^2 = .03$, and interest, $F(1, 150) = 14.82$, $p < .001$, $\eta_p^2 = .09$. As can be seen in Figure 2, artworks were liked more in the museum ($M = 4.44$, $SD = 0.64$) than in the laboratory ($M = 4.20$, $SD = 0.67$). In addition, they were rated as more interesting in the museum ($M = 4.68$, $SD = 0.62$) than in the laboratory ($M = 4.28$, $SD = 0.70$), shown in Figure 3. As can be seen in Figures 2 and 3 and in Table 1, the difference between the museum and laboratory contexts was more pronounced for the reproductions than for the genuine artworks.

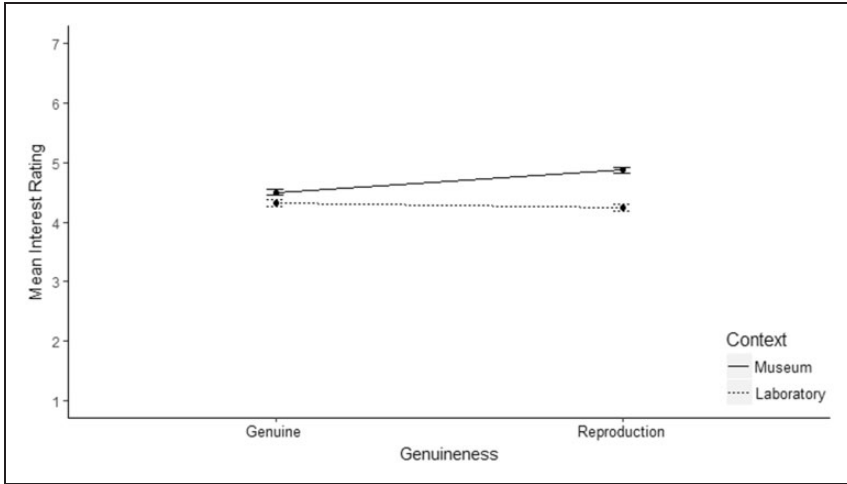


Figure 3. Mean interest rating shown for context and genuineness, with error bars.

Nonetheless, there was no main effect of genuineness, Wilks' $\Lambda = .96$, $F(5, 146) = 1.31$, $p = .265$, $\eta_p^2 = .04$. Nor was there an interaction between context and genuineness, Wilks' $\Lambda = .97$, $F(5, 146) = .88$, $p = .497$, $\eta_p^2 = .03$.

To test for the robustness of the effect, we conducted follow-up analyses. In a first step, we computed a multivariate analysis of variance without art interest as covariate. This was done because, in contrast with Brieber, Leder, et al. (2015), we did not observe a difference in art interest between conditions. This analysis did not change the reported effects. Second, we repeated the initial MANCOVA with exclusion of one artwork that was not presented in the correct size in the reproduction condition; our results did not change when this artwork was excluded.

Explorative Analysis: Art Style

In addition, we wanted to explore the effect of art style. By design, we included both abstract and figurative artworks to enable us to investigate the role of style in art experience in different conditions of art presentation (museum/laboratory as well as genuine/reproduction). To analyze the role of style, we used a mixed-effect MANCOVA with art style as a within-subject factor, context (museum/laboratory) and genuineness (genuine/reproduction) as between-subjects factors, and the rating scales (measured arousal, valence, liking, interest, and understanding) as dependent variables. To ensure comparability, we included art interest as a covariate.

There was a main effect of art style, explaining 44% of variance in the model, Wilks' $\Lambda = .57$, $F(5, 146) = 22.46$, $p < .001$, $\eta_p^2 = .44$. Follow-up analyses showed that art style explained 42% variance of understanding, $F(1, 150) = 107.61$, $p < .001$, $\eta_p^2 = .42$, one can also express the effect size in terms of a standardized mean difference, in this case that would translate to a Cohen's d of 2.09. The understanding of figurative artworks ($M = 5.07$, $SD = 0.82$) was rated significantly higher compared with abstract artworks ($M = 3.09$, $SD = 1.06$).

No interaction was observed between art style and context or genuineness.

Discussion

In contrast with the study of Brieber, Leder, et al. (2015), but in line with the previous literature (Brieber, Nadal, et al., 2015; Brieber et al., 2014; Locher & Dolese, 2004; Locher et al., 1999, 2001), this study found an effect of context on art experience. Our main finding was that people found artworks more interesting and liked artworks more in a museum context than in a laboratory context.

This suggests that the improvements made regarding the manipulation of context were effective. We worked in a real museum, instead of a commercial gallery, to control for confounding variables like physical characteristics or expectancy of the participants (Cupchik, 2002; Locher et al., 2010; Scherer, 2005). This shows that for the effects of context to occur, the manipulation of context should be strong enough. When a museum just does not feel like a museum, or more applicable to the discussion of Brieber, Leder, et al. (2015), when participants still have the sense of being in a gallery context rather than in a *real* laboratory (inverse-with-cube effect), the effect will not appear.

In line with the study by Brieber, Leder, et al. (2015), we did not find an effect of genuineness. This differs from assumptions regarding the positive effect of genuineness, as seeing a genuine artwork is known as one of the main motivations, which drives people to visit a museum in the first place (Mastandrea, Bartoli, & Bove, 2007, 2009). In this study, we aimed to enhance the manipulation of genuineness by using artworks painted on canvas rather than photographs under the assumption that this difference in medium would highlight the genuineness (or nongenuineness) of the artworks.

One possibility is that our manipulation was still not effective. A potential explanation for this is that a potential effect of genuineness is subverted by the high quality of the reproductions. The reproductions were professional photographs shown on a high-resolution 4k monitor with 40-in. diagonal size. Due to the presentation of all artworks in their original size, presentation sometimes took up the full size of the screen. This high quality may have therefore reduced the difference between seeing the reproduction or the genuine or even artificially increased the pleasure in the art experience of the reproductions.

Another possibility for the lack of an effect is the facsimile-accommodation hypothesis as discussed in the introduction of this article. Locher et al. (2001)

argue that if people are aware that they are looking at a reproduction rather than a genuine artwork, they will attempt to adjust their experience accordingly. This hypothesis may be an explanation of the lack of a genuineness effect in general but may be particularly relevant in our case. In this study, participants in the reproduction condition reported that they assumed the artworks to be larger in real life than presented on the screen, even though they were explicitly told that this was not the case and that the presented artworks were shown in their original size. This implicit assumption was probably created by the fact that reproductions are usually smaller than genuine artworks. Due to monitor size, one would need a pretty large monitor to be able to show a relatively small canvas artwork in its original size, leading to reproductions consistently being presented as smaller. This is the case not only for digital reproductions but can also be observed in analog reproductions such as postcards. However, if the facsimile accommodation is correct, our participants will have tried to accommodate for the fact that they were seeing a reproduction, specifically, participants imagined the artworks larger. This fault in accommodation (the genuine artworks were actually not larger in real life) may have obscured the effects of genuineness. This is possible, as artwork size can influence the art experience (Pelowski et al., 2017).

Finally, the lack of an effect of genuineness may be due to our sample consisting mainly of laypeople. The results of Locher et al. (2001) suggest that only art-trained people rated genuine artworks differently than reproductions. It may be the case that genuineness matters for people with a high knowledge of art but not for laymen. If this is the case, then any effect of genuineness can only be found in an art-trained sample.

Nonetheless, it is worth asking if this potential effect disappears when genuine artworks are compared with high-quality reproductions. It is likely that trained observers pick up on subtle visual cues (such as brushstrokes) that are in genuine artworks but not normally visible in low-quality reproductions due to the medium or limitations in technical aspects (e.g., display resolution). The presence (or lack) of these visual cues may be the underlying cause of the effects of genuineness. If this is the case, progress in technology and the digital display of images will lead to the ability to create reproductions that have the same subtle visual cues as the genuine artwork. This technical progress has the potential to change the way we experience art.

Furthermore, it is possible that effects of genuineness depend on the artistic quality of the artwork. In the case of famous artworks, such as the Mona Lisa, people will have explicit expectations, knowledge-based emotions, or other valuations, which make them differently and especially valued by people when seeing them in real life. As our sample of artworks included neither well-known artworks nor well-known artists, it may be that the effects of genuineness only come into play when people are looking at famous artworks or artworks by famous artists.

Nonetheless, it is crucial to discuss the possibility of the absence of a genuineness effect, in general, which is suggested by the findings of Brieber, Leder, et al. (2015) and this present study. Perhaps previous genuineness effects may have been caused by confounds.

As a new finding and extension of the previous literature, our study showed a strong effect of art style. Art style explained 44% of variance in the model, illustrating the importance of this factor. Interestingly, style did not interact with conditions of context or genuineness. This result indicates that style effects are present regardless of context or genuineness. This suggests that style effects previously found in a laboratory setting using reproductions are generalizable to people engaging with genuine artworks in a museum. This knowledge supports further investigations on style in laboratory designs. An interesting direction for the future would be to assess how important art style is for people with training in art (Leder & Nadal, 2014). Laypeople tend to consistently rate abstract art lower than figurative art on a variety of measures (Locher et al., 2001), which may result in art style being a more important aspect to laypeople than to experts, especially because art style mainly influences to what extent people understand the artworks. Figurative art may be easier to understand for laypeople because the representation of figurative scenes can be easier linked to scenes of everyday life (see “Cognitive Mastering” in Leder et al., 2004). Due to their low level of knowledge about art, the link between art and everyday life may act as an important resource for laypeople to understand art. In contrast, it seems likely that people with art training do not need to rely on figurative content as much, as their higher level of knowledge of art can act as an alternative resource for understanding art.

In sum, our study emphasizes the importance of art style in art experience. In addition, it provides further insight into the (lack of) effects of genuineness of artworks. Finally, it supports previous research by replicating a positive effect of the museum context. Our results show that artworks are liked more and are seen as more interesting in a museum context rather than a laboratory context. These findings show the importance of a museum as institution and physical space that create an optimal environment for the experience of art.

Acknowledgments

The authors thank the MUSA for allowing the use of its exhibition space, genuine artworks, as well as their professional photographs as reproductions. A special thanks goes to the former director of the MUSA Dr. Berthold Ecker and the photographer Heimo Watzlik for their support and highly valuable expert advices. Finally, the authors would like to thank the anonymous reviewers for their constructive comments and suggestions that helped to improve this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

- Benjamin, W. (2010). The work of art in the age of its reproducibility. In M. Jennings (Ed.), *The work of art in the age of its technological reproducibility* (pp. 11–37). Cambridge, MA: Grey Room and Massachusetts Institute of Technology.
- Brieber, D., Leder, H., & Nadal, M. (2015). The experience of art in museums: An attempt to dissociate the role of physical context and genuineness. *Empirical Studies of the Arts*, 33, 95–105.
- 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.
- Brieber, D., Nadal, M., Leder, H., & Rosenberg, R. (2014). Art in time and space: Context modulates the relation between art experience and viewing time. *PLoS One*, 9(6), e99019.
- Cupchik, G. (2002). The evolution of psychical distance as an aesthetic concept. *Culture Psychology*, 8, 155–187.
- Dutton, D. (2009). *The art instinct: Beauty, pleasure, & human evolution*. Oxford, England: Oxford University Press.
- Griswold, W., Mangione, G., & McDonnell, T. (2013). Objects, words, and bodies in space: Bringing materiality into cultural analysis. *Qualitative Sociology*, 36, 343–364.
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95, 489–508.
- 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.
- Locher, P., & Dolese, M. (2004). A comparison of the perceived pictorial and aesthetic qualities of original paintings and their postcard images. *Empirical Studies of the Arts*, 22, 129–142.
- Locher, P., Overbeeke, K., & Wensveen, S. (2010). Aesthetic interaction: A framework. *Design Issues*, 26, 70–79.
- Locher, P., Smith, J., & Smith, L. (2001). The influence of presentation format and viewer training in the visual arts on the perception of pictorial and aesthetic qualities of paintings. *Perception*, 30, 449–465.
- Locher, P., Smith, L., & Smith, J. (1999). Original paintings versus slide and computer reproductions: A comparison of viewer responses. *Empirical Studies of the Arts*, 17, 121–129.

- Mastandrea, S., Bartoil, G., & Bove, G. (2009). Preferences for ancient and modern art museums: Visitor experiences and personality characteristics. *Psychology of Aesthetics, Creativity, and the Arts*, 3, 164–173.
- Mastandrea, S., Bartoli, G., & Bove, G. (2007). Learning through ancient art and experiencing emotions with contemporary art: Comparing visits in two different museums. *Empirical Studies of the Arts*, 25, 173–191.
- Mastandrea, S., Maricchiolo, F., Carrus, G., Giovannelli, I., Giuliani, V., & Berardi, D. (2018). Visits to figurative art museums may lower blood pressure and stress. *Arts & Health*, 1–10. doi:10.1080/17533015.2018.1443953
- Newman, G., & Bloom, P. (2012). Art and authenticity: The importance of originals in judgments of value. *Journal of Experimental Psychology: General*, 141(3), 558. doi:10.1037/a0026035
- Pelowski, M., Forster, M., Tinio, P., Scholl, M., & Leder, H. (2017). Beyond the lab: An examination of key factors influencing interaction with ‘real’ and museum-based art. *Psychology of Aesthetics Creativity and the Arts*, 11, 245.
- Scherer, K. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44, 695–729.
- Smith, J. (2014). *The museum effect: How museums, libraries, and cultural institutions educate and civilize society*. Lanham, MD: Rowman & Littlefield.
- Smith, J., & Wolf, L. (1996). Museum visitor preferences and intentions in constructing aesthetic experience. *Poetics*, 24, 219–238.
- Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H. (2018). The Vienna Art Interest and Art Knowledge Questionnaire (VAIAK): A nified and validated measure of art interest and art knowledge. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. doi:10.1037/aca0000205
- Specker, E., Tinio, P. P. L., & van Elk, M. (2017). Do you see what I see? An investigation of the aesthetic experience in the laboratory and museum. *Psychology of Aesthetics, Creativity, and the Arts*, 11(3), 265–275. doi:10.1037/aca0000107
- Tröndle, M., & Tschacher, W. (2012). The physiology of phenomenology: The effects of artworks. *Empirical Studies of the Arts*, 30, 75–113.

Author Biographies

Susanne Grüner is a PhD student at the Faculty of Psychology at the Technische Universität Braunschweig.

Eva Specker is a PhD student at the Faculty of Historical and Cultural Studies at the University of Vienna.

Helmut Leder is professor of Cognitive Psychology and Head of the Department of Basic Psychological Research and Research Methods at the University of Vienna.

2.3.4

Specker, E., Stamkou, E., Pelowski, M., & Leder, H. (*under review*). Radically revolutionary or pretty flowers? An experimental museum study of the impact of curatorial narrative highlighting artistic deviance on the visitor's assessment of artist influence. *Curator*. Submitted: 05.04.2019

My Contribution:

In this case, the concept and idea underlying the manuscript was developed by me based on Stamkou et al. (2018) and as such on previous work by the second author. I designed the study, and collected and analyzed the data. ES, MP, HL all gave feedback on the design. MP and HL assisted in establishing the contact with the Albertina Museum. I was the main author of the manuscript though the section on Stamkou et al. (2018) was written by ES and MP had a strong contribution to the introduction. All co-authors provided feedback on consecutive versions.

**Radically revolutionary or pretty flowers?
An experimental museum study of the impact of curatorial narrative highlighting
artistic deviance on the visitor's assessment of artist influence**

Eva Specker^{1,3}

Eftychia Stamkou²

Matthew Pelowski¹

Helmut Leder^{1,4}

¹ Department of Basic Psychological Research and Research Methods, Faculty of Psychology,
University of Vienna, Vienna, Austria

² Department of Social Psychology, Faculty of Social and Behavioural Sciences, University
of Amsterdam, Amsterdam, The Netherlands

³ Department of Art History, Faculty of Historical and Cultural Studies, University of Vienna,
Vienna, Austria

⁴ Research Platform Cognitive Science, University of Vienna, Vienna, Austria

Author Note:

Correspondence concerning this article should be addressed to Eva Specker,
Department of Basic Psychological Research and Research Methods, Faculty of Psychology,
University of Vienna, Liebiggasse 5, 1010 Vienna, Austria.

E-mail: eva.specker@univie.ac.at.

Abstract

The importance of curatorial narrative—the overarching theme or meaning of an exhibition that informs all of the micro-decisions made in an exhibition design—is clear in practice but has not been empirically investigated. Especially, how does the simple ordering and grouping of artworks impact the way that the artist is seen? In the current paper, we discuss a systematic empirical investigation of the impact of curatorial narrative on visitor experience meant to test these questions. To do this, we build on recent lab-based research from social psychology that showed that by embedding target artworks with others meant to highlight the artist’s difference (or ‘deviance’) in terms of new styles, led to viewers judging the artist as more influential. We adapt this paradigm to a museum study, reporting findings of a field experiment conducted in the Albertina Museum in Vienna ($N=216$) in which we aimed to translate the lab-based paradigm into the museum context. To do this, we used an artwork placed within a temporary exhibition on Claude Monet that showed his stylistic progression throughout his life as a real-life equivalent of the deviance condition from the lab. This condition was compared to the same artwork placed within a permanent exhibition that showed the artwork amongst other examples of impressionism without suggestions of deviance. We hypothesized that the artist is seen as more influential in the temporary exhibition (deviance) than in the permanent exhibition (no deviance). Our results confirm this hypothesis. These findings provide quantification for how curatorial narratives influence whether and when artists are considered influential, beyond the level of didactic extra-exhibit information, and offers guidelines for evidence-based curatorial practices.

Keywords: Artistic influence, deviant artists, curator narrative, evidence-based curatorial practice, exhibition design.

Curators play an important role in shaping visitor experience in museums. They make decisions about exhibition design such as lighting, wall colors, accompanying texts, but also play a role in deciding what objects are displayed, and—most importantly—make decisions about the ‘curatorial narratives’ of exhibitions themselves. That is, when designing an exhibition, curators must think of an overarching exhibition concept: What is this exhibition about; What do I want people to learn or take away from their experience? These questions inform what items are selected and how they are displayed. Even more, curators might consider how they want the selected items to be ordered within an exhibition, how an individual might move between rooms, between themes, or how items might combine together so that an overall experience and understanding can unfold as individuals move throughout an exhibition space.

Especially this last, temporal aspect of exhibition design poses a number of possibilities and questions for the curator and for curatorial research. Often, as in many exhibitions of art, which might target a specific theme, style, or group of one or more artists, exhibitions may have a linear narrative, which starts in the first room of the exhibition (e.g., an artist’s early work) and ends at the last room of the exhibition (e.g., an artist’s late work). However, it must be noted that some exhibitions allow the visitor to follow their own path (a strategy often seen in science museums, for example). Nonetheless, with emerging ideas of the role of curators, to highlight novel interpretations or reactions to sets of artworks, as well as to create interesting or unexpected juxtapositions or connections between differing works, this curatorial device is becoming more prominent. This also fits important discussions in the more general topic of empirical aesthetics or social psychology, which increasingly suggest the important role of context and conceptual framing as a key modulator of the reaction to and appreciation of art (Bullock & Reber, 2013; Luring et al., 2016; Pelowski, Markey, Forster, Gerger, & Leder, 2017).

However, despite a handful of studies on, for example, the role of curatorial narratives in inclusion-exclusion dynamics (e.g., Weil, 2002) or exploratory assessments of exhibition design (e.g. Pelowski, Liu, Palacios, & Akiba, 2014; Tröndle, Greenwood, Bitterli, & van den Berg, 2014; for a comprehensive overview of visitor studies in general, see Schiele, 2016), this aspect of art perception remains almost completely empirically unexplored. Most importantly, because of their exploratory nature the majority of past curatorial-impact investigations and results represent interesting data for specific museums, however, it remains unclear how the results would generalize to other museums. This calls for a confirmatory approach or a “experimentalist view of museum evaluation” (Shettel, 1968; Schiele, 2016; Screven, 1976) to formulate specific hypotheses, based on previous empirical or theoretical findings, that can then be systematically tested and with generalizable results to provide possible direction or ideas for future curatorial work.

Such an approach is important not only for better understanding the role of curatorial narrative, but also for our general assessment of the efficacy of curatorial and museum practice. As recently put by Schiele (2016), “the fact remains that the lack of objective data on visitors [...] in the context of ... the educational role of museums, [makes it presently] impossible to prove the efficiency of an exhibition” (p. 338). This itself is a re-phrasing of an argument stemming back decades, to for example Parr (1962): “we have no evidence of what effects good design may have upon our ability to communicate knowledge and create understanding by visual means” (p. 36-37).

In the present paper, we report the results of such an experimental comparison targeting the curatorial narrative topic and applied to an actual ongoing museum exhibition with masterpieces of art. This is based on a confirmatory methodological approach that focuses on a common curator device—involving the hanging order of artworks to communicate a historical narrative regarding important changes or developments throughout the career of a major artist (Claude Monet). This study further draws from social psychological theory and

recent lab-based studies (Stamkou et al., 2018) that suggested that by simply embedding target artworks in a narrative highlighting stylistic changes or deviance, when compared to a similar grouping that did not highlight such deviance, the artist was seen as more important and revolutionary. Here, we duplicate this in a major museum (the Albertina in Vienna) using the same works of art within both a temporary exhibit showcasing a major Monet retrospective with an implicit focus on deviance, and a permanent exhibition without deviance. We test whether the simple act of changing the juxtaposition of artworks shown has a measurable impact on appraisals of artists.

The current text will follow the following outline: we will first briefly review the literature on narrative or context effects, as well as the lab-based empirical work by Stamkou et al. (2018). We will then discuss how this can be applied/translated to a museum context as well as how these findings might be relevant for curators, and finally introduce the systematic investigation of the impact of curatorial narrative on visitor experience, concluding with a general discussion of the results and implications for future curator research.

Background—historical narrative and perceived deviance and importance of artists

The theoretical point of departure for this study focuses on the impact from embedding art within a wider narrative structure as might be created by a curator for an exhibition—i.e., through selecting and ordering artworks, dividing pieces into different spaces or groupings, or more didactically through the use of accompanying “extra-exhibit media” (Bitgood, 1992) to communicate an exhibition theme. Certainly, the potential of such activities for impacting art perception is a well-recognized aspect of curatorial work. It might be said to be one of the major jobs for a curator him or herself. This emphasis also occurs in tandem with a similar emerging academic awareness in empirical art studies of the importance of many types of context as a modulator of viewer actions, expectations, and artwork assessments (see Bitgood, 1992; Pelowski, Forster, Tinio, Scholl, & Leder, 2017 for museum-focused reviews).

There are several emerging studies that highlight specific context effects that might directly or indirectly touch aspects involved in designing an exhibition. For example, giving background information before the display of art that a work is by a master (e.g., Rembrandt) as opposed to an apprentice (Huang, Bridge, Kemp, & Parker, 2011) leads people to like the art more. In addition, giving information that the artwork is highly regarded by museum curators (Lauring et al., 2016) also leads people to like the art more. And can even, lead to neurophysiological activation in the reward areas of the brain (see also Noguchi & Murota, 2013). Similar results have been found when informing viewers that a work is relatively more or less expensive (Lauring et al., 2016). Studies have found that leading viewers to assume that an artist had a high degree of hands-on contact with an artwork (Newman & Bloom, 2012) or even that a piece took longer to create (Kruger, Wirtz, Van Boven, & Altermatt, 2004) leads to ratings of the art as more aesthetically and economically valuable. Even providing a label with the name of the artist, in a laboratory context, has been shown to lead to higher hedonic ratings (Gerger & Leder, 2015; Mullennix, Pilot, Steeves, & Burns, 2018; Park, Yun, & Jeong, 2015; Russell & Milne, 1997); though it has also been argued that that labels may not have as much or even detrimental impact in museum exhibitions (Bitgood & Patterson, 1993; Smith, Bousquet, Chang, & Smith, 2006) for arguments. Similar positive effects might also translate to ratings of artists. For example, informing individuals that an artwork is authentic versus a copy (say by a skilled forger or apprentice) leads both to higher perceived art quality and to higher perceived talent of an artist (Wolz & Carbon, 2014).

Research also shows that the simple act of placing art within a gallery or museum has similar impacts—with comparisons displaying the same artworks in a lab versus a museum showing higher ratings in the latter case (Brieber, Nadal, & Leder, 2015; Brieber, Nadal, Leder, & Rosenberg, 2014; Grüner, Specker, & Leder, 2019). This is perhaps related to implicit assumptions regarding importance or even curator opinions about the art. Studies have also found similar effects from hanging objects higher than others within a gallery space,

or hanging large object with smaller others, which may increase focus and perceived importance (Miles, Alt, Gosling, Lewis, & Tout, 1982).

These findings suggest many possible impacts when considering a wider exhibition structure. However, again, with the exception of a few studies on viewing time or movement patterns (Bitgood & Patterson, 1993; Pelowski, 2015; Pelowski, Akiba, & Palacios, 2012; Tröndle et al., 2014) what has not been considered is how a wider specific exhibition structure and the interrelation or ordering of artworks may impact the assessment of artists in general. This aspect too, presumably plays a key role in art perception and the overall experience and reaction of a visitor.

One particular aspect that is often highlighted when composing an overall narrative for exhibiting art, and which may provide an intriguing first entry into the empirical investigation of this topic, regards the historical progression within the work of an artist. The historical position of art has been particularly highlighted as a main feature of art appreciation in theoretical arguments by Bullo and Reber (2013). These authors suggest that historical position, along with other context, is taken into account when forming an interpretation or even when deciding how much effort to give to an artwork. Certainly, historical importance is implicitly touched upon in many of the authenticity or other experimental manipulations reviewed above, and could be equally primed via providing labels or other accompanying information (e.g., wall texts or audio guides). Perhaps even more intriguing, a historical narrative could also be manipulated by the very ordering and juxtapositions of art pieces in an exhibition that trace the trajectory of an artist throughout their career or that highlight certain stylistic developments.

Intriguingly, such a narrative may mirror both what is a fundamental aspect of the history and production of art, and also impact the perceived importance of art and artists. Presenting context information by visual means as in the current study and the Stamkou et al. (2018) thus represents a way of making the art historical context of artworks accessible to art viewers. As

noted by Stamkou et al. (2018), the historical context plays an essential role in what people consider to be revolutionary or deviant—i.e., art or other actions that question or often radically change prevalent conceptions or styles. Impressionist art, for example, may only appear deviant when one compares it to the Salon style of painting that was the mainstream style at the time. This constitutes ‘interpersonal deviance’—deviating from the actions of others around one. Not only the broader historical context plays a role. Artists can also deviate from their own previous styles. Take, for example, Picasso who employed numerous artistic styles ranging from his “Blue Period” to cubism throughout his lifetime. This constitutes ‘intrapersonal deviance’—deviating from what you yourself have done before.

Importantly, both forms of deviance are intimately tied to arguments regarding the essential nature of art making and are a potential mark of successful art making. “To produce unique works of art” argues Becker, (1982, p. 204) “artists must unlearn a little of the conventionally right way of doing things they have learned. Totally conventional pieces bore everyone and bring the artist few rewards. So artists, to be successful in producing art, must violate standards more or less deeply internalized.” Crowther (1993, p. 116), too, notes if art is to “stand out as highly accomplished; then the work must be recognized as not simply repeating what has gone before. It must be seen to be in some sense original, innovatory, or exemplary, and this means that it must be judged with reference to the empirical history of the medium” (see also Bourdieu, 1968). In social psychology as well, studies suggest that deviance leads to the perception of power and influence (Stamkou & Van Kleef, 2014; Van Kleef, Homan, Finkenauer, Gündemir, & Stamkou, 2011). In turn, Stamkou et al. (2018) suggested that artists who are perceived by a viewer to deviate from previous means of expression by adopting distinct styles within their career (intrapersonal deviance) and/or to deviate from their contemporaries’ styles (interpersonal deviance) are perceived as more influential and their art products to be more valuable.

To test these hypotheses, Stamkou et al. (2018) created a laboratory paradigm that also provides an intriguing possibility for a similar museum investigation, and which will serve as the essential basis for the present study. Stamkou et al. (2018) performed several investigations in which they showed viewers a selection of artworks (two to five) either all by the same (intrapersonal) or several different artists (interpersonal deviance). Within the selection of artworks, they asked participants to focus on one target image which was either similar in style to all of the others (control condition) or in some way stylistically different. Importantly, all artworks or artists were suggested as contemporaries—i.e., embedded in the same historical time period or reflecting a historical sequence in the intrapersonal condition. Participants were then asked to rate the artist for perceived influence (famousness, influence on contemporaries, made an important contribution to art) and deviance, as well as to assign a monetary value to the target artwork. Artists were routinely found to be more influential when their work was shown in a narrative that focused on deviance from both their own and others' styles. Participants also assigned more monetary value to the artworks and spent more time looking at them.

Although the findings of Stamkou and colleagues (2018) were observed in a lab setting and focused on testing general principles in social psychology, these can be highly relevant for curators. Curatorial narratives—especially in temporary exhibitions focused on one artist—often include deviance. Such exhibitions classically follow a curatorial narrative that tends to be historical—one starts with the early career of the artist and so follows the artist throughout his or her lifetime. This constitutes a prototypical example of intrapersonal deviance: we see an artist's style change in comparison to their own earlier style. Another common narrative that could have the same effect includes interpersonal deviance. This is often established in exhibition design by first showing the art of that period and then showing the art of an artist that deviates from the mainstream art of that period. If the curatorial goal is to highlight the importance of an artist, the findings of Stamkou and colleagues (2018)

provide a basis to create a curatorial narrative that is particularly effective. Stamkou and colleagues (2018) would predict that these curatorial approaches would lead visitors to see the artist as more influential and thus provide a basis for predicting how curatorial narratives influence visitor experience. If the findings of Stamkou et al. (2018) translate to a museum context this can provide important information for curators. This would also support an evidence-based approach to curation in which curators can use empirical findings to achieve their curatorial goals and shape visitor experience. As well as provide an intriguing measure of an art education approach for framing art experience—simply embedding an artist with similar or differing styles—that goes beyond didactic label or accompanying information and which is also a largely overlooked aspect of exhibition design and curatorial work.

Present Study

To test if the paradigm of Stamkou et al. (2018) would translate to a museum setting we conducted a field experiment in the Albertina Museum in Vienna. We used paintings by Claude Monet, who was the focus of a comprehensive retrospective from the 21st of September 2018 to the 6th of January 2019 in a temporary exhibition. This temporary exhibition followed a curatorial narrative that neatly followed the stereotypical historical model and the specific study of Stamkou et al. (2018) suggested above. It started from Monet's early work using a relatively realistic style in the first rooms of the exhibition, and then following Monet's career, progressed to a more impressionistic approach. Ending with pieces that become increasingly abstract in their focus on specific light effects and dabs of multiple layered colors. According to the curator press release (Albertina, 2018), although notably not necessarily known by the typical visitor, a specific focus of the show was precisely to highlight Monet's progression and also his, rather deviant, effort to “pave the way for abstract expressionism” (p. 4).

To mirror the lab design, we chose as a target artwork *The Water Lily Pond* (1917-1919) from the Batliner collection in the Albertina that was in the second to last room of the

exhibition. This artwork was suitable for many reasons. First, the room in which it was displayed showed—within the curatorial narrative—the point within the exhibition where Monet's work became (increasingly) more abstract, thus highlighting his deviance from his past approaches. One of the goals of the exhibition was in fact to highlight Monet's essential deviance or revolutionary changes which were argued to set up abstract expressionist movements to come, and thus cementing his lasting legacy as forever influential artist.

At the same time, the water lily theme itself, and more broadly the work of Claude Monet, also provides a nice test case because of the popular conception surrounding such art. At this point in popular culture, the water lily paintings especially have become a 'classic' Monet or even a classic example of impressionism, which is often highly liked by lay viewers, most especially because they are considered to be beautiful, decorative, pretty, and, most probably, non-deviant or threatening. The water lily is a top selling art print on multiple online sites, reproduced on cups, umbrellas, and other knickknacks in museum shops across the world (www.allposters.com; www.art.com), and the focus of coloring books or acrylic DIY kits for hobby artists and children. Thus, despite the actual historical deviance of the painting (see also, e.g., Rollins, 2004), when viewed without the historical framing, we expected this might show a quite different response. Relating to this last point, this particular artwork was part of the Albertina's permanent collection (part of a large gift of multiple paintings typically shown together). Though the permanent exhibition is loosely ordered by style, it does not follow a curatorial narrative that emphasizes deviance (neither intrapersonal nor interpersonal), with the specific Monet painting shown alongside several other, somewhat similar impressionist works. This allowed for a relatively controlled study design where we could test responses to the exact same artwork in both a deviant condition (temporary exhibition) as well, via a re-testing at a later date, a control condition (permanent exhibition), the only systematic difference being the curatorial narrative.

Based on the findings of Stamkou et al. (2018), we hypothesized that Claude Monet, and the water lily painting, would be seen as more influential in the temporary exhibition than in the permanent exhibition. To exploratively look at other aspects of visitor experience we also asked participants to evaluate the artwork itself. Though we do not have a theoretical mechanism that would explain why the same artwork would be evaluated differently in the temporary as compared to the permanent exhibition we were interested to see if we could find an influence of curation on this aspect of visitor experience.

Participants

The study employed a final total sample of 216 participants (78 males, 136 females, and 2 people identifying with neither gender), with an average age of 36.2 ($SD = 17.1$) with an age range of 18-80. These individuals were roughly split between the two conditions of: (1) a temporary Monet retrospective exhibition with curatorial narrative highlighting intrapersonal deviance ($N = 104$; 42 males, 61 females, 1 person identifying as “other”; $M_{age} = 31.5$, $SD = 15$; range = 18-80), or (2) a control condition, shown at a later date, with the same artwork embedded in a permanent collection of works by other artists without any specific curatorial narrative focused on deviance ($N = 112$; 36 males, 76 females, 1 person identifying as “other”; $M_{age} = 41.3$, $SD = 17.8$; range = 18-77). All demographics for the whole sample and split up by exhibition are reported in Table 1. The majority of visitors in both groups (74 % combined total) were from Europe. Table 1 shows the cultural regions (based on Anděl, Bičík, & Bláha, 2018) our participants belonged to and Table S1 in the supplementary material shows the same information for nationality.

Participants were selected at random from all visitors who appeared over 18 years of age within the target gallery space, with the only exclusion of individuals participating in guided tours. Participants were approached by the researchers and asked if they wanted to participate. Visitors within tours were not approached because it was assumed that the guide would influence their visitor experience and thus potentially interfere with the curatorial

narrative otherwise present in the exhibition. Participants were rewarded with a postcard of an artwork, from the museum shop, for their completion of all tasks.

Materials—exhibition structure and target artwork

The procedure focused on one target artwork (*The Water Lily Pond* by Claude Monet, 1917-1919, see Figure 1 panel A), embedded within two exhibition set-ups. This painting depicts, in a very impressionistic style, a view into a body of water reflecting a blue sky and green-yellow trees with several patches of blue, green, and pink water lilies on the surface. The exhibition conditions were as follows:

(1) the target exhibition, *Claude Monet: A floating world* (21 September 2018 to 6 January 2019) marked the first Monet retrospective to be held in Austria in 20 years, with 100 Monet artworks on loan from over 40 international museums within the Albertina. The target artwork was shown in the second to last room encountered from the entrance (“*The Late Work at Giverny*”, see Figure 1 panel D), again devoted to highlighting Monet’s progressive switch to highly impressionistic/nearly abstract works. As suggested by the wall text (which, importantly, a viewer did not read before participating, see further below), the curator desired that these works would be seen as “...revolutionary in terms of theme, composition, and painting technique, ...solely guided by his inner sensations” and with “[giving] another fresh turn to his painting, whose radical modernism would only bear fruit decades after the death of this great master...”

(2) The control exhibition, “*Monet to Picasso: The Batliner Collection*,” consisted of the permanent display of roughly one hundred pieces donated by a prominent collector in Austria. The artworks range from impressionism to post-impressionism (Degas, Cézanne, Toulouse-Lautrec, Gauguin), to a room devoted to German Expressionism (e.g., Brücke and Der Blaue Reiter), to Austrian artists (Kokoschka, Egger-Lienz), to Russian avant-garde, to numerous works by Picasso ranging from Cubism to the 1940s. The exhibition is generally divided into separate rooms with the above groupings, however without any

obvious attempts by a curator to suggest particular inter-relation of the art. Rather, it is shown as a collection of various masterworks from one individual's lifetime collecting. The target artwork, as seen in Figure 1 (panel D and B) was exhibited with other impressionist paintings (e.g., by Renoir etc., as well as two other paintings by Monet, *View of Vétheuil*, 1881; *The House among the Roses*, 1925) importantly with a similar style.

It should also be noted that while the Monet temporary exhibition did have advertising and posters throughout Vienna, so too is the permanent collection advertised. In addition, this advertising again specifically highlights Monet with a picture of this artwork on the posters. In addition, the poster reads: "Monet to Picasso: Collection Batliner" (Monet bis Picasso: Sammlung Batliner, in German) thus emphasizing Monet specifically even more (see Figure 1, panel C). Thus, it was expected that the simple aspect of Monet being the focus of an exhibition, which could of course influence visitor ratings, would essentially be equivalent.

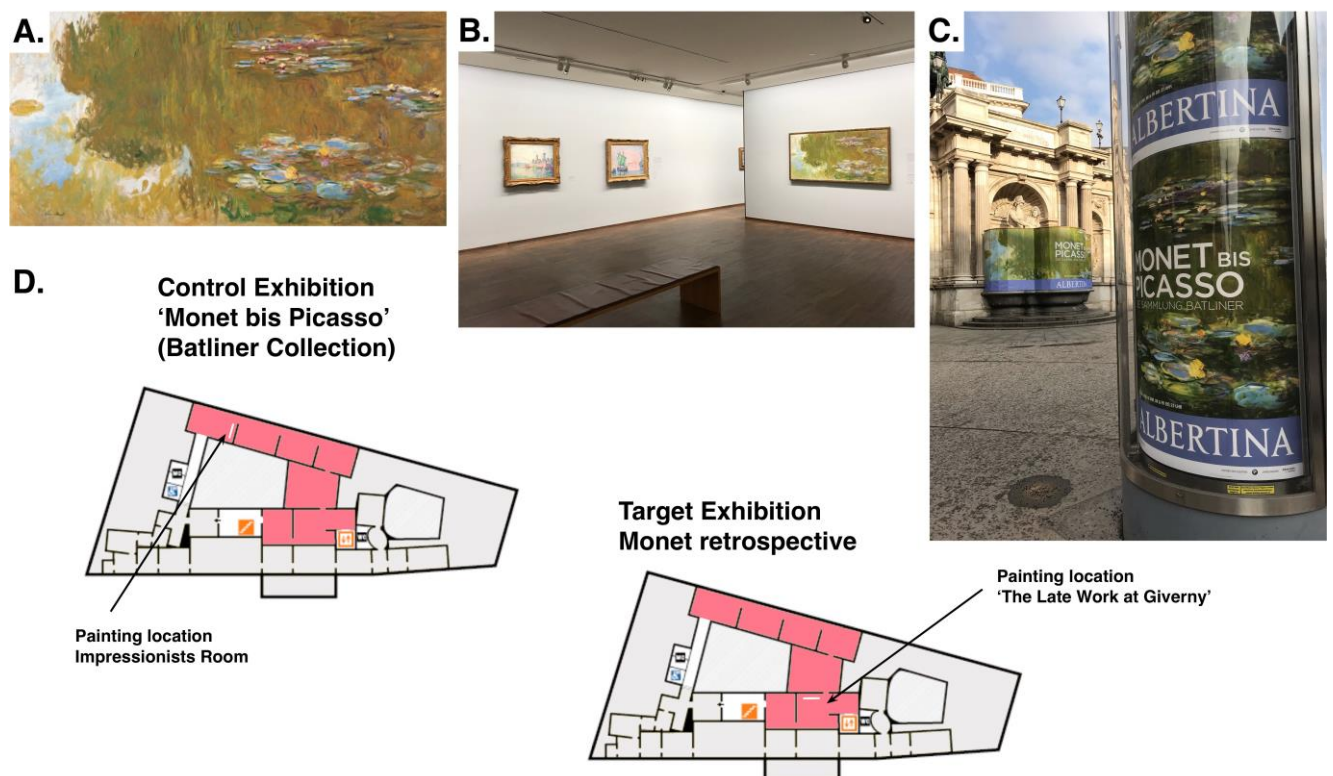


Figure 1. (A) shows the artwork we used, (B) shows the artwork in the Batliner Collection, (C) shows the advertising of the Batliner Collection outside of the Albertina at the time of

data collection for the Control Exhibition, and (D) shows the layout of the rooms in both exhibitions.

Materials—Questionnaire

Perceived general inter-artist deviance: As a manipulation check we asked for perceived intra-personal deviance. In the temporary exhibition, we asked this was phrased as: “How *different* do you perceive this artwork to be from the other artworks made by Monet that hang in the beginning of the exhibition?” This question was accompanied by two images of artworks from the beginning of the exhibition representing Monet’s early representative work (*The La Rue Bavolle at Honfleur*, 1864; *Quai du Louvre*, 1867). In the control condition we asked: “How *different* do you perceive this artwork to be from the other two artworks by Monet that hang in this room?” Again, this question was accompanied by two images of the artworks (*View of Vétheuil*, 1881; *The House among the Roses*, 1925). In both conditions answers were given on a 7-point scale from 1 (not at all) to 7 (very much).

Perceived Influence: As a measure of perceived influence of the artist we asked, “What do you think of Monet as a painter?” followed by four statements: “I think that ... Monet will continue to make a great contribution to art even after many generations of painters”; “... Monet will remain a famous artist in the future”; “... the movement that Monet started will continue to inspire future generations of artists”; and “... Monet’s artistic vision will remain impactful in the future”. Following Stamkou et al. (2018), these questions were provided with the expectation of creating an averaged composite score of perceived influence. All statements were rated on a 7-point scale ranging from 1 (not at all) to 7 (very much). Participants’ responses on these statements were averaged to create a mean score of perceived influence. The overall reliability was $\omega=.85^1$ (split by exhibition: $\omega=.85$ for the temporary exhibition and $\omega=.85$ for the permanent exhibition).

¹ We report omega instead of Cronbach’s alpha due to the numerous deficiencies in Cronbach’s alpha outlined in detail in Dunn, Baguley, & Brunsden (2014).

Art Evaluation. Art evaluation was measured by ratings of the following adjective pairs: ugly-beautiful, passive-active, don't like-like, sad-happy, weak-strong, meaningless-meaningful, calm-lively, boring-interesting, bad-good. They were rated on bipolar 7-point scales from -3 (left adjective) to +3 (right adjective). These scales, loosely based on the Evaluation, Potency, Activity: EPA-model of Osgood et al. (1957), have been routinely used in previous lab and museum art studies (e.g., Pelowski, 2015; Pelowski et al., 2018) and are shown to provide a good overview of individual's general appreciation of a work of art. In addition to the basic hedonic scales (beauty, liking, good, interest), the assessment of especially potency (in this case "strong/weak") was expected to specifically address the aspect of novelty and deviance. As noted by Carroll (1959, p. 113; see also Berlyne, 1974) this factor may indicate "the amount of adjustment that is made or must be made" to a stimulus. Relatively higher assessments with potency factors in conjunction with artworks seen as particularly deviant were previously documented in empirical art study by Pelowski (2015; Pelowski et al., 2012).

Art interest/Art knowledge. To measure the level of art interest and art knowledge we used the Vienna Art Interest and Art Knowledge questionnaire (VAIAK; Specker et al., 2018). This is a new and validated measure that reliably estimates these constructs.

Visit intention/expectation. Since we know that visitor expectations and visitors reasons for visiting can influence visitor experience (Falk, 2009; Pekarik, Doering, & Karns, 1999; Pekarik & Schreiber, 2012), we asked participants if they especially came to see the exhibition they were in when participating with a yes/no answer format. That is, in the temporary exhibition we asked: "For your visit to the Albertina today, did you come especially to see the temporary exhibition on Claude Monet that you are in now?" and in the permanent exhibition we asked: "For your visit to the Albertina today, did you come especially to see the permanent exhibition (Collection Batliner) that you are in now?"

Familiarity. Since we anticipated that knowing the artist beforehand might influence visitor experience of works by that artist we asked participants if they knew Claude Monet beforehand in a yes/no answer format. In both conditions we asked: “Before your visit to the Albertina today, did you know the artist (Claude Monet)?”

Procedure

The procedure and questions followed the laboratory study of Stamkou et al. (2018), adapted for the museum space, and with a very simple between-participants design which, however, allowed for the testing of overall exhibition context on appreciation of a target work of art. All visitors for both conditions were approached when they entered the room containing the target artwork (following the selection criteria above), and without any prior knowledge that they might participate in a study. They were asked if they wanted to participate and, if they said ‘yes,’ were led to the target artwork and asked to make an evaluation using a questionnaire. Notably, by approaching visitors when they entered the room (and because visitors could only enter the room from one side) participants did not have the opportunity to read the wall text referenced earlier which was placed further along in the room, thus reactions were determined only by the visual assessment of the new artworks in light of those previously seen.

The questionnaire could be completed in either German (41.67%) or English (57.87%), and filled out either on a participant’s own phone by scanning a QR-code provided by the researchers (44%) or on a provided tablet computer (56%). Participant’s attention on the target artwork was checked by the researcher before beginning in order to make certain that the evaluations were directed at the correct painting. People first evaluated the artwork itself. This was mainly done to make participants focus on this artwork making our field setting more comparable to the lab-based study design. Afterwards, participants filled out the questions about perceived influence and deviance. People then filled out the VAIK (art interest/art knowledge questionnaire) and finally reported demographics, if they knew the

artist before visiting, if the purpose of their visit was to see the exhibition they were in at the moment, and which exhibitions in the Albertina they visited that day before entering this exhibition.

Results

All data and analyses scripts from the study are available at the Open Science Framework (<https://osf.io/uf6wk/>). Unsurprisingly, most participants (97.7%) knew Claude Monet before their visit and conditions did not differ in their familiarity with Monet, $t(170.68) = 1.30$, $p = .1952$, $g = .17$. Another unsurprising finding was that more people in the temporary exhibition (83%) came specifically to see the exhibition in which they were surveyed than in the permanent exhibition (61%), $t(206.04) = 3.59$, $p = .0004$, $g = .48$. Though it has to be noted that even in the permanent exhibition the majority of participants came to the Albertina to see this one exhibition, which is in line with our previous argument made about the similarity of the advertising for both exhibitions. Importantly, participants in both exhibitions had similar levels of art interest, $t(214) = -.016$, $p = .8686$, $g = .022$. Participants in the temporary exhibition did have a significantly higher level of art knowledge, $t(210.99) = -2.637$, $p = .0089$, $g = .36$. Corresponding means and standard deviations for all descriptive variables can be found in Table 1.

Our manipulation seemed to be successful: participants in the temporary exhibition rated Claude Monet as more deviant ($M = 5.60$, $SD = 1.45$) than participants in the permanent exhibition ($M = 4.89$, $SD = 1.32$), $t(207.95) = 3.72$, $p = .0002$, $g = .51$. Confirming our hypothesis, a one-way ANOVA² showed that participants in the temporary exhibition rated Claude Monet as a more influential artist ($M = 6.31$, $SD = .84$) than participants in the permanent exhibition ($M = 5.93$, $SD = .99$), $t(211.8) = 3.03$, $p = .0027$, $\eta^2 = .041$.

² Normally, one would expect an independent samples t-test here, however, since we wanted to run a follow-up analysis with visitor intentions and art knowledge as covariates we chose an ANOVA to make it easier to compare the two analyses. When running a t-test over the same data the results remain the same: $t(211.8) = 3.03$, $p = .0027$, $g = .41$.

Since our two conditions (permanent and temporary exhibitions) differed in visit intention and art knowledge, we re-ran a series of ANCOVAs, once with visit intentions as a covariate, once with art knowledge as a covariate, and once with both combined. All models slightly changed the results mainly leading to a smaller effect size estimate, but importantly the direction and significance remained the same—visit intention, $F(1,214) = 7.56, p = .0064, \eta^2 = .034$; art knowledge, $F(1,214) = 7.80, p = .0057, \eta^2 = .035$; combined, $F(1,214) = 6.41, p = .0120, \eta^2 = .029$. Since participants did not differ in how familiar they were with the artist we did not run similar analyses with familiarity.

Exploratory Analyses.

We were also interested to see if we could find an influence of curation on how the artwork was evaluated. In addition, one could hypothesize that the nature of the temporary exhibition (an impressive and comprehensive retrospective) may be a rough manipulation that would enhance all aspects of art perception rather than influencing perceived deviance (and consequently influence) of an artist. Testing if people also appraised the specific artwork differently could strengthen the argument for a specific influence of the curation on perceived deviance. As a general test, we ran a MANOVA with all evaluation scales as dependent variables and comparing between the two exhibitions. However, this was not significant: $F(1,214) = 1.54, p = .1369$. It is worth mentioning that the four scales assessing general hedonic responses (ugly-beautiful, don't like-like, boring-interesting, bad-good) showed indications of a potential ceiling effect with the lowest average rating being 1.61 for boring-interesting in the permanent exhibition (1.82 in temporary exhibition) on a scale from -3 to +3. Thus, the fact that the artwork is highly liked overall may have limited the potential effect of a manipulation.

In addition, as expected the curation also influenced the rating of potency, as assessed by the strong-weak scale, $F(1, 214) = 9.01, p = .0030, \eta^2 = .040$, with people in the temporary

exhibition rating the artwork as stronger ($M=1.26$, $SD=1.51$) than people in the permanent exhibition ($M=.65$, $SD = 1.47$).

[Insert Table 1]

Discussion

This study used a rather simple manipulation to assess a key aspect of curatorial work and museum or exhibition design, which, surprisingly, has almost never been empirically quantified in a laboratory much less in a real art museum. We tested the hypothesis that by simply embedding artworks within an exhibition structure meant to provide a curatorial narrative that highlights major stylistic changes throughout an artist's career, compared to a similarly sized exhibition that simply features the work of multiple masters without any clear narrative, the artist would be seen as more impactful in the history of art. To assess this, we used an artist (Claude Monet, exhibited in both a permanent collection of the Albertina in Vienna and in a temporary retrospective) and an artwork (an iconic painting of water lilies) that are undeniably revolutionary and deviant within an art historical context. But, are also seen within popular culture and by many lay visitors as one of the most reliably pleasant, safe, and steadily decorative artists.

The results showed a significant influence of curatorial narrative on how influential and deviant the artist was perceived. As was hypothesized, artists are perceived as more influential when their work is exhibited in a context of (intrapersonal) deviance than when exhibited in a non-deviant context.

The finding have important implications for a number of areas related to art display and curatorial research. First, this study marked one of the first onsite verifications of a long held assumption in curatorial practice—that the decisions of the curator when composing an overall narrative for an exhibition—i.e., the selection of pieces, the division into thematic rooms, or the ordering in which pieces are seen—does in fact lead to measurable differences in the visitor reception of artists. The fact that curatorial narrative can influence visitor's

perceptions of an artist as famous as Claude Monet is noteworthy. It seems more than plausible that the influence of curation would be larger for lesser known artists that visitors may not know before visiting and for whom visitors have no previous knowledge to base their perceptions on. In general, we did not find an effect on how the artwork itself was evaluated—focusing specifically on hedonic or enjoyment/preference scales. However, it has to be noted that this may have been because of a ceiling effect. As noted, the water lilies by Monet are generally highly liked pieces. At the same time, we did find significant increase in potency/strength of the artwork, which again was suggested to be an indication of a felt need, among visitors, to adapt to the art, and further supporting our detected change in artist influence.

This also verified previous intriguing findings, in the lab, of similar effects by Stamkou et al. (2018), suggesting that this may indeed be a rather robust result. Although effects are of course not massive—roughly half of a point on a 7-point scale—when one takes into account that over a million people visit the Albertina Museum in a year, thus approximately 2740 people visit the Albertina on any given day of the year, and that 490.000 people visited during the Monet exhibition, it becomes clear that many people are affected by these curatorial interventions in the Albertina alone. In addition, given how often this particular curatorial narrative is used in exhibitions worldwide this would mean that potentially hundreds of thousands of people are affected each day and that these findings are relevant to all curators.

These findings also provide a basis for an evidence-based approach to curation wherein curators can use empirical findings to achieve or modify curatorial goals and shape visitor experience. The present study suggests that, indeed, if a curator wants to highlight the influence of a particular artist, highlighting the artist's intrapersonal deviance by showing their stylistic progression is a viable strategy. Speculatively, it would also work to highlight

interpersonal deviance by showing how that artist differed from other artists at the time. This would be a fruitful direction of future research.

The findings are also relevant for museum educators who wish to provide a learning experience—say, involving the historical importance and novelty of an artist and specific revolutionary artworks. ‘Learning something new’ or experiencing a new viewpoint or understanding of art or artists is given as a common goal by visitors in studies of motivations for visiting museums (e.g., Chang, 2006; Falk, 2009; Jansen-Verbeke & van Rekom, 1996). This study again quantifies exactly how this may be achieved in the course of a typical artist retrospective.

Even more, the present analysis also suggests how the ability of an overall ambient arrangement design—the actual selection and ordering of artworks—may provide important information to visitors above and beyond any didactic labels or guiding information. Formal knowledge transfer in art museums still occurs often through language-based mediums such as wall texts, guided tours, and audio guides, etc. Due to the nature of the objects on display in art museums, engagement with the objects on display is often hard to achieve through other means. Our findings provide evidence for knowledge transfer—‘this artist/artwork is important,’ ‘this artist did something new and interesting’—by only visual means. The fact that people pick up on these differences without (in depth) knowledge of the corresponding art styles and use these visual cues in their evaluation and experience shows that knowledge is transferred through visual means and shapes visitor experiences. Though it might seem obvious that people can—only by looking—see the difference between, for example, a fauvist and an impressionist work, visitors without the art historical background and concepts may not be able to verbalize this difference. If people can pick up on this difference without having to learn (in depth) about the corresponding styles and can use these visual cues in their evaluation and experience, this would show that knowledge is transferred in this way and shapes visitor experiences. As such, this provides evidence that up-until-now has been lacking

(Schiele, 2016). Our results, along with Stamkou et al. (2018) suggest that, in line with Bulot and Reber (2013), art viewers use this type of information in their aesthetic experiences.

Of course, it should be noted that it is not necessarily our intention to argue that curators should always strive for such impacts or attempt to fit a narrative to art experiences. As noted by Pekarik (2004, p.12) an important question for museum studies is “when does an interpretation [enhance or] diminish the fundamental nature of . . . experience and place limits on independent judgment?” We agree that this is an important question—however, in order to address this, it is equally important for researchers and curators to have better understanding and quantification of possible effects as were found here. Once again, this calls for much future research.

Finally, this study of course is only one step towards a better understanding of curator practice and the relation between museum display and the experience of visitors. Future studies should also test these effects with other artworks, artists, differing backgrounds of visitors or perhaps considering within-participant changes or other measures. However, as a first step toward systematic experimental considerations of art viewing within museums we hope that this study will provide an important basis and framework for much future research.

References

- Albertina. (2018). *Presskit Claude Monet*. Retrieved from https://www.albertina.at/site/assets/files/1811/presskit_claude_monet.pdf
- Anděl, J., Bičík, I., & Bláha, J. D. (2018). Concepts and delimitation of the world's macro-regions. *Miscellanea Geographica*, 22(1), 16–21. <https://doi.org/https://doi.org/10.2478/mgrsd-2018-0001>
- Becker, H. S. (1982). *Art worlds*. University of California Press.
- Berlyne, D. E. (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation*. Washington, DC: Hemisphere.
- Bitgood, S. (1992). The Anatomy of An Exhibit. *Visitor Behavior*, 7(4), 4–15.
- Bitgood, S., & Patterson, D. D. (1993). The Effects of Gallery Changes on Visitor Reading and Object Viewing Time. *Environment and Behavior*, 25(6), 761–781. <https://doi.org/10.1177/0013916593256006>
- Bourdieu, P. (1968). Outline of a Sociological Theory of Art Perception. *International Social Science Journal*, (20), 589–612.
- 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. <https://doi.org/10.1016/j.actpsy.2014.11.004>
- Brieber, D., Nadal, M., Leder, H., & Rosenberg, R. (2014). Art in Time and Space: Context Modulates the Relation between Art Experience and Viewing Time. *PLoS ONE*, 9(6), e99019. <https://doi.org/10.1371/journal.pone.0099019>
- Bullot, N. J., & Reber, R. (2013). The artful mind meets art history: Toward a psycho-historical framework for the science of art appreciation. *Behavioral and Brain Sciences*, 36(02), 123–137. <https://doi.org/10.1017/S0140525X12000489>
- Carroll, J. B. (1959). Review of the measurement of meaning. In J. G. Snider & C. E. Osgood (Eds.), *Semantic differential technique; a sourcebook*. Chicago: Aldine Publishing Company.
- Chang, E. (2006). Interactive Experiences and Contextual Learning in Museums. *Studies in Art Education*, 47(2), 170–186. <https://doi.org/10.1080/00393541.2006.11650492>
- Crowther, P. (1993). *Art and embodiment: From aesthetics to selfconsciousness*. Oxford, UK: Oxford University Press.
- Dunn, T. J., Baguley, T., & Brunnsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Falk, J. H. (2009). *Identity and the museum visitor experience*. Walnut Creek, CA: Left Coast Press.
- Gerger, G., & Leder, H. (2015). Titles change the esthetic appreciations of paintings. *Frontiers in Human Neuroscience*, 9, 464. <https://doi.org/10.3389/fnhum.2015.00464>
- Grüner, S., Specker, E., & Leder, H. (2019). Effects of Context and Genuineness in the Experience of Art. *Empirical Studies of the Arts*, 027623741882289. <https://doi.org/10.1177/0276237418822896>
- Huang, M., Bridge, H., Kemp, M. J., & Parker, A. J. (2011). Human Cortical Activity Evoked by the Assignment of Authenticity when Viewing Works of Art. *Frontiers in Human Neuroscience*, 5, 134. <https://doi.org/10.3389/fnhum.2011.00134>
- Jansen-Verbeke, M., & van Rekom, J. (1996). Scanning museum visitors: Urban tourism marketing. *Annals of Tourism Research*, 23(2), 364–375. [https://doi.org/10.1016/0160-7383\(95\)00076-3](https://doi.org/10.1016/0160-7383(95)00076-3)
- Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of Experimental Social Psychology*, 40(1), 91–98. [https://doi.org/10.1016/S0022-1031\(03\)00065-9](https://doi.org/10.1016/S0022-1031(03)00065-9)
- Lauring, J. O., Pelowski, M., Forster, M., Gondan, M., Ptito, M., & Kupers, R. (2016). Well,

- if they like it... Effects of social groups' ratings and price information on the appreciation of art. *Psychology of Aesthetics, Creativity, and the Arts*, 10(3), 344–359.
<https://doi.org/10.1037/aca0000063>
- Miles, R., Alt, M., Gosling, D., Lewis, B., & Tout, A. (1982). *The design of educational exhibits*. London, UK: Unwin.
- Mullennix, J. W., Pilot, K. M., Steeves, T. A., & Burns, J. C. (2018). The effects of cognitive load on judgments of titled visual art. *Psychology of Aesthetics, Creativity, and the Arts*, 12(2), 166–176. <https://doi.org/10.1037/aca0000128>
- Newman, G. E., & Bloom, P. (2012). Art and authenticity: The importance of originals in judgments of value. *Journal of Experimental Psychology: General*, 141(3), 558–569.
<https://doi.org/10.1037/a0026035>
- Noguchi, Y., & Murota, M. (2013). Temporal dynamics of neural activity in an integration of visual and contextual information in an esthetic preference task. *Neuropsychologia*, 51(6), 1077–1084. <https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2013.03.003>
- Park, S. A., Yun, K., & Jeong, J. (2015). Reappraising Abstract Paintings after Exposure to Background Information. *PLOS ONE*, 10(5), e0124159.
<https://doi.org/10.1371/journal.pone.0124159>
- Parr, A. E. (1962). Some Basic Problems of Visual Education by Means of Exhibits. *Curator: The Museum Journal*, 5(1), 36–44. <https://doi.org/10.1111/j.2151-6952.1962.tb01573.x>
- Pekarik, A. J. (2004). To explain or not to explain. *Curator: The Museum Journal*, 47(1), 12–18. <https://doi.org/10.1111/j.2151-6952.2004.tb00363.x>
- Pekarik, A. J., Doering, Z. D., & Karns, D. A. (1999). Exploring Satisfying Experiences in Museums. *Curator: The Museum Journal*, 42(2), 152–173.
<https://doi.org/10.1111/j.2151-6952.1999.tb01137.x>
- Pekarik, A. J., & Schreiber, J. B. (2012). The Power of Expectation. *Curator: The Museum Journal*, 55(4), 487–496. <https://doi.org/10.1111/j.2151-6952.2012.00171.x>
- Pelowski, M. (2015). Tears and transformation: feeling like crying as an indicator of insightful or “aesthetic” experience with art. *Frontiers in Psychology*, 6, 1006.
<https://doi.org/10.3389/fpsyg.2015.01006>
- Pelowski, M., Akiba, F., & Palacios, V. (2012). Satori, koan and aesthetic experience: Exploring the “realization of emptiness” in Buddhist enlightenment via an empirical study of modern art. *Psyke & Logos*, 33, 236–268.
- Pelowski, M., Forster, M., Tinio, P. P. L., Scholl, M., & Leder, H. (2017). Beyond the lab: An examination of key factors influencing interaction with ‘real’ and museum-based art. *Psychology of Aesthetics, Creativity, and the Arts*, 11(3), 245–264.
<https://doi.org/10.1037/aca0000141>
- Pelowski, M., Liu, T., Palacios, V., & Akiba, F. (2014). When a Body Meets a Body: An Exploration of the Negative Impact of Social Interactions on Museum Experiences of Art. *International Journal of Education & the Arts*, 15(14). Retrieved from <https://eric.ed.gov/?id=EJ1045933>
- 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.
<https://doi.org/10.1016/J.PLREV.2017.02.003>
- Rollins, M. (2004). What Monet Meant: Intention and Attention in Understanding Art. *The Journal of Aesthetics and Art Criticism*, 64(2), 175–188.
- Russell, P. A., & Milne, S. (1997). Meaningfulness and Hedonic Value of Paintings: Effects of Titles. *Empirical Studies of the Arts*, 15(1), 61–73. <https://doi.org/10.2190/EHT3-HWVM-52CB-8QHJ>
- Schiele, B. (2016). Visitor studies: A short history. *Loisir et Société / Society and Leisure*,

- 39(3), 331–356. <https://doi.org/10.1080/07053436.2016.1243834>
- Screven, C. G. (1976). Exhibit Evaluation-A Goal-Referenced Approach. *Curator: The Museum Journal*, 19(4), 271–290. <https://doi.org/10.1111/j.2151-6952.1976.tb00507.x>
- Shettel, H. H. (1968). An Evaluation of Existing Criteria for Judging the Quality of Science Exhibits1. *Curator: The Museum Journal*, 11(2), 137–153. <https://doi.org/10.1111/j.2151-6952.1968.tb01625.x>
- Smith, L. F., Bousquet, S. G., Chang, G., & Smith, J. K. (2006). Effects of Time and Information on Perception of Art. *Empirical Studies of the Arts*, 24(2), 229–242. <https://doi.org/10.2190/DJM0-QBDW-03V7-BLRM>
- Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H. (2018). The Vienna Art Interest and Art Knowledge (VAIAK) Questionnaire: A Unified and Validated Measure of Art Interest and Art Knowledge. *Psychology of Aesthetics, Creativity, and the Arts*. <https://doi.org/10.1037/aca0000205>
- Stamkou, E., & Van Kleef, G. A. (2014). Do we give power to the right people? When and how norm violators rise to the top. In J.-W. van Prooijen & P. A. M. van Lange (Eds.), *Power, Politics, and Paranoia* (pp. 33–52). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139565417.005>
- Stamkou, E., van Kleef, G. A., & Homan, A. C. (2018). The art of influence: When and why deviant artists gain impact. *Journal of Personality and Social Psychology*, 115(2), 276–303. <https://doi.org/10.1037/pspi0000131>
- Tröndle, M., Greenwood, S., Bitterli, K., & van den Berg, K. (2014). The effects of curatorial arrangements. *Museum Management and Curatorship*, 29(2), 140–173. <https://doi.org/10.1080/09647775.2014.888820>
- Van Kleef, G. A., Homan, A. C., Finkenauer, C., Gündemir, S., & Stamkou, E. (2011). Breaking the Rules to Rise to Power. *Social Psychological and Personality Science*, 2(5), 500–507. <https://doi.org/10.1177/1948550611398416>
- Weil, S. (2002). *Making Museums Matter*. Washington, DC: Smithsonian.
- Wolz, S. H., & Carbon, C.-C. (2014). What's Wrong with an Art Fake? Cognitive and Emotional Variables Influenced by Authenticity Status of Artworks. *Leonardo*, 47(5), 467–473. https://doi.org/10.1162/LEON_a_00869

Table 1. Sample Characteristics.

		Exhibition		
		Total	Temporary	Permanent
		(N=216)	(N=104)	(N=112)
Age	Mean	36.2	31.5	41.3
	SD	17.1	15	17.8
	Range	18-80	18-80	18-77
Gender	Female	136 (63%)	61 (59%)	75 (67%)
	Male	78 (36%)	42 (40%)	36 (32%)
	Other	2 (1%)	1 (1%)	1 (1%)
Cultural Region	Australia/Oceania	3 (1%)	1 (1%)	2 (2%)
	Central America & S. America	9 (4%)	3 (3%)	6 (5%)
	East Asia	9 (4%)	0	9 (8%)
	Europe	160 (74%)	89 (86%)	71 (63%)
	North Africa / SW. Asia	7 (3%)	1 (1%)	6 (5%)
	North America	12 (6%)	6 (6%)	6 (5%)
	Russia	13 (6%)	4 (4%)	9 (8%)
	South-Saharan Africa	1 (1%)	0	1 (1%)
	German	90 (42%)	59 (57%)	31 (28%)
	English	126 (58%)	45 (43%)	81 (72%)
Device used	Own Phone	95 (44%)	35 (34%)	60 (54%)
	Tablet	121 (56%)	69 (66%)	52 (46%)
Visit Intention	Yes	155 (72%)	86 (83%)	69 (62%)
	No	61 (28%)	18 (17%)	43 (38%)
Familiarity	Yes	211 (98%)	103 (99%)	108 (96%)
	No	5 (2%)	1 (1%)	4 (4%)
Art interest	Mean	52.5	52.4	52.6
	SD	11.5	12	11.1
Art Knowledge	Mean	8.4	7.5	9.4
	SD	5.3	5.1	5.3
How often do you visit art museums or art galleries on average?	Mean	3.9	4	3.9
	SD	1.4	1.4	1.3

Note: Visit Intentions refers to the question: “For your visit to the Albertina today, did you come especially to see the temporary exhibition on Claude Monet that you are in now?” for the temporary exhibition and “For your visit to the Albertina today, did you come especially to see the permanent exhibition (Collection Batliner) that

you are in now?” for the permanent exhibition. Familiarity refers to the question: “Before your visit to the Albertina today, did you know the artist (Claude Monet)?”

Supplementary Material

Table S1.
Nationality Distribution

Nationality	Total	Exhibition	
		Temporary	Permanent
Algerian	1	1	
American (USA)	10	6	4
Argentinian	1		1
Austrian	40	30	10
Australian	3	1	2
Belgian	1		1
Bolivian	1	1	
Bosnian	2		2
Brazilian	4	1	3
British	5	5	
Canadian	2		2
Chilean	2		2
Chinese	5		5
Croatian	3	3	
Czech	3	1	2
Danish	2	1	1
Dutch	6	6	
Estonian	2		2
French	4		4
German	34	21	13
Greek	1		1
Hungarian	2	2	
Irish	1	1	
Israeli	2		2
Italian	14	5	9
Japanese	3		3
Lithuanian	2	1	1
Luxemburgish	1	1	
Mexican	1	1	
Polish	7	1	6
Portuguese	3	2	1
Romanian	6	2	4
Russian	13	4	9
Swiss	6	2	4
Serbian	1	1	
Slovakian	1	1	
Slovenian	1	1	
South African	1		1
South Korean	1		1

Spanish	7	1	6
Turkish	4		4
Ukrainian	5	1	4
Unknown	2		2

Book Chapters

2.3.5.

Pelowski, M., & **Specker, E.** (in press). The general impact of context on Aesthetic Experience. In M. Nadal, & O. Vartanian (Eds.), *The Oxford Handbook of Empirical Aesthetics*.

Summary:

In this chapter, we discuss the general impact of context on the aesthetic experience. Beginning with the interest in context throughout the history of aesthetics, we build to present empirical approaches and especially theory, focusing on context's main layers and points of influence. We note that from the beginning aesthetics have been considered within a tension between on one hand argues that relevant properties of an object are "directly exhibited and can be registered with the senses at anytime, anywhere, by observer" (Deines, 2013) and on the other hand argues that there is an influence of context by ways of e.g. "the use of taste, perceptiveness, and sensitivity" (Sibley, 1959). As empirical aesthetics slowly emerged as a practice, we find a similar context focus. Fechner (1876) considered them in "aesthetics from above", the early mid century included research on taste, in the 1970s Berlyne also noted its importance, and cross-cultural research started as early as the late 40s (e.g. Abel & Hsu, 1949). In modern empirical aesthetics it is commonly accepted that context is important (also shown by a recent review by Pelowski et al., 2017) as such we then discuss how key context issues might be considered in models of aesthetic processing (e.g. Chatterjee, 2004; Leder et al., 2004; Locher et al., 2010; Pelowski, 2018), with the goal of providing a frame for better approaching context aspects in this book and in one's own future studies. We conclude with some major issues such as how we assess context and how context can act as a confounding variable in designs primarily testing something else, some candidates for future consideration such as illnesses like dementia or Parikson, or the use of hormones, as well as body-perception or crying, and suggestions for further reading and education. Reiterating, for example, the question originally asked by Leder and Nadal (2014): "To what extent does the laboratory distort or attenuate aesthetic episodes?" Where we, as also noted in the other chapter with Pablo Tinio, argue that even though laboratory work is extremely valuable, there are certain aspects of aesthetic experiences that are best examined when art is presented in more natural contexts.

My Contribution:

In this case, MP was responsible for the core idea of the chapter as well as the underlying literature. My contribution was in co-writing the paper. Once MP had an initial draft we went to multiple versions in editing, rewriting, cutting, etc. of the chapter to end up at a final version.

The Oxford Handbook of Empirical Aesthetics

Marcus Nadal, Oshin Vartanain, Editors

The General Impact of Context on Aesthetic Experience

Matthew Pelowski, Eva Specker

University of Vienna, Faculty of Psychology

“Context and approach, antecedent moment and goal, are part of the aesthetic experience...”

—Michael Mitias (1986, p. 77)

“A minute in the world’s life passes! ...To become that minute, to be the sensitive plate. What we make of that ...moment when it is before our eyes depends upon what we expect ... and that in turn depends today upon what we have already experienced...”

—Paul Cezanne (in Berger, 1972)

Context is the ever-present counterpoint to the person and the stimuli discussed in the previous chapters, mediating how these come into relation to create an aesthetic experience. Researchers are becoming increasingly aware that our preferences, our tastes, our emotions and decisions are not stable but are constructed during an interaction and from a complex interplay of factors (Reber, Schwarz, & Winkielman, 2004; Pelowski, Forster, Tinio, Scholl, & Leder, 2017 for reviews). Notable context effects—which will be tackled in this and in the forthcoming chapters—include, but are definitely not limited to: participant mood, labels or contextual information, suggested cost or authenticity, the social environment or knowledge of peer ratings, museums and theaters versus laboratory settings, personality, previous training and interactions; not to mention the outside weather, odors; even doing jumping jacks or walking instead of sitting. All may change what we think is beautiful, important, or how we classify and engage with a stimulus.

This suggests an almost endless supply of potential research directions. Indeed, in composing an introductory chapter on context, it is tempting to write “context is potentially everything, have a nice day!” and leave it at that. However, there is much to be said about the interaction of context. First and foremost, this involves moving beyond certain studies or identified effects to how these might be fit together into present knowledge of the processing and sequences in aesthetic experience. This can also help us to anticipate new context issues and consider how and where they might play a role and be investigated. At the same time, aesthetics is also itself rather unique in its relationship to

context and its analysis. Built from a philosophy which argues against context effects, much aesthetic writing suggests, on the other hand, a highly personal or, transversely, *universal*, context-free basis for our judgments (Leder & Nadal, 2014 for review). To borrow a few famous clichés, either there is ‘no accounting for taste’—and therefore not much to really talk about using typical empirical methods—or our aesthetic reactions are made to qualities inherent in objects and thus “objective,” “unprejudiced” (Hume, 1757), “unchanged by circumstance” (Bell, 1914, p. 37). This attitude can be traced into the present day in the design of highly controlled, reductive studies. Thus, not only how do we approach context, but are there certain features that are immutable? How do we design empirical and consider studies in this domain?

In this chapter, we offer an introduction to this complex topic. This will serve as a gateway to the next several chapters. As many of context’s specific areas—the social, the physical or institutional, background or personality-related features—have been considered or will be tackled later, we will offer a more general consideration. Beginning with the interest in context throughout the history of aesthetics, we build to present empirical approaches and especially theory, focusing on context’s main layers and points of influence. We then discuss how key context issues might be considered in theories or models of aesthetic processing, with the goal of providing a frame for better approaching context aspects in this book and in your own future studies. This is interspersed with what we consider to be some of the more intriguing books and papers—spanning both literature on empirical and psychological aesthetics, as well as the humanities, art criticism, and sociology—in order to hopefully spur your thinking about the potential for studying context. We conclude with major issues, candidates for future consideration, and suggestions for further reading and education.

1. A brief history of context in aesthetics

Context—at least philosophically within the field of aesthetics—is marked as much by its ambivalence as it is by its ever-present acknowledged importance. It is therefore useful to briefly

consider history as this does inform our approach and understanding of context studies. This is by no means an exhaustive review but is designed to give a taste of the prevalent discussion and, one might say *the context*, for present issues in empirical study.

Where did we come from? Empiricist perspectives... but only for the right people

The beginning of aesthetics, again, stems from an essentially empiricist or formalist perspective. This claims that the aesthetically relevant properties of an object—such as its beauty—are innate; “they are directly exhibited and can be registered with the senses at any time, anywhere, by any observer” (Deines, 2013). For example, Cicero (roughly 100-50BC; paraphrased by Summers, 1987), suggests that aesthetic experience begins in “seeing, hearing, and understanding hidden and wonderful things.” These “are seen at once” without modulation by outside factors. Similar arguments can be traced through Plato (e.g., 1987) and Aristotle, and a long line of philosophers suggesting immutable but perhaps intangible qualia that exist at the basis of the world and our sense perceptions (Summers, 1987; Dennet, 1988). Aesthetics, in turn, was essentially being the pursuit of such qualities or their perception and appreciation.

Yet, if one looks through the same history of writings with a contemporary perspective, we do find a continuous if implicit argumentation for context—mixed with a dollop of classicism. To continue the quote of Cicero, formal qualities may be argued to be immutable, but “nature and reason” are required to ““transfer’ ...what we gain from sense from a lower to a higher sphere, ‘from the eyes to the mind’” (Summer, 1987, p. 50). That is, by having the necessary knowledge, life experiences, or interest, or perhaps even by being born with the correct abilities, one could properly tune to and then appreciate the qualities of the world and in a sense get at their underlying importance or “truth”. This is explicit with Plato, whose iconoclastic idea of the danger of images, art, or poetry—expressed via his allegory of the cave (Plato, 1987)—explicitly suggested that one might develop enough knowledge to “see past” the artifice. “The intelligence,” notes Murdoch (1977, p. 2), “moves from uncritical acceptance of sense experience ... to a more sophisticated and morally

enlightened understanding.” In turn, artists were not banned—another famous suggestion—because they made objects that were deficient, but because they were so convincing that they could easily seduce the novice masses—“simple-minded fellow[s]” taken in by a “charlatan” (Plato, 1987, p. 340).

Similar arguments can be traced through the Middle Ages, to the Renaissance (see Summers, 1987), to Descartes (e.g., 1965), and Leibniz (1902). Context also played an implicit role in Baumgarten’s (1750) introduction of aesthetics as a unique field (Osborne, 1979 for review). His argument for aesthetic experience, what Baumgarten called the “perfection of sensate cognition” (1954), although focused again on low-level features or immutable sensual perceptions, implied an element of training or other interpersonal difference. Although formal properties (e.g., shape and color) can be perceived by anyone, “the use of taste, perceptiveness, and sensitivity is necessary” for aesthetic perception or discerning what made the properties special (Sibley, 1959). These were acquired through learning or education. In fact, according to Poppe (1907, p. 47) Baumgarten may have been specifically motivated by a pamphlet from the late 17th century (Pierre Bonhours') that Germans, including Baumgarten, were “incapable” of appreciating art and beauty. One might therefore say that our entire field is based on a key context element—nationalism. (Incidentally, as an American and Dutch author, happily working under a German boss, it should be noted that whether Germans can actually appreciate art has still, to our knowledge, not been considered.)

Early empirical study to 1980—differences in taste, aesthetic sensitivity, art training, culture

As empirical aesthetics slowly emerged as a practice, we find a similar context focus. Fechner (1876) with the introduction of empirical aesthetics as a science, talks of both lower relationships involving sensual materials but also of the formation and development of higher relationships—manifest in top down associations that can create differing meanings or appreciations. These involve the cultivated understanding of differences—“In short, not only width but height,” that is, experience and education (see also Deines, 2013; Harris, 2013 for similar discussion of Hume, etc.).

In the early to mid-20th century, the early assessments on aesthetic sensitivity or ability (Barron & Welsh, 1952; Child, 1962; Iwawaki, Eysenck, & Gotz, 1979), which might be said to mark some of the first artwork rating studies, and which asked individuals to choose a “superior” (more pleasing, more artistic, more satisfying) image, also typically used previous ratings by experts (those trained, or at least those studying art history, art making, etc.) as the point of reference for “correct” judgments. Child (1970) went even further, positing what he called the “t” factor, which was argued to be separate from intelligence or creativity. He explicitly connected to context in crafting his empirical studies. “Some people feel that a scientific approach requires giving equal status to everyone. I would argue on the contrary that people greatly interested and ...knowledgeable in the fine arts are likely to respond to a painting or sculpture in especially illuminating ways” (p. 47). Interestingly, he also raised the possibility that aesthetic ability was not necessarily the product of training, but came about naturally through personality— involving a liking for challenge, novelty, independent judgment, emotional awareness (Child, 1970, p. 48; see also Eysenck, 1972)—which would both drive individuals to obtain more art exposure or to seek out careers in aesthetics.

With the 1970s, and the wave of researchers at the cusp of contemporary empirical aesthetics study, we find an even more overt focus. Parson’s (1989) suggested a cognitive developmental model of art appreciation becoming more nuanced as one had more and more exposure, especially in adult stages. Berlyne (1974) introduced the importance of background in his discussion of aesthetic enjoyment. His four channels of information by which we attend to stimuli, for example, listed a semantic source consisting of the meaning of things in ‘reality,’ an expressive and a cultural aspect requiring awareness of art and social norms, and a syntactic source that gives significance to formal elements. In his empirical studies, such as with assessment of paintings (Berlyne, Ogilvie, & Parham, 1968), he also specifically excluded individuals who did show art interest or recruited art interested individuals. See also Martindale (1988) who explicitly suggested the role of background and expertise in major processing stages.

Moving back through the 20th century, context also became a factor in studies of cultural differences, which might be said to mark some of the first explicit context-focused empirical investigations. (Incidentally, this area is still extremely under-investigated). In an early study, Abel and Hsu (1949) showed that Chinese-born participants perceived ink blots as a whole pattern more frequently than did American-born participants. Chiu (1972) showed that Chinese children tended to group two objects on the basis of the relational-contextual information; American children tended to group objects based on shared analytic features or categories. Berlyne (1976) found that judgements of similarity for art works (colored post-card reproductions of Western paintings) differed between Indian villagers and Canadian participants. The villagers also showed higher preference for brightness. Osgood and colleagues (e.g., Tanaka, Oyama, & Osgood, 1963) considered cultural influence on topics such as color perception and assessment of art. He reported more universal consistent way of making evaluations—although not necessarily similar evaluations themselves—suggesting that our ratings might divide into hedonic/evaluative, potency, and activity clusters. He also reported a shared preference for blue (Adams & Osgood, 1973). See Komar and Melamid (1997) for a particularly compelling modern parallel in cross-country preferences in painting.

Other notable studies include McArthur and Berry (1987), who considered facial attractiveness and report a shared response to baby-faces in Koreans and Westerners, or the ecological approach to vision by Gibson (1955), which dispelled notions that the eye could be reduced to a camera and highlighted context aspects of mobility or expectations to guide perception.

1990—expertise, labels, museums

This path continued into the burgeoning of psychological aesthetics in the 1990s. Early context-related studies looked at the importance of art expertise as modulating expectations and visual artwork ratings (Winston & Cupchik, 1992; Cupchik, 1995; Hekkert & Van Wieringen, 1996; see also Getzels & Csikszentmihalyi, 1969 for a very early study). Another main early empirical

focus considered the impact of accompanying labels or titles, which were shown to increase an artwork's felt importance, encourage meaning making (Cupchik, Shereck, & Spiegel, 1994), and can also boost a work's ratings (Russell & Milne, 1997). Branching afield to general museum studies, art education, and sociology, researchers assessed the role of museum or gallery spaces (discussed further in a forthcoming chapter and below). Notably, Bitgood (1992; Bitgood & Patterson, 1993; see also Falk, 1993) offered a number of studies looking at display conditions, noting that, for example different lighting, placing an object on a higher plane as compared to other objects, changing wall colors, etc., could increase the object's importance and viewer attention or even enjoyment. Studies also have considered "museum fatigue" (e.g., Melton, 1972) or the effect from repeated art viewing. Moving back a bit, in what must be one of the earliest context studies, Robinson (1928) showed that, upon showing participants a succession of art prints and asking them to look as long as they wished, they showed a progressive decrease in attention.

This brings us, essentially, to the present day. Building on past history, there is an awareness that context is important, and especially that it may manifest in aspects of training, experience, and even personality. There is also an awareness that certain modes of perception, setting, or culture may be important. Of course, some researchers did and perhaps still do also argue against the importance of especially background in appreciation of art. For example, the gestalt psychologist Arnheim (see Funch, 1997) considered the perception of visual forces to be direct and without the influence of knowledge (see also Ramachandran & Hirstein, 1999 for a few such arguments regarding the brain).

How is context assessed in empirical studies?

It is also useful to give a brief note about how context is empirically studied. There are several main methods that have emerged throughout the above review. These include the use of two or more conditions, varied between- or within-participants, and whereby one or more factors are changed (say abstract verses representational art; five versus ten second presentations; museum versus lab setting;

Japanese versus Western viewers). Another way is to simply manipulate the prompting or supporting information that is given to participants. This is often done with cues or labels, for example, showing art in tandem with elaborative or descriptive titles (as reviewed above). Similar studies may involve social or setting information (telling participants that “this came from a museum” versus “this is computer generated” (Kirk et al., 2009). Researchers can also manipulate presentation. Gartus and Leder (2014), for example, showed art images with either a street background or with white walls of a typical museum. Causative studies (e.g., tDCS, TMS, pharmacology) might use repeated designs with a sham versus experimental condition, or single or double-blind procedures so as to eliminate other, unwanted, aspects of context and priming.

2. So, context is important and has always been—how is this presently considered?

Context is (probably most often) important; but, this also raises the important question of how can this be more carefully considered? Where do these different aspects fit and how can they be understood within present understanding of aesthetic experience? It is this question—asking us to move beyond an intuited idea of certain factors’ modulation, or an empirical study’s identification of certain effects, to a more cohesive approach of theoretical and empirical studies that, we would argue, marks the next step for context study, and the primary topic for today.

This can be answered by walking through several emerging discussions regarding context importance, models of aesthetic processing, and notable aspects, all of which we will consider below.

What are the general categories for Context studies?

To begin a discussion of present approaches, we might first consider the present arguments for main context areas. A review of literature, does suggest that context arguments have been grouped into a few main clusters. Many of these are noted above. Note also that comprehensive discussions of general context layers, or early review papers, are also mostly found outside the explicit area of empirical aesthetics. Rather, they begin in general considerations of museums, sociology, or other

reviews of human interactions with the environment, and which have gradually been incorporated into explicit discussion of media experience.

Among notable examples, Falk and Dierking (1992), in their review for general interactions in museums, suggested three overlapping sectors: once again, following the above past studies, first an individual's interests and motivations; as well as social context; and the physical context/space. They argue that it is primarily the interplay of these factors that can shape our enjoyment, behavior, and understanding. Bitgood (1992) in a similar review for museums and public spaces, suggested both the setting/architecture, object, display configuration, social/cultural factors, and “extra-exhibit” aspects—including labels, catalog copy, maps, audio guides, text information (see also Griswold et al., 2013). Moving into the focus on visual aesthetics, Gartus and Leder (2014) note that current psychological models divide factors “roughly into three groups”: basic visual characteristics of the processed art; characteristics of the viewer; and other surrounding social or physical aspects of the environment. Recently, we (Pelowski et al., 2017a) reviewed these and other studies, highlighting five areas (shown in Figure 1): (1) once again the *Viewer context*—personality, demographics, etc. (2) the *Object context*, including the formal or physical features of a stimulus, whether art, music, or other media; (3) *Presentation context*, physical aspects of the setting, other stimuli or objects; (4) *Social context* or interactions between individuals; (5) the *Cultural context*, including learned attitudes towards the stimulus and aspects of the interaction, as well as, for example, stimulus history.

Looking to Figure 1, with just these general layers one can of course think of multiple questions or potential effects of context. Whether it be a work of music or art, a football match or a ballet, experienced in a theater or gallery or alone in our homes, in person or on TV, our interactions are always embedded within such a rich, multi-layered experience. Where and how is a stimulus presented? Are we alone or with other people? What is the size, color, texture, or resolution if on a screen? What are your mood and your expectations?

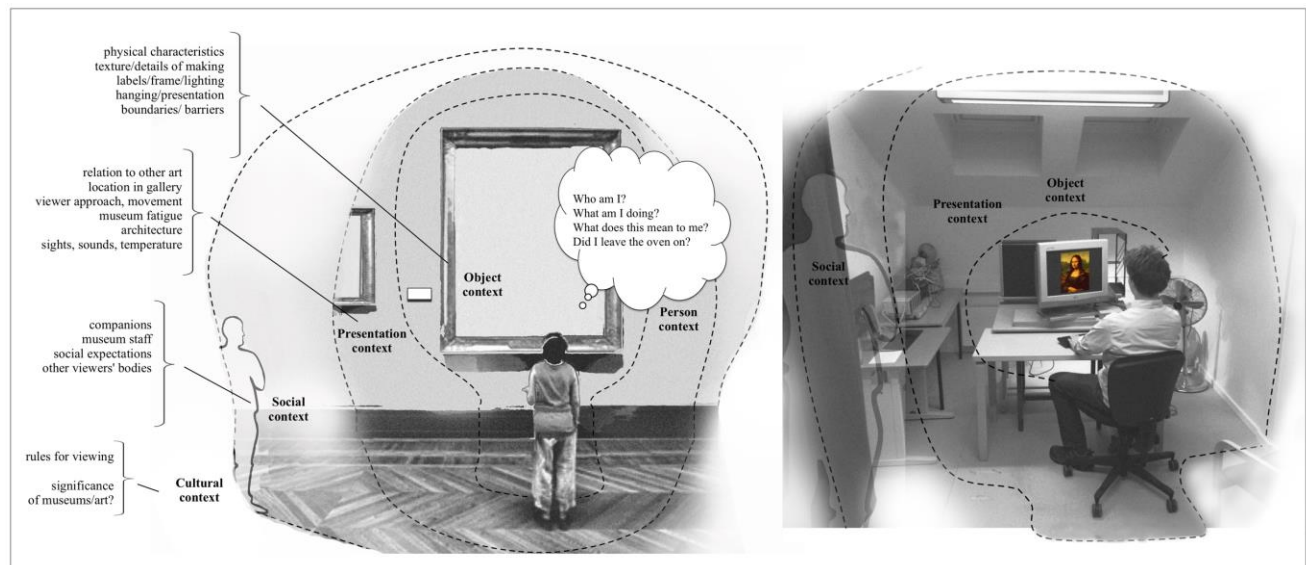


Figure 1. An example of main layers of context in aesthetic experience

Moving beyond a collection of effects—Embedding context *within* aesthetic experience

To better consider the actual psychological or processing implications, the above layers can be combined with present models of aesthetic experience. Models were reviewed in an earlier chapter (see also Pelowski et al., 2016), therefore we will not delve deeply into their nuance here. Suffice it to say, for the purpose of beginning to think about the multiple impacts of context, models are useful because they give at least a theoretical frame for the multiple points and processes that might be modulated. A few examples of main models can be seen in Figure 2, with their explicit mention of context layers circled so that the reader can have a taste of what we mean.

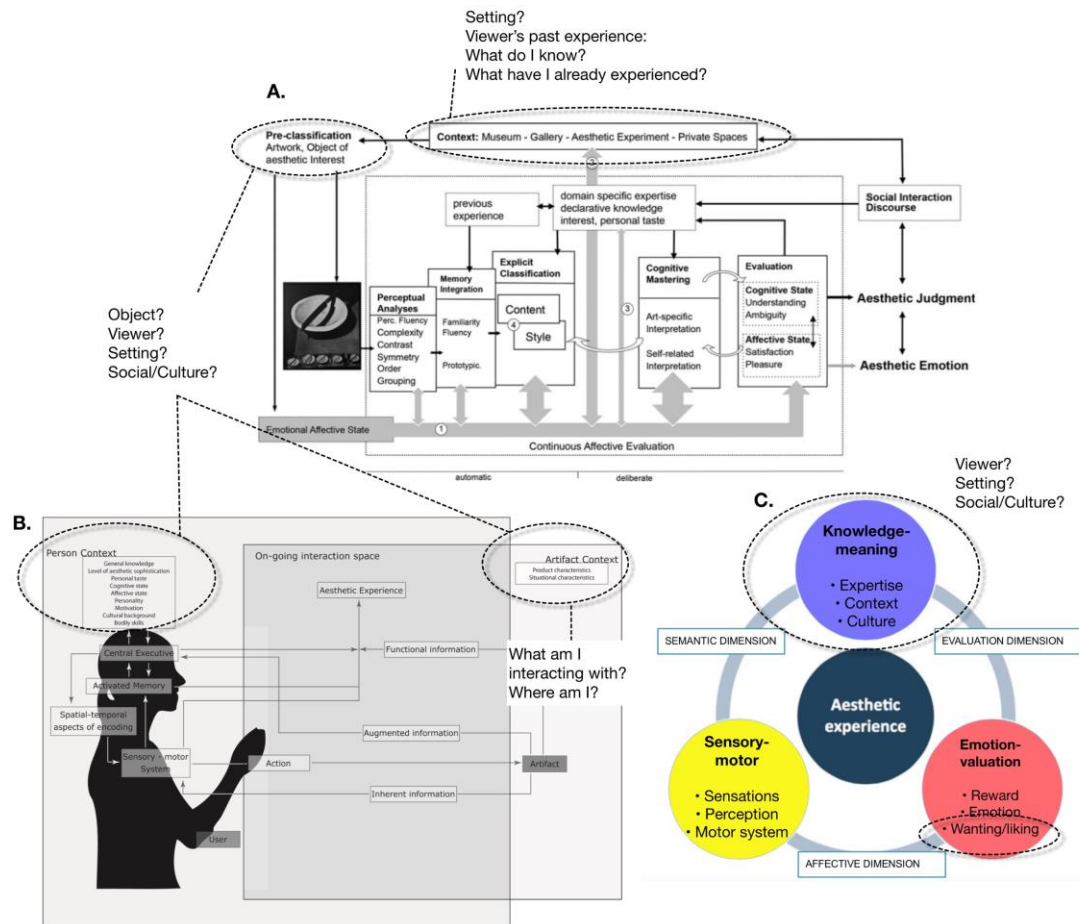


Figure 2. Examples of context layers and key aspects in models of aesthetic experience. **A.** The model of aesthetic appreciation and aesthetic judgment of Leder et al. (2004; updated by Leder & Nadal, 2014), explicitly integrates an *object context* in a stage of preclassification, as well as *setting context* (“museum, gallery, aesthetic experiment, private spaces”), and a *social context* (in this case framed as person-person or person-artists interaction). The authors also add key features of *person context*, domain specific expertise or training, especially in mid-stages. **B.** More geared to design and general object usage, Locher, Overbeeke, and Wensveen (2010, p. 77) also note “two driving forces,” the “artifact itself”—modulated by aspects such as quality impression, assumed function, novelty, etc.—and a “person context” (p. 72), which might be modulated by previous cultural, social, and personal expectations. Similarly, use of an object is argued to activate featural and semantic information in the user’s knowledge base, influenced by the person context including level of aesthetic sophistication, experience, personal tastes, cultural background, personality, and emotional and cognitive state during the aesthetic experience, which influence how viewers “perceive,” and “evaluate” (p. 73). **C.** Similar considerations can also be found in models of neural processes and key correlates of aesthetic processing in the brain. Chatterjee and Vartanian’s (2014) recent ‘aesthetic triad,’ highlights the role of several features. They suggest that aesthetic experiences arise from the interaction among sensory-motor, emotion-valuation and meaning-knowledge neural systems, with especially the latter tied to context, but where all may be modulated by person, stimulus, and environment.

For the purpose of a more detailed review, we have created a hybrid model, shown in Figure 3. This is based on the work of Leder et al. (2004), Chatterjee (2004), Locher et al. (2010) etc. (see also Pelowski et al., 2017 for a similar model to the one used in this paper in the context of museum context studies). Thus, it is especially geared, in early stages, to visual stimuli and to the processing of art. However, when considering its general progression and inputs/outputs, it can apply to other

non-visual or everyday stimuli. It also explicitly lays out outputs and inputs at each stage of experience, again based on a review of empirical literature and present models. Obviously, this discussion cannot hope to be exhaustive, nor do we suggest that this is necessarily the best or only way to explain specific aspects. Rather, treat this as a thought experiment, with suggestions that will hopefully be teased out further in later chapters or within your own future research!

This model proposes six stages, beginning with an initial pre-classification or set of conditions and expectations, moving to an initial processing stage of Perceptual analysis, where we first become aware of and attend to the low-level features of a stimulus (e.g., shape, contrast, lines colors, etc.). This is followed by a stage of Memory integration, during which a features might be grouped into more complex aspects based on previous experiences, expertise, and the particular schema held by the viewer, leading then to an Explicit classification, which involves attention to conceptual or formal aspects (style/content) and where we also may become aware of higher level supporting features such as history, artist, and even begin to identify evoked emotions. These aspects then feed into a stage of Cognitive mastering, in which individuals combine elements and create meaning, and culminating in the outputs of understanding, aesthetic judgment, and affective response.

Depending on the outcome of an initial cycle through such processing, which itself may be accomplished in a matter of seconds (see Pelowski et al., 2017b), individuals may stop, disengaging or moving onto a new stimulus, or may again cycle through various stages, as might occur if individual factors draw specific interest, or if one finds there interactions particularly discrepant. Such an outcome may lead one to feel negative emotion and attempt to disengage or escape, often leading to negative emotion or appraisals, or may also engender more overt top-down, viewer-centered outputs in which one may undergo periods of reflection, adjustment of schema, and even highly pleasurable or transformative experience (e.g., Pelowski & Akiba, 2011). As can be seen in the top part of Figure 3, at each stage we can talk of impact from most if not all context layers. Below we discuss a few main aspects, walking in general through each stage.

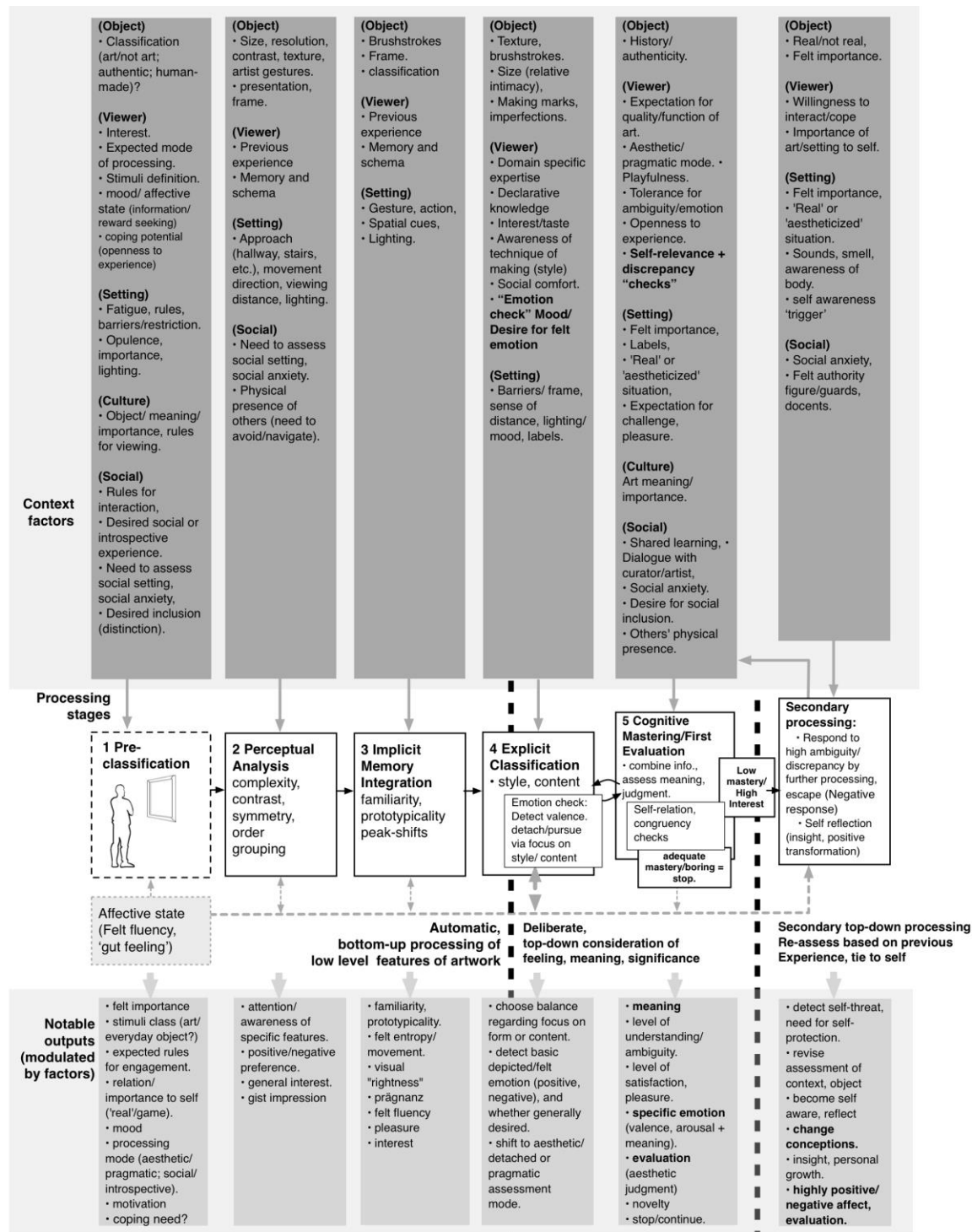


Figure 3. General model of aesthetic processing experience (main focus on visual stimuli), with posited context inputs and implications (outputs) for each stage. (Model adapted from Pelowski et al's. (2017) review of museum context factors and based originally on visual processing model by Leder et al., 2004).

Pre-state—What is it? who am I? What should happen? How do I respond?

First, in the pre-state, before even meeting a stimulus, context (noted as inputs) may play a major role in shaping our forthcoming engagement. This state might be thought of as defining the conditions whereby we engage with a stimulus (noted as outputs at the bottom of the model). Aspects may include our level of interest, our expectations, the expected stimuli type, which can direct our behavior, drive what we look for or our attention, how we respond, or influence the nature of processing itself. The results may then themselves feed into the next stages.

Viewer, Object, Setting Context—Motivations and expectations

One of the most basic aspects here pertains to our expectations for what we will be doing or experiencing. If we expect a stimulus to be important, novel, boring, mundane, or challenging we might both look for differing features or respond differently. For example, Hendon et al. (1989) found that galleries containing art that participants were most interested in before a visit were rated as most interesting at the end of their visit as well. These can be modulated by, say, the *Setting*, *Person*, or the *Object* (as in the box above the pre-state in Figure 3). Visiting a museum or concert hall versus a laboratory, or engaging within one's home, might engender very differing expectations, as would the level of opulence, cultural importance, or even size of an object.

The very decision to visit aesthetic media may set the basis for the entire experience. For example, a good deal of studies have explored visitors' motivations and expectations for going to museums or other cultural events (Tinio et al., 2014; Pelowski et al., 2017). For example (see more in the *Museum Context* chapter), Falk (2009) identified five motivations: intention to explore, have new experiences, recharge/ rejuvenate, facilitate art experience for friends or children, and see specific artworks or exhibits. Pekarik and Schreiber (2012) suggested that visitors may expect to learn new information, make an emotional connection, find beauty, reflect, feel specific reactions such as awe, imagine exotic events or locations, or have interactions with rare, real objects.

The above expectations may provide a particular cognitive frame of reference through which information is processed (see e.g., Darley et al., 1988). Individuals who expect learning or

exploration may be more likely to deeply engage, cognitively process, and seek out and respond positively to challenge or confusion. Those seeking pleasure may respond more positively to positively-valenced stimuli, seek out emotion, or consider only style and visual aspects (see the discussion of emotion below). The level of enjoyment that individuals report may also directly tie to whether or not their specific expectations are met (e.g., Tinio et al., 2014). Aspects may certainly be modulated by *Culture* or *Social* context as well: think of expectations for what reactions *should* occur, or rules for engaging—be quiet, heed other people, do or do not show emotion (see Figure 4).

Personality and background

Of course, multiple studies have also looked at personality determinants of aesthetic processing (Hekkert & Van Wieringen, 1996; Myszkowski et al., 2014 for reviews). Particularly relevant might be the Big Five factor of “openness to experience” (McCrae, 2007). This involves a willingness to seek out novel encounters and has shown correlation with deeper appreciation for and engagement with the arts (Myszkowski et al., 2014; Fayn et al., 2015), higher general art preference (Chamorro-Premuzic et al., 2009). It also significantly predicts physiological reactions when viewing art (chills, feeling touched, absorption, Silvia & Nusbaum, 2011). Similar results may be found for “sensation seeking” (Swami, Stieger, Pietschnig, & Voracek, 2010). Conversely, the need for cognitive closure—suggesting a disliking of ambiguity or desire for clear classifications—might show negative correlation with the above aspects (Roets & Van Hiel, 2011). Another difference may involve, say, personal independence (Markus & Kitayama, 1991), which may interact with social setting to increase anxiety or context impact. Once again, just about any personality construct might be a worthy candidate!

Differences may also be found in demographics. In general, as noted by Hanquinet (2013), in an assessment of engagement with museums or other types of aesthetic media, visitors fit a general pattern of likely white-collar professionals, Euro-American with generally higher levels of income and education, which may be the best predictor of attendance (Chang, 2006). These can be further

broken into sub-patterns (Hanquinet, 2013, p. 809)—for example, “omnivorous,” educated, younger visitors who frequently visit and seek both high- and low-brow culture; individuals with classical, conservative art tastes, who do or do not seek out other cultural experiences. The demographic profile of a typical visitor is presumably quite different from the typical laboratory participant, who is most probably an undergraduate student, while the profiles again may help to direct our engagement. As one example, Hanquinet notes that younger or omnivore viewers may be more likely to be active rather than passive cultural consumers, actively questioning or being critical of aesthetic media (see also Bitgood, 1992).

Object classification

Another important aspect in regards to expectations involves the classification of what an individual will experience. For example, a main topic might be classification of something as an “aesthetic event” or a work of “art”. In most cultures, art often demands a certain reverence and implicit ties to luxury, beauty, or importance (Hagtvedt & Patrick, 2008). Becker (1982, p. 133) notes that many may “believe art is better, more beautiful, and more expressive” than other objects. Art-viewing, when contrasted to viewing other objects thought not to be art, has been empirically related to higher ratings of beauty, pleasure, and liking (Locher et al., 2001). Furthermore, telling someone—via contextual information— that an object is “art” can lead to a higher positive aesthetic rating than when the exact same image is presented with other priming (Arai & Kawabata, 2016).

Returning to the model and the pre-classification stage, this may tie to certain expectations of reward or engagement—an argument supported both behaviorally where an art context may lead to expectations for having pleasurable and positive emotional experience (Leder et al., 2004), and by recent imaging findings that show that viewing objects expected to be art correlates to higher activation of reward and vision areas in the brain (Kirk et al., 2009; Lacey et al., 2011; Kühn and Gallinat, 2012).

An art context can also change processing on a cognitive level, which would manifest throughout the subsequent stages. Perceivers attend more to stylistic and formal properties of stimuli (Cupchik, Vartanian, Crawley, & Mikulis, 2009; Jacobsen et al., 2006; Kirk et al., 2009) and show a more elaborative processing style which goes beyond simple object recognition (Nadal, Munar, Capo, Rossello, & Cela-Conde, 2008). Art or aesthetic contexts may also involve a “transfiguration” or suffusing of ordinary objects and events with deeper meaning (Danto, 1974). This might cause viewers to look harder for significance, beyond an initial visual impression, or give more attention and veneration (Bailey, 2000). This may also allow them to enjoy even ambiguous, challenging, or negative images (for a laboratory study example see Gerger et al., 2014, Wagner, Menninghaus, Hanich, & Jacobsen, 2014). Arthood can encourage a search for intention or prime an expectation of communication. Boas (1943, p. 116) notes, for many viewers, “the fundamental distinction” of art/not-art is between controlled and random behavior. This “faith in the artist” (Parsons, 1987, p. 74) may motivate individuals to “persist in looking . . . when otherwise [they] would be tempted to pass onto something more meaningful” (see Zeglin Brand, 2000 for this aspect and *Cultural* context).

Artistic expression, often from a real human, may also be important in prompting empathic responses. The anthropologist Gell (1997) believing that a stimuli is in some way a product of or signal from another may be key for engaging our artistic interest or for looking for meaning. An interesting study by Umiltà et al. (2012), which showed individuals photographs of artworks by Lucio Fontana, composed of slashes cut into a canvas, as well as black and white line representations of the same slashes, showed higher Mu suppression over bilateral premotor and motor regions in the original case. The authors suggest this showed so-called mirroring of the artist actions. The authors also report that the participants viewed the photographs as more “real art”.

The determination of object classification itself could then be modulated by context. Often empirical studies may explicitly provide supporting information as in a label or cue that a stimulus is art, or may do so more implicitly through an instruction, say, that a participant “will be shown some

works of art” or explicitly asked to “rate how much you like the artwork.” Arthood may also be evoked via setting, or modulated by interpersonal differences. The individual probably carries an understanding of what constitutes art—to them, personally (see for example Pelowski, Gerger, Chetouani, Markey, & Leder, 2017 for a recent study of art/not art classifications and tie to both liking and interpersonal differences).

Authenticity; gallery/theater versus lab

A related issue involves authenticity. Much as individuals may respond in certain ways to art, seeing objects as ‘real’—one-of-a-kind objects from artists with a historical provenance—may also feed into perceived importance or aesthetic merit. Kirk et al. (2009) found that paintings (shown in digital reproductions), which were thought to be created by esteemed artists and borrowed from museums, as opposed to images made by a researcher, were evaluated as more appealing. Similarly, Wolz and Carbon (2014) found that paintings presented as veridical artworks as opposed to forgeries received significantly lowered estimations of quality. An fMRI study by Huang et al. (2011) found that images of Rembrandt portraits, when viewers were told that they were going to be originals vs. forgeries, resulted in higher activation of orbitofrontal areas connected to reward (see also Noguchi and Murota, 2013).

This may also be found in comparison between “real” (i.e., corporeal) art and reproductions in the lab. As will likely be explored more in the forthcoming chapter on museums, researchers often find that museum versus lab art presentations, while not changing assessment of pictorial content or composition, may lead to higher felt interest, pleasantness, surprise or novelty (Locher et al., 1999; Locher and Dolese, 2004), to longer viewing durations (Brieber et al., 2014), and to better recall for what one had seen (Brieber et al., 2015). It is also interesting to consider the discussion of a “transferability thesis” (Currie, 1985), which questions whether viewers are willing to treat an underlying digital reproduction as an artwork. Several studies, which simplified artworks or

contrasted physical objects against images, have found that these are considered less “real” by viewers (e.g., Lacey et al., 2011; Umiltà et al., 2012).

Processing mode—Aesthetic or pragmatic approach

A related aspect to framing also involves the processing mode. It has often been argued that individuals may adopt two general approaches (e.g., Cupchik, 2013; Gerger, Leder, & Kremer, 2014): (1) a “pragmatic” mode where we tend to perceive and respond to a stimulus based on its utility, meaning, and importance or relation to the self, and often argued to align to our typical or ‘everyday’ way of engaging the environment—“can I use it? What does it mean? will it hurt or help me?”; and (2) a more “aesthetic” focus, wherein we may pay less attention to meaning or use and instead focus on an objects sensory or emotive aspects. This latter approach is also often suggested to coincide with a certain “psychological distance” (e.g., Bullough, 1912) or detached perspective in which one can process without consideration of personal threat or relevance, or even to approach engaging in fiction or play without ‘real life’ consequence.

This may modulate all the following stages, impacting what features individuals attend to—style or formal aspects versus meaning or content (Jacobsen et al., 2006; Cupchik et al., 2009). Individuals may be less goal oriented and may be more tolerant for surprise, ambiguity (e.g., Wagner et al., 2014), and especially negative emotions or content. Psychological distance from emotionally negative events reduces signs of arousal and increases relative pleasantness (Lazarus & Alfert, 1964; Speisman et al., 1964). Such a mode may coincide with classification of objects as art (Dissanayake, 2007; Tan, 2000; Cupchik, 2002), engendered by explicit contextual information or, for example, with entering a museum or other media space.

General importance of framing and pre-expectations

Of course, the above discussion does not have to be specific to art or any other media. Rather classifications and contextual framing might be thought of as one way of assigning a particular level of importance. This can of course be done with any object, and in myriad ways—hanging something

high on a wall, placing on a pedestal, using expensive looking materials (e.g., Locher, Overbeeke, & Wensveen, 2010 for discussion with design). Think also of the importance we might give to the football pitch—a field of grass with some white chalk lines marked upon it. Consider also how someone classifies ‘music’ and how they might respond differently to the dissonant compositions of Charles Ives or 4’33” by John Cage (e.g., see Becker, 1982 for an extended discussion).

The social setting can also do the same. Luring et al. (2016) showed that telling participants that a painting they are about to see was expensive or liked by an expert or peer group led to higher ratings of the same art. Interestingly, when telling individuals that art was disliked by an undesirable group, participants again gave the art higher ratings (see also Pinney & Thomas, 2001 for philosophical discussion of social context and framing). Similarly, following the work of Bourdieu (1984), we might talk of “distinction,” or seeking out media as a means of joining a desired group or distancing oneself from others, which may also dictate how we behave or look for. Returning to the model, note that while these factors may play a role especially in the first stage, they can also interject at points later in the experience—in relation to how we respond to emotion or certain content (e.g., in Cognitive Mastery); in whether we keep interacting or disengage if faced with a negative stimulus (Secondary Processing); see further discussion below.

Perceptual Analysis/Implicit Memory Integration—low level features and framing

When we then engage and begin to process, we encounter several more aspects of context. As suggested by the boxes regarding outputs, we might first process low-level features, group them together and generate a gist impression (involving basic identification or negative/positive valence). Interestingly, occasionally it might be thought that these early features are less impacted by context—indeed, it is these aspects that might align with historical discussion of immutable, context-free formal aspects of a stimulus. However, research does suggest many modulating effects.

Here, we might attend to aspects such as complexity, contrast, order and grouping, all of which might be impacted stimulus size, resolution, presentation, as well as the ambient environment.

Visual surroundings are important in perception and object recognition (Bar, 2004; Todorović, 2010). Think, for example, of the Ebbinghaus illusion (Figure 4, A.), which has shown that perception of size can be modulated by surrounding visual information. A similar impact might be had from a picture frame, which may “tell the eye” where to look or potentially “complete” a painting (Ensor & Hamilton, 2014, p. 121), or, in one of the only empirical studies, might reduce recall of an artwork (Koutstaal, 1998).

Features may also once again tie to aspects of classification or felt importance, which can be further modulated within a processing experience. Texture, size, or felt authenticity may give a sense that something is special or made by a human (see Figure 4). These latter aspects may then impact early processing experience (see Hagtvedt & Patrick, 2008) or may be recalled in stages involving memory integration and knowledge of the participant. We may even talk of social context. In perception of faces, the so-called “cheerleader effect” where we tend to rate others as more attractive when surrounded by others in a peer situation (see e.g., Walker & Vul, 2014). Framing can also occur via external information such as with packaging and marketing programs, which can influence value and even meaning or symbolic associations (Locher, Overbeeke, & Wensveen, 2010), or even the approach, whereby we come into a first impression of a work or where it may be slowly revealed. Think for example, the staircase in the Louvre and the Nike of Samothrace (Figure 4, C.). Here, we may also attend to aspects such as brushstrokes or other features that tell us we are engaging with the work of another human, or cues regarding gestures.

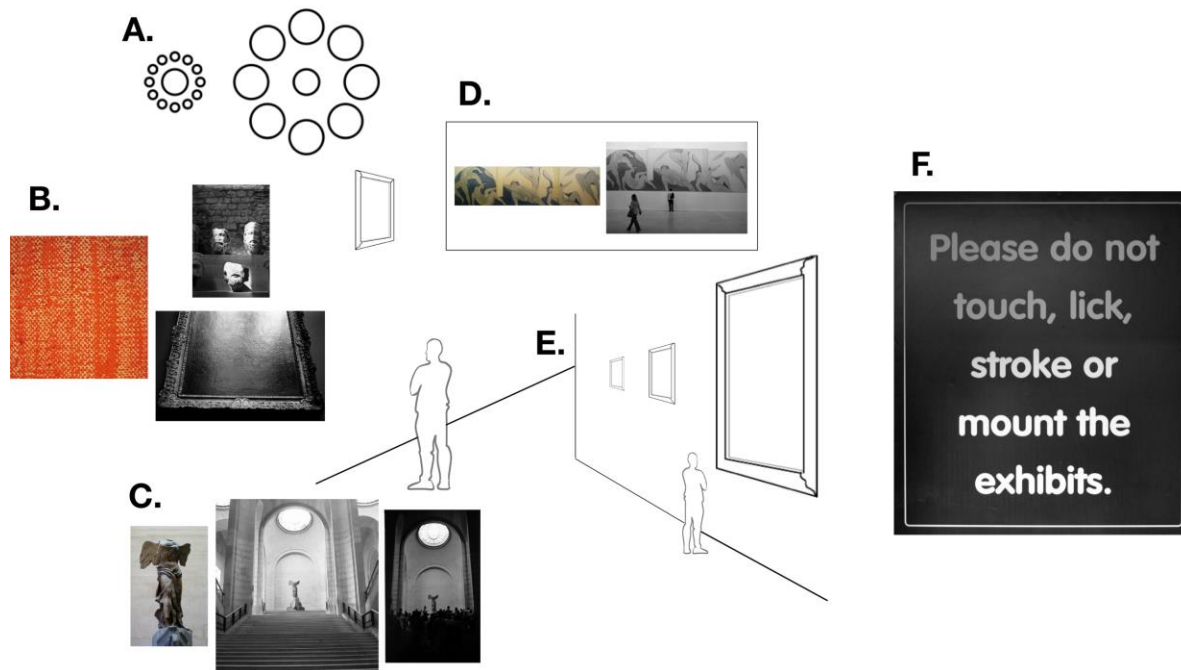


Figure 4. Examples of physical context and impact on classification, expectations, framing, perceived importance, even visual recognition and attention to objects. **A.** The Ebbinghaus illusion which can change the perceived size of the middle circle, although both left and right are the same actual size. **B.** Texture or lighting. **C.** Approach and physical or social/cultural context (Nike of Samothrace in the Louvre, Paris). **D.** Image size (The Unfinished Dance (1931) by Matisse: left image might be how this appears in a typical laboratory study; right side shows its actual appearance in the Palais de Tokyo, Paris). **E.** Hanging and impact on importance. **F.** Museum rules, posted at the Museum of Sex, New York City. (all photos by the first author).

Explicit classification—memory, expertise, training

Intermediary stages of vision should involve processing that recruits access to memory and or involve higher-order cognitions such as the perceiver's knowledge and background experiences. For example, certain features may lead to felt familiarity or prototypicality/novelty of media or contribute to assessed meaning in the final model stages (see also Tinio, 2013). We may also consider style of an aesthetic piece.

A particularly well-travelled though still very much fruitful area of study involves expertise or particular knowledge about a topic. Numerous studies have shown that art expertise—especially involving past training in the arts or aesthetic contexts, as well as working artists—influences our perception, understanding, and evaluation of art (Augustin & Leder, 2006; Brieber, Leder, & Nadal, 2015; Chamberlain & Wagemans, 2015; Leder, Gerger, Brieber, & Schwarz, 2014). E.g., experts may

have an emphasis on formal, stylistic, and relational properties rather than content, craftsmanship or mimesis (Hekkert & Van Wieringen, 1996). Cupchik and Gebotys (1988) showed that art naïve viewers may take an approach to art that is essentially an extension of everyday perception. Thus, they mainly search for recognizable elements that can elicit pleasurable associations.

Becker (1982, p. 48) notes that specialized audience members may also use “the history of attempts to make similar works in that medium or genre; characteristic features of different styles and periods in the history of the art; ...an acquaintance with various versions of the same work; and an ability to respond emotionally and cognitively to the manipulation of standard elements in the vocabulary of the medium.” Deines (2013, p. 29) calls such contextual factors “relational properties”: “If I dislike a work because I think it deals with important [issues] in too simple a manner, if I find it boring because the artist repeats him or herself ... if I find it witty ... all these properties ... depend on my knowledge of its context, of the time of production, of the artist and her or his previous work.” Fenner (2003, p. 48), similarly speaking from the purview of art criticism, also suggests numerous other “contexts under which we consider aesthetic objects” including ethnic, racial, class, gender (“Do I see Milos Forman's *One Flew Over the Cuckoo's Nest* as essentially misogynist?”), national or political (think back to Baumgarden). We might also connect knowledge to specific *Objects*. Certain pieces or styles may require certain backgrounds or can lead to very different reactions for different *Persons*. (For more on this aspect, focusing on art history, see Bullot & Reber, 2013; for a discussion tied to design, see Locher, Overbeeke & Wensveen, 2010).

The role of expertise is also far from settled. Locher (2015) suggests that experts may give more importance to the initial impression and resulting affective reaction when appraising value or authenticity of artworks. Experienced viewers may also be more likely to see an artwork as an autonomous entity, and to perceive artworks “for their own sake” (Hekkert & Van Wieringen, 1996).

Expertise and knowledge throughout stages of experience

Interestingly, like many factors, previous knowledge may also play a role throughout several model stages. For example, expertise can impact initial framing. Winston and Cupchik (1992) suggest novices hold beliefs that art should please, evoke peaceful feelings, positive memories, and appeal to many people. Experienced people often expect that art will challenge the viewer's conception, express the artist's deep feelings, and require some effort. This then again may determine what we look for, what we perceive, and how we respond to perceived aspects.

Differences as a function of expertise can even manifest in low-level visual processing. Experts tend to look more at a painting's periphery, evaluate the entirety of the canvas, and use a global or visual "gist" perspective, whereas art naïve viewers focus more on the center of a work (Pihko et al., 2011). Experts have even been shown to be less swayed by context or priming—i.e., as in labels or added social information (Kirk, Harvey & Montague, 2011).

Mood, desired emotion

One factor that presumably plays a role in all stages, but may be especially important in explicit classification, involves mood or emotional experience. As can be seen in the processing model, as well as that of Leder et al. (2004), mood/emotion might be thought of as its own channel of important information, continuously updated through our interactions. Before beginning an encounter, mood or emotion carried over from prior interactions may influence experience. Cognition can be mediated through mechanisms of arousal or general valence tied to how one is feeling (Derryberry, 1988; Gendolla, 2000). Mood may also modulate processing via "directive impact" (Gendolla, 2000) by influencing behavioral preferences in "compliance with a hedonic motives"—i.e., avoiding displeasure, seeking pleasure. Individuals who are in a bad mood may seek out happy or positive experiences so as to regulate their feelings. We may then use this information to modify our engagement—i.e., continuing to interact with or leaving alone an artwork. Derryberry (1988, p. 26) for example notes a "mood congruity effect," a phenomenon in which individuals tend to recall information consistent with their ongoing mood.

We may also modify our processing mode so as to diminish the emotion-arousing potential. Such a “selective attention” (Derryberry, 1988) may essentially align with taking a more detached aesthetic approach, focusing on formal aspects instead of emotion-related content. For example, Cupchik and Wroblewski-Raya (1998; Cupchik, 2013 for review) considered participants who were assessed on a measure of loneliness, and who were asked to assess paintings of solitary (presumably lonely) figures. Already lonely subjects preferred stylistic qualities, such as color or composition, over the subject matter. This may influence selection of media in the pre-state, but may be particularly important in the explicit classification stage, where Pelowski and colleagues (2017) suggested that there may be an “emotion check” wherein individuals first become aware of more complex emotional content or evoking potential—beyond basic valence (see also Cupchik, 2013)—and then modulate their mode of engagement to maximize or minimize their resulting feelings. Individuals may also not always seek out pleasure or avoid negative affect. They may do the opposite if this fits the particular personality and context—think of an individual listening to sad music after a breakup (e.g., see Kemp & Cupchik, 2007 for empirical evidence). Moods may also change depending on weather, odors, lighting (Gendolla, 2000 for review).

Cognitive Mastery and Secondary processing—judgments, specific emotions

Moving onto the later model stages, it is argued that the individual will essentially build up a collection of information from the earlier processing (low-level sensory aspects, identified objects, felt familiarity/novelty, arousal and valence, meaning and associations) and combine these to create final meaning or understanding, to define specific judgements, or to further engage or to process particularly compelling or difficult aspects. Here, we can also talk of some key aspects of context.

Processing success, congruency, fluency/discrepancy

First, the felt ease or difficulty one has had in their interaction up until this stage may be key, especially in formulating a judgment. This can be used in a declarative fashion, as in many studies—answering if something is beautiful, pleasing, interesting—or more organically can guide future

behavior—do I purchase it, do I want to see it again?—or to drive further engagement. Empirical and theoretical work has suggested that evaluations—i.e., whether we like or find something pleasing—might often be shaped by our own felt difficulty within the processing experience. This is often connected to “fluency” (Forster, Leder, & Ansorge, 2016; see also the dedicated chapter here). As noted by Graf and Landwehr (2015, p. 6), fluency might occur on two or more levels— perceptual fluency is concerned primarily with identifying the physical identity of a stimulus; conceptual fluency is concerned primarily with stimulus meaning or associations. Especially in cases of quick responses—such as a typical 1-5s lab rating task—or where one does not have much vested interest, we may default to our fluency to make an assessment (i.e., easy = good; see Schwarz, 2011). Fluency may also fit into more general discussions of a ‘gut’ response. Gendlin (1995; also Schwarz, 2011) explicitly connects this to aesthetic acts such as finding the ‘right’ word for a poem. This may also be a source for aesthetic sensitivity wherein we choose a superior piece, perhaps without being able to truly articulate why (think back to mid-20th century empirical studies—perhaps a nice future target).

Disfluency, or discrepancy can also be important. This may serve as a cue that leads us to engage further or adopt a systematic approach—carefully perceiving and thinking about a stimulus or renewing processing through subsequent cycles –whereas fluency leads to more superficial processing (Bulot & Reber, 2013). Silvia (2009) also suggests that felt ease or perhaps more often difficulty are a basis of appraising relative novelty, familiarity, challenge, as well as general positive or negative affective valence.

While acting as a component of context in its own right, fluency can also of course be impacted by other context. As noted by Graf and Landwehr (2015, p. 6), this might occur on two or more levels— perceptual fluency is concerned primarily with identifying the physical identity of a stimulus, and may be impacted by aspects such as high contrast, color saturation, symmetry. Conceptual fluency is concerned primarily with stimulus meaning and modulated by one’s training and previous exposure (see also Reber, Schwarz, & Winkielman, 2004). Fluency can also be

manipulated by other ambient factors. In one intriguing study, Lee, Kim, and Schwarz (2015) suggested that the environment might be changed to make something feel a bit “off”. By putting fish oil under a desk—creating the ambient feeling of something fishy or strange, even though this occurred below the level of overt awareness—they increased participants’ likelihood of detecting a semantic distortion or to engage in negative hypothesis testing (falsifying their own initial hunch).

Self-importance, perceived detachment/threat—specific emotions or aesthetic outcomes

At the same time, especially when interaction is disfluent or discrepant, felt ease may also be complimented by another component involving whether or not they really care about the outcome of their viewing, or whether or not they really have an interest in or need to process. If one is only minimally engaging, or finds their interaction particularly fluent, they may have a generally positive appraisal and move on. However, one may also find that the event is deeply important, or they may have a vested interest in the outcome; it may also be dangerous.

Notably, Silvia (2009) in his discussion of “appraisal theories” suggests two checks—for “congruency” and “self-relevance”—which may guide the actual types of emotions we may experience and report (determined in Cognitive Mastery stage) beyond basic arousal and valence. For example, in the case of interest, Silvia (2005) suggests that our appraisal structure would consist of: (1) a judgment of high novelty/complexity (low congruence), combined with (2) low self-relevance or little importance for one’s goals, expectations, or low threat. In contrast, anger (Cooper and Silvia, 2009) would combine appraising an event as (1) inconsistent with one’s schema (low schema congruence), but also with (2) close tie to one’s goals/self, and thus leading to a vested interest. He also adds the third element of “coping potential” (Silvia, 2005) or the need to stick out an engagement or take responsibility; low in the anger case (i.e., an experience is someone else’s fault). Because of this structure, Silvia concludes that in any situation, different people will have different responses to these processing checks, and thus different emotions to the same stimulus, or the same person may even have different emotions depending on context.

This role of self-relation, and resulting adjustment, may also tie to the topic of aesthetic processing mode, which would essentially be a detached low self-relation event, or may be sought out to create such conditions by switching attention to formal aspects. This may also relate to felt versus perceived (or so-called “aesthetic”) emotions. Several theories propose that emotions in the arts can be distinctively different from emotions occurring in everyday life (Frijda, 1988; Scherer, 2005; see Gerger et al., 2014 for review). Presumably this is because the aesthetic or art setting is disengaged from personal importance or implications, or one can perceive what emotion a work is ‘trying’ to communicate without actually feeling it (see Barwell, 1986 for a nice art critical discussion). However this could probably be duplicated in other contexts.

Pelowski and Akiba (2011; refined in Pelowski et al., 2017) go further and argue that self-schema matching and relation to the environment may be used to differentiate broad varieties of experience—from mundane to profound. Following previous researchers such as Carver (1996), they suggest that, in addition to general personality characteristics or differences in background and motivation, individuals hold this set of postulates—an “ideal self image”. This can be thought of as a hierarchical set of postulates, headed by core goals (who am I and who do I want to be?) branching down into goals for general actions (what should I do to protect and advance my core aspirations and beliefs?) and then further branching down into lower and more specific beliefs and schema. Interacting would then always entail the matching of the self, via specific schema, to the ‘reality’ of the environment, and thus itself a core aspect of context from the pre-state onward.

They suggest the same essential congruency and self-relevance checks, situated between memory integration and cognitive mastery, and divide outcomes of experience into five varieties—(1) cases where individuals have a low self relevance and low discrepancy, leading to rather facile or surface engagement with little emotion; (2) discrepancy but also low self relevance, which may allow individuals to find and enjoy novelty or small insights (discovering a new way to depict something or an unexpected word use, but which is not terribly upsetting to one’s entire worldview); (3) cases with

high self relevance but also low discrepancy, which, especially where one can process with ease or finds a resonant emotion, may lead to felt harmony or even “flow” type experiences; and cases where one has both discrepancy and high self relevance which lead to a felt threat, potentially leading either to (4) negative emotion and evaluations as one tries to extricate themselves and protect the self; or (5) a potential for self revision and transformative experience. As can be seen in Figure 5, at all stages, the way that feelings and reactions are assessed can be explicitly tied to context, as well as several levels of information that all may be assessed to determine an outcome or guide response.

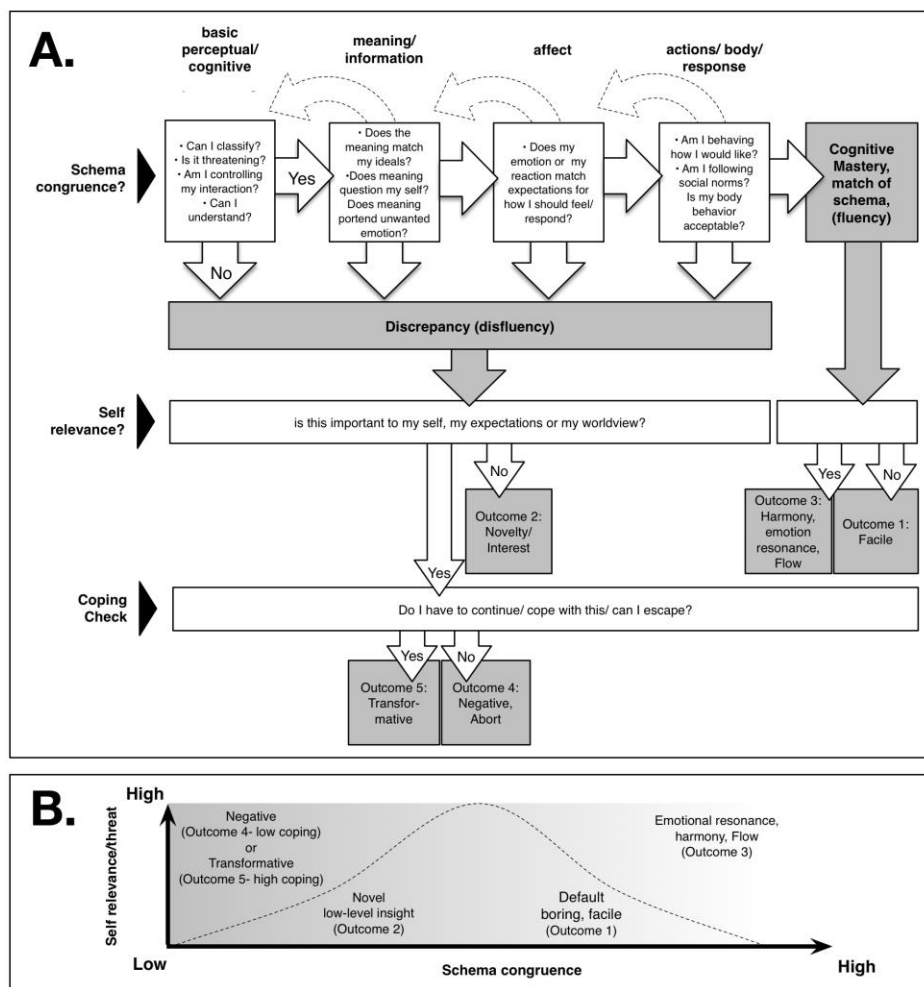


Figure 5. Proposed tie between self-relevance, schema congruency, and need for coping “checks” and outcomes of aesthetic experience as well as examples of processing levels. (adapted from Pelowski et al., 2017).

The determinant of whether or not one feels self-relevance or how they cope with threat may in turn be modulated by factors. This could vary greatly between laboratory and in situ engagements. Individuals may not feel personal involvement or threat (see also Mandel, 2003). How we respond to particularly strong emotions or reactions could also tie to context. On one hand they may resonate or create particularly pleasurable experience. At the same time, if one feels such a response in cases where they are inappropriate due to social norms or other factors, they may cause individuals to disengage or negatively respond (Gendolla, 2000). The sociologist Goffman (1974, p. 353) talks of “flooding out,” which “occurs when an individual must accept restraints on bodily behavior over an appreciable portion of his body” and may thus lead to negative response. When facing particularly challenging and self-important situations, contextual sensations (hearing one’s own footsteps, sensation of blown air conditioning) may also help to shift attention from coping or avoiding issues to one’s current actions, encouraging self-reflection or insight (e.g., Pelowski et al., 2012 for discussion with art). These physical “triggers” maybe important for future research (see also Rothbaum et al, 1982 for discussion outside art).

Time; General movement through stages

Finally, by placing the above myriad context features into a general frame, we can of course also consider context aspects that might provide important tools to unlocking or assessing the above progressions or processing sequences. For example, time may be a key factor that is only beginning to be fully understood (see e.g., Locher, Overbeeke, & Wensveen, 2010). Truncating time for viewing (e.g., below about 5 seconds) may lead to higher variance in appraisals due to inability to conduct all processes or integrate all information that individuals may normally use (Bachmann & Vipper, 1983; Augustin, Leder, Hutzler, & Carbon, 2008). This also nicely fits the above processes to topic of the underlying functional neurobiology (Cela-Conde et al., 2013).

4. Major Challenges, Goals, and Suggestions

This was again but a small taste of the context discussion, but hopefully can be useful in making clear: (1) how prevalent and important context is in aesthetic or basic perceptual and psychological experience. (2) By tying this discussion to several models and theoretical frameworks, it is hoped that the reader might go further and have a sense of how and where context might play specific roles in modulating processes or outputs. This is not to say that the above models and discussion are necessarily perfect or could benefit from other explanations. To conclude this chapter, we consider some other questions and avenues for future research:

Other aspects of context—brains, bodies, an endless list of possibilities.

First, there are of course innumerable other areas deserving of research that could not be mentioned in this chapter. Emerging evidence suggests that individuals who develop brain damage or disease (e.g., fronto-temporal dementia, Parkinson's disease) can show marked changes in the way they produce or appreciate art (Chatterjee, 2011; Grettton & Ffytche, 2014; Lauring et al., submitted). Studies also document changes that can occur, temporarily, via causative manipulation of brain activity (Cattaneo et al., 2014a, 2014b, 2015)—for example modulating assessed beauty or liking. Neuropharmacology or modulating the levels of hormones (e.g., testosterone) or other neurotransmitters such as dopamine, serotonin may also greatly impact experience (Spee, Ishizu, Leder, & Pelowski, 2018 for aesthetic-related review), but have so far not been applied empirically. Other causative interventions such as drugs or alcohol may play important roles—for example, drinking a glass of wine may increase our likelihood of rating others as attractive (Van Den Abbeele et al., 2015).

Other intriguing areas involve the role of and awareness of the body and how this may tie to felt valence, difficulty, or drive perception (e.g., Gibson, 1979; Turner & Penn, 2002). For example, Kapoula et al. (2011) showed that postural sway and stability when individuals stand before paintings can be modulated depending on local depth information and that individuals unconsciously adopt leaning or upright postures of tall sculptures by Richard Serra. Moving one's body in accordance with

the brushstrokes of a painting (pointillism or postimpressionism) positively affected liking of art in a style corresponding to the movement (Leder et al., 2012). Walking as opposed to sitting may also modulate engagement. This leads to increased activation in the prefrontal cortex; even mentally preparing to walk to a target, as might occur upon seeing a painting across a gallery, leads to increased activation (Suzuki et al., 2008).

We might also look at specific physiological responses, such as tears, which could be a signal to the individual that something profound is happening and they should withdraw or introspect (Cotter et al., 2017; Pelowski, 2015). Other feelings such as chills may also signal a harmony or peak emotional or resonant experience (in a song, movie, or art; McCrae, 2007; Pelowski, Markey, & Leder, 2018). There may also be specific ties between emotions—for example, Eskine, Kacinik, and Prinz (2012) showed that priming individuals with fearful movies increased their likelihood of rating images as sublime, as did exercise (jumping jacks) before viewing.

It is also important to consider how we assess context. Carefully weigh the costs and benefits of certain designs. A researcher might in fact try to weigh multiple possible context features to decide which can and cannot be manipulated. At the same time, researchers, in study planning, should also spend time brainstorming context-effects that they had not considered but which might have influence or confound findings. Since context can by definition be almost everything an experimenter has to select specific properties out of numerous possibilities to make context into an experimental factor that can be investigated. At least for the first author, the failure to catch a conflating explanation is a common reason for rejecting journal paper submissions! Context studies are also impacted by what we don't know about aesthetic processing. For conducting experimental laboratory research, it is necessary to know in advance what the key variables are, but it is not clear if empirical aesthetics has already figured out all of these variables. Therefore, it still remains a great challenge to investigate the effect of context on aesthetic processing.

On site aesthetics, what can be done in the lab?

The experimental setting also needs consideration. As asked by Leder and Nadal (2014, p. 453-4) “To what extent does the laboratory distort or attenuate aesthetic episodes?” This question also contains within it the same ambivalence as found throughout aesthetics’ history. “...Given how much ... work has been performed in laboratory settings, it is also clear that the psychology of art and aesthetics has traditionally accepted the formalist argument that aesthetic experience is, to a large extent, contextually impermeable.” This may not be true.

It is also important to stress that these layers and issues are *not* only present in so-called ecologically-valid interactions or with a ‘real’ stimulus. They can and do manifest in the laboratory (see left side Figure 1). It may also be that there are certain reactions—intense emotion, feelings of self threat, change, transformation—that are very difficult to evoke in a lab, precisely because the context is not of the right type (Tallis, 2008; Cela-Conde et al., 2011). The next decade plus of study must go out into the field to capture and understand such rich experiences. This is already beginning with mobile measures of eye-tracking, brain imaging (EEG, NIRS), and physiology that can bring complex data from the “real world”.

Predictive coding and top-down/bottom-up processing

Another important area for considering context involves the topic of “predictive coding” or the possibility that, especially when considered at the neural level, much of our interactions or perceptions—at least in initial stages—are driven largely by prior schema or internal models which guide and adjust perception (Clark, 2013). This at once offers the theoretical promise of fitting aesthetic models or processes to mathematical functions which may provide a much higher level of nuance. It also raises the questions of what can truly be thought of as bottom-up, environment-derived perceptions versus top-down aspects of the perceiver.

What is not impacted by context—a search for immutable aesthetic factors

Another important question involves the topic of what factors are or are not ultimately modulated by context differences. As made clear in the above introduction, the field of aesthetics

derives from a history of, at least intuited if not often empirically verified, ideas that some reactions are innate or immutable. It would be fascinating to find out what factors might be more or less stable across differing conditions, and which may not. Some intriguing directions include evidence in aesthetic ratings of paintings by dementia or Alzheimer's patients, who, despite no longer remembering past trials tend to show both consistency in their judgments and also with healthy control groups (Graham et al., 2013; Halpern et al., 2008). There is also need for more cross-cultural study. For example, curves as opposed to sharp edges may be universally preferred (Gómez-Puerto, Rosselló, Corradi, Acedo-Carmona, Munar, & Nadal, 2017). A similar point of general agreement might involve color. For example, following the early shared taste (for blue) found by Osgood, similar more recent studies have shown a general dislike of dark yellow/olive (Palmer & Schloss, 2010; however see Saito, 1996 for differences in appreciation of white). Intriguingly, it may be that which remains when context is accounted for that can be a window into our shared aesthetic response. As put by Jerry Fodor (in Rollins, 2004, p. 177) "The universality of an artwork consists precisely in its independence from its historical [and, presumably other] context." We might also ask if is meaningful to ever assess reactions to stimuli divorced from context? Why and what could this tell us about psychology or empirical aesthetics?

5. Application to the classroom or the applied setting

Last, many of the above topics can also make their way into the classroom for discussing context. Once one considers the scope whereby context impacts our aesthetic interpretations, the potential for application in the classroom or in the world is, of course, endless. However, to provide a few more examples: It may be interesting as a learning device to take well-grounded studies or ideas from aesthetics (pre-conceptions, mood, expectations) and discuss how these would translate to the reactions and evaluations we make everyday. How does what we know about the impact of social priming on our evaluations of others' beauty impact our shopping or job hiring decisions? How does

mood or our response to others impact how we ‘see’ and respond to the world. Not only at the level of our hedonic appraisals or emotions, but also at the cognitive level regarding what we pay attention to or how we respond to cognitive processes?

We would also advise students to look outside empirical aesthetic books and journals to a number of other sources. Some of the best descriptions of context importance, and research ideas, in the first author’s own training have come from sociology, anthropology, etc. For particularly interesting reads, try *The Meaning of Things; Domestic Symbols and the Self* (Csikszentmihalyi & Rochenberg-Halton, 1981); *Frame Analysis; An Essay on the Organization of Experience* (Goffman, 1974); see also Alexander (1977) for great consideration of the built environment; Gell (1998) for an anthropological perspective of the co-relationship with art; *Art Worlds* (Becker, 1982); *Nudge: Improving decisions about health, wealth, and happiness* (Thaler & Sunstein, 2008); *The transfiguration of the commonplace* (Danto, 1981). Those interested in context might also consider the discussions of cultural literacy (Hirsch, 1987), visual literacy, techno literacy (Dakers, 2006), etc., and how what “we all know” or more likely what we “should know” in order to “properly respond” in aesthetics may reveal what is going on under the surface.

References

- Abel, T., & Hsu, F. L. K. (1949). Some aspects of personality of Chinese as revealed by the Rorschach test. *Rorschach research exchange and journal of projective techniques*, 12, 285-301.
- Adams, F. M., & Osgood, C. E. (1973). A cross-cultural study of the affective meanings of color. *Journal of cross-cultural psychology*, 4(2), 135-156.
- Alexander, C. (1977). *A Pattern Language; Towns, Buildings, Construction*. Oxford University Press.
- Arai, S., & Kawabata, H. (2016). Appreciation contexts modulate aesthetic evaluation and perceived duration of pictures. *Art & Perception*, 4(3), 225-239.
- Augustin, M. D., & Leder, H. (2006). Art expertise: A study of concepts and conceptual spaces. *Psychological Science*, 48, 135–156.
- Augustin, M. D., Leder, H., Hutzler, F., & Carbon, C. C. (2008). Style follows content: On the microgenesis of art perception. *Acta Psychologica*, 128(1), 127-138
- Bachmann, T., & Vipper, K. (1983). Perceptual rating of paintings from different artistic styles as a function of semantic differential scales and exposure time. *Archiv fur Psychologie*, 135, 149–161.
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5(8), 617–629.
- Barron, F., & Welsh, G. S. (1952). Artistic perception as a possible factor in personality style: Its measurement by a figure preference test. *The Journal of Psychology*, 33(2), 199-203.
- Barwell, I. (1986). How does art express emotion?. *The Journal of aesthetics and art criticism*, 45(2), 175-181
- Baumgarten, A. G. (1750). *Aesthetica*. Hildesheim: Olms.
- Baumgarten, A.G. (1954). *Reflections on poetry: Meditationes philosophicae de nonnullis ad poema pertinentibus*, K. Aschenbrenner and W.B. Holther (Eds.), Berkeley: University of California Press.
- Becker, H. S. (1982). *Art Worlds*. University of California Press.
- Bell, C. (1924). *Art. 1914*. London: Chatto and Windus.
- Berger, J. (1972). *Ways of Seeing*. London: BBC.
- Berlyne, D. E. (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation*. Washington, DC: Hemisphere.
- Berlyne, D. E. (1976). Similarity and preference judgments of Indian and Canadian subjects exposed to Western paintings. *International Journal of Psychology*, 11(1), 43-55.
- Berlyne, D. E., Ogilvie, J. C., & Parham, L. C. C. (1968). The dimensionality of visual complexity, interestingness, and pleasingness. *Canadian Journal of Psychology*, 22, 376-387.
- Bitgood, S. (1992). The anatomy of an exhibit. *Visitor Behavior*, 7, 4–15.
- Bitgood, S., & Patterson, D. D. (1993). The effects of gallery changes on visitor reading and object viewing time. *Environment and Behavior*, 25, 761–781.
- Boas, G. (1943). Art and reality. *College Art Journal*, 2, 115.
- 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.
- Brieber, D., Nadal, M., Leder, H., & Rosenberg, R. (2014). Art in time and space: Context modulates the relation between art experience and view- ing time. *PLoS One*, 9, e99019.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste* (R. Nice, Trans.). Cambridge, MA: Harvard University Press.
- Bullot, N. J., & Reber, R. (2013). The artful mind meets art history: Toward a psycho-historical framework for the science of art appreciation. *Behavioral and brain sciences*, 36(2), 123-137
- Bullough, E. (1912). “Psychical distance” as a factor in art and an aesthetic principle. *British Journal of Psychology*, 2, 87-118.

- Carver, C. S. (1996). Cognitive interference and the structure of behavior. In: Sarason IG, Pierce GR, Sarason BR, editors. *Cognitive interference; theories, methods, and findings* (pp. 25–46). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cattaneo, Z., Lega, C., Flexas, A., Nadal, M., Munar, E., & Cela-Conde, C.J. (2014a). The world can look better: enhancing beauty experience with brain stimulation. *SCAN* 9, 1713–1721.
- Cattaneo, Z., Lega, C., Gardelli, C., Merabet, L.B., Cela-Conde, C.J., & Nadal, M. (2014b). The role of prefrontal and parietal cortices in esthetic appreciation of representational and abstract art: a TMS study. *Neuroimage* 99, 443–450.
- Cattaneo, Z., Lega, C., Ferrari, C., Vecchi, T., Cela-Conde, C.J., Sivanto, J., & Nadal, M. (2015). The role of the lateral occipital cortex in aesthetic appreciation of representational and abstract paintings: a TMS study. *Brain and Cognition*. 96, 44–53.
- Cela-Conde, C. J., Agnati, L., Huston, J. P., Mora, F., & Nadal, M. (2011). The neural foundations of aesthetic appreciation. *Progress in Neurobiology*, 94, 39–48.
- Cela-Conde CJ, García-Prieto J, Ramasco JJ, Mirasso CR, Bajo R, Munar E, et al. Dynamics of brain networks in the aesthetic appreciation. *Proc Natl Acad Sci USA* 2013;110:10454–61.
- Chamberlain, R., & Wagemans, J. (2015). Visual arts training is linked to flexible attention to local and global levels of visual stimuli. *Acta Psychologica*, 161, 185–197
- Chamorro-Premuzic, T., Reimers, S., Hsu, A., & Ahmetoglu, G. (2009). Who art thou? Personality predictors of artistic preferences in a large UK sample: The importance of openness. *British Journal of Psychology*, 100(3), 501–516.
- Chang, E. (2006). Interactive experiences and contextual learning in museums. *Studies in Art Education*, 47, 170–186.
- Chatterjee, A. (2004). Prospects for a cognitive neuroscience of visual aesthetics. *Bull. Psychol. Art* 4, 55–60.
- Chatterjee, A. (2011). Visual Art. In J.A. Gottfried (Ed.), *Neurobiology of sensation and reward*. Boca Raton, FL: CRC Press.
- Chatterjee, A., & Vartanian, O. (2014). Neuroaesthetics. *Trends in cognitive sciences*, 18(7), 370–375.
- Child, I. L. (1962). Personal preferences as an expression of aesthetic sensitivity 1. *Journal of Personality*, 30(3), 496–512.
- Child, I. L. (1970). Aesthetic judgment in children. *Transaction*, 7(7), 47–51.
- Chiu, L. H. (1972). A cross-cultural comparison of cognitive styles in Chinese and American children. *International journal of psychology*, 7(4), 235–242.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavior and Brain Science*, 36(03), 181–204.
- Cotter, K. N., Silvia, P. J., & Fayn, K. (2017). What does feeling like crying when listening to music feel like? *Psychology of Aesthetics, Creativity, and the Arts*.
- Csikszentmihalyi, M., & Rochenberg-Halton, E. (1981). *The Meaning of Things; Domestic Symbols and the Self*. Cambridge University Press.
- Cupchik, G. C. (1995). Emotion in aesthetics: Reactive and reflective models. *Poetics*, 23(1-2), 177–188.
- Cupchik, G. C. (2002). The evolution of psychical distance as an aesthetic concept. *Culture & Psychology*, 8(2), 155–187.
- Cupchik G. C. (2013). I am, therefore I think, act, and express both in life and in art. In: Roald T, Lang J, (Eds.). *Art and identity: essays on the aesthetic creation of mind: Consciousness, literature the arts* (pp. 67–92.). Amsterdam: Rodopi.
- Cupchik, G. C., & Gebotys, R. J. (1988). The search for meaning in art: Interpretive styles and judgments of quality. *Visual Arts Research*, 14, 38–50.
- Cupchik, G. C., Shereck, L., & Spiegel, S. (1994). The effects of textual information on artistic communication. *Visual Arts Research*, 62–78.

- Cupchik, G. C., Vartanian, O., Crawley, A., & Mikulis, D. J. (2009). Viewing artworks: contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain and cognition*, 70(1), 84-91.
- Cupchik, G. C., & Wroblewski-Raya, V. (1998). Loneliness as a theme in painting. *Visual Arts Research*, 24, 65-71.
- Currie, G. (1985). The authentic and the aesthetic. *American Philosophical Quarterly*, 22, 153-160.
- Dakers, J. R. ed. (2006). *Defining Technological Literacy: Towards an Epistemological Framework*. Palgrave.
- Danto, A. C. (1981). *The transfiguration of the commonplace: a philosophy of art*. Harvard University Press.
- Darley, J. M., Fleming, J. H., Hilton, J. L., & Swann Jr, W. B. (1988). Dispelling negative expectancies: The impact of interaction goals and target characteristics on the expectancy confirmation process
- Deines, S. (2013). Art in context: On cultural limits to the understanding, experience and evaluation of works of art. In J. deMul & R. van de Vall (Eds.). *Gimme shelter: Global discourses in aesthetics, international yearbook of aesthetics*, Volume 15 (pp. 23-40). Amsterdam University Press.
- Dennett, D. (1988). Quining qualia. in A. Marcel and E. Bisiach (Eds.), *Consciousness in Modern Science*, Oxford: Oxford University Press.
- Derryberry D. (1988). Emotional influences on evaluative judgments: roles of arousal, attention, and spreading activation. *Motivation and Emotion*, 12(1), 23-55.
- Descartes, R. (1965). *Discourse on method, optics, geometry and meteorology*, P. J. Olscamp (Trns.), New York: Hackett.
- Dissanayake, E. (2007). What art is and what art does: An overview of contemporary evolutionary hypotheses. In C. Martindale, P. Locher, & V. Petrov (Eds.), *Evolutionary and neurocognitive approaches to aesthetics, creativity, and the arts* (pp. 1-14). Amityville, NY: Baywood.
- Ensor, K., & Hamilton, M. (2014). Effect of distinctive frames on memory for pictures. *Empirical Studies of the Arts*, 32, 121-131.
- Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2012). Stirring images: fear, not happiness or arousal, makes art more sublime. *Emotion*, 12(5), 1071.
- Eysenck, H. J. (1972). Personal preferences, aesthetic sensitivity and personality in trained and untrained subjects. *Journal of Personality*, 40(4), 544-557.
- Falk, J. H. (1993). Assessing the impact of exhibit arrangement on visitor behavior and learning. *Curator*, 36, 133-146.
- Falk, J. H. (2009). *Identity and the museum visitor experience*. Walnut Creek, CA: Left Coast Press.
- Falk, J., & Dierking, L. (1992). *The museum experience*. Washington, DC: Whalesback Books.
- Fayn, K., MacCann, C., Tiliopoulos, N., & Silvia, P. J. (2015). Aesthetic emotions and aesthetic people: Openness predicts sensitivity to novelty in the experiences of interest and pleasure. *Frontiers in Psychology*, 6, 1877.
- Fechner, G. T. (1876). *Vorschule der aesthetik*. Leipzig: Breitkopf und Härtel.
- Fenner, D. E. (2003). Aesthetic experience and aesthetic analysis. *Journal of Aesthetic Education*, 40-53
- Forster, M., Leder, H., & Ansorge, U. (2016). Exploring the subjective feeling of fluency. *Experimental Psychology*, 63(1), 45.
- Frijda, N. H. (1988). The laws of emotion. *American psychologist*, 43(5), 349.
- Funch, B. S. (1997). *The psychology of art appreciation*. Copenhagen: Museum Tusculanum Press, University of Copenhagen.
- Gartus, A., & Leder, H. (2014). The white cube of the museum versus the gray cube of the street: The role of context in aesthetic evaluations. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 311-320.

- Gell, A. (1998). *Art and Agency, An Anthropological Theory*. Oxford.
- Gendlin, E. T. (1995). Crossing and dipping: Some terms for approaching the interface between natural understanding and logical formulation. *Minds and Machines*, 5, 547-560.
- Gendolla, G. H. E. (2000). On the impact of mood on behavior: An integrative theory and a review. *Review of General Psychology*, 4, 378–408.
- Gerger, G., Leder, H., & Kremer, A. (2014). Context effects on emotional and aesthetic evaluations of artworks and IAPS pictures. *Acta Psycho- logica*, 151, 174–183.
- Getzels, J. W., & Csikszentmihalyi, M. (1969). Aesthetic opinion: An empirical study. *Public Opinion Quarterly*, 33(1), 34-45.
- Gibson, J. J. (1955). *The ecological approach to visual perception*. London: Lawrence Erlbaum Associates.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press
- Gómez-Puerto, G., Rosselló, J., Corradi, G., Acedo-Carmona, C., Munar, E., & Nadal, M. (2017). Preference for Curved Contours Across Cultures. *Psychology of Aesthetics, Creativity, and the Arts*.
- Graf, L. K., & Landwehr, J. R. (2015). A dual-process perspective on fluency-based aesthetics: the pleasure-interest model of aesthetic liking. *Personality and Social Psychology Review*, 19, 395–410.
- Graham, D. J., Stockinger, S., & Leder, H. (2013). An island of stability: Art images and natural scenes—but not natural faces—show consistent esthetic response in Alzheimer’s-related dementia. *Frontiers in psychology*, 4, 107.
- Gretton, C. & Ffytche, D. H. (2014). Art and the brain: A view from dementia. *International Journal of Geriatric Psychiatry*, 29(2), 111-126. doi: 10.1002/gps.3975.
- Hagtvedt, H., & Patrick, V. M. (2008). Art infusion: The influence of visual art on the perception and evaluation of consumer products. *Journal of Marketing Research*, 45, 379 –389.
- Halpern, A. R., Ly, J., Elkin-Franklin, S., & O’Connor, M. G. (2008). I know what I like: Stability of aesthetic preference in Alzheimer’s disease. *Brain and Cognition*, 66, 65–72.
- Hanquinet, L. (2013). Visitors to modern and contemporary art museums: Towards a new sociology of “cultural profiles”. *The Sociological Review*, 61, 790–813.
- Harris, J. A. (Ed.). (2013). *The Oxford handbook of British philosophy in the eighteenth century*. OUP Oxford.
- Hekkert, P., & Van Wieringen, P. C. (1996). Beauty in the eye of expert and nonexpert beholders: A study in the appraisal of art. *The American Journal of Psychology*, 389-407.
- Hendon, W. S., Costa, F., & Rosenberg, R. A. (1989). The general public and the art museum: Case studies of visitors to several institutions identify characteristics of weir publics. *American Journal of Economics and Sociology*, 48, 230–243.
- Hirsch, Jr., E. D. (1987). *Cultural Literacy: What Every American Needs to Know*. Houghton Mifflin Company.
- Huang, M., Bridge, H., Kemp, M. J., & Parker, A. J. (2011). Human cortical activity evoked by the assignment of authenticity when viewing works of art. *Frontiers in Human Neuroscience*, 5, 134.
- Hume, D. (1757). *Four dissertations*. London: A. Millar.
- Iwawaki, S., Eysenck, H. J., & Götz, K. O. (1979). A new visual aesthetic sensitivity test (VAST): II. Cross-cultural comparison between England and Japan. *Perceptual and Motor Skills*, 49(3), 859-862.
- Jacobsen, T., Schubotz, R. I., Höfel, L., & Cramon, D. Y. V. (2006). Brain correlates of aesthetic judgment of beauty. *Neuroimage*, 29(1), 276-285.
- Kapoula, Z., Adenis, M.-S., Lê, T.-T., Yang, Q., & Lipede, G. (2011). Pictorial depth increases body sway. *Psychology of Aesthetics, Creativity, and the Arts*, 5, 186–193.

- Kemp, S. W. P., & Cupchik, G. C. (2007). The emotionally evocative effects of paintings. *Visual Arts Research*, 33, 72–82.
- Kirk, U., Skov, M., Hulme, O., Christensen, M. S., & Zeki, S. (2009). Modulation of aesthetic value by semantic context: An fMRI study. *NeuroImage*, 44, 1125–1132.
- Kirk, U., Harvey, A., & Montague, P. R. (2011). Domain expertise insulates against judgment bias by monetary favors through a modulation of ventromedial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 10332–10336.
- Komar, V., & Melamid, A. (1997). *Painting by numbers: Komar and Melamid's scientific guide to art*. University of California Press.
- Koutstaal, W. (1998). Memory for picture frames. *Empirical Studies of the Arts*, 16, 47–57.
- Kühn, S., & Gallinat, J. (2012). The neural correlates of subjective pleasantness. *Neuroimage*, 61(1), 289–294.
- Lacey, S., Hagtvædt, H., Patrick, V. M., Anderson, A., Stilla, R., Deshpande, G., . . . Sathian, K. (2011). Art for reward's sake: Visual art recruits the ventral striatum. *NeuroImage*, 55, 420 – 433.
- Lauring, J. O., Ishizu, T., Kutlikova, H. H., Dörflinger, F., Haugbøl, S., Leder, H., Kupers, R., Pelowski, M. (submitted). Why would Parkinson's disease lead to sudden changes in creativity, motivation, or style with visual art?: Case evidence and neurobiological, contextual, and genetic hypotheses. *Perspectives on Psychological Science*.
- Lauring, J. O., Pelowski, M., Forster, M., Gondan, M., Pito, M., & Kupers, R. (2016). Well, if they like it... Effects of social groups' ratings and price information on the appreciation of art. *Psychology of Aesthetics, Creativity, and the Arts*, 10(3), 344
- Lazarus, R. S., & Alfert, E. (1964). Short-circuiting of threat by experimentally altering cognitive appraisal. *The Journal of Abnormal and Social Psychology*, 69(2), 195
- Leder, H., Bär, S., & Topolinski, S. (2012). Covert painting simulations influence aesthetic appreciation of artworks. *Psychological Science*, 23, 1479–1481.
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95, 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(6), 1137–1147
- 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(4), 443–464.
- Lee, D. S., Kim, E., & Schwarz, N. (2015). Something smells fishy: Olfactory suspicion cues improve performance on the Moses illusion and Wason rule discovery task. *Journal of Experimental Social Psychology*, 59, 47–50.
- Leibniz, G. W. (1902). *Discourse on metaphysics: Correspondences with Arnauld, and monadology*. G. Montgomery (Trns.), P. Janet (ed.), Chicago: Kessinger.
- Locher, P. J. (2015). “The aesthetic experience with visual art ‘at first glance’,” in *Investigations into the Phenomenology and the Ontology of the Work of Art: What are Artworks and How do we Experience Them?*, eds P. F. Bundgaard and F. Stjernfelt (New York, NY: Springer), 75–88.
- Locher, P., & Dolese, M. (2004). A comparison of the perceived pictorial and aesthetic qualities of original paintings and their postcard images. *Empirical Studies of the Arts*, 22, 129–142.
- Locher, P., Overbeeke, K., & Wensveen, S. (2010). Aesthetic interaction: A framework. *Design Issues*, 26(2), 70–79.
- Locher, P., Smith, L. F., & Smith, J. K. (1999). Original paintings versus slide and computer reproductions: A comparison of viewer responses. *Empirical Studies of the Arts*, 17, 121–129.
- Locher, P. J., Smith, J. K., & Smith, L. F. (2001). The influence of presentation format and viewer training in the visual arts on the perception of pictorial and aesthetic qualities of paintings. *Perception*, 30, 449–465.

- Mandel, D. (2003). Counterfactuals, emotions, and context. *Cognition & Emotion*, 17, 139-159.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, 98(2), 224.
- McArthur, L. Z., & Berry, D. S. (1987). Cross-cultural agreement in perceptions of babyfaced adults. *Journal of cross-cultural psychology*, 18(2), 165-192.
- McCrae, R. R. (2007). Aesthetic chills as a universal marker of openness to experience. *Motivation and Emotion*, 31(1), 5-11.
- Melton, A. W. (1972). Visitor behavior in museums: Some early research in environmental design. *Human Factors*, 14, 393-403.
- Myszkowski, N., Storme, M., Zenasni, F., & Lubart, T. (2014). Is visual aesthetic sensitivity independent from intelligence, personality and creativity?. *Personality and Individual Differences*, 59, 16-20.
- Nadal, M., Munar, E., Capó, M. À., Rosselló, J., & Cela-Conde, C. J. (2008). Towards a framework for the study of the neural correlates of aesthetic preference. *Spatial vision*, 21(3), 379-396.
- Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, 200906172.
- Parsons, M. J. (1987). *How we understand art; a cognitive developmental account of aesthetic experience*. Cambridge, England: Cambridge University Press.
- Plato. (1987). *The republic*. 2nd ed. D. Lee (Trns.), London: Penguin.
- Pekarik, A. J., & Schreiber, J. B. (2012). The power of expectation. *Curator*, 55, 487-496.
- Pelowski, M. J. (2015). Tears and transformation: Feeling like crying as an indicator of insightful or “aesthetic” experience with art. *Frontiers in Psychology*, 6, 1006.
- Pelowski, M., & Akiba, F. (2011). A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas in Psychology*, 29(2), 80-97.
- Pelowski, M., Forster, M., Tinio, P. P., Scholl, M., & Leder, H. (2017). Beyond the lab: An examination of key factors influencing interaction with ‘real’ and museum-based art. *Psychology of Aesthetics, Creativity, and the Arts*, 11(3), 245.
- Pelowski, M., Gerger, G., Chetouani, Y., Markey, P. S., & Leder, H. (2017). But is it really art? The classification of images as “Art”/“Not Art” and correlation with appraisal and viewer interpersonal differences. *Frontiers in psychology*, 8, 1729.
- 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.
- 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.
- Pelowski, M., Markey, P., & Leder, H. (2018). Chills, aesthetic experience, and new versus old knowledge—What do chills actually portend?: Comment on “Physics of mind: Experimental confirmations of theoretical predictions” by Schoeller et al. *Physics of Life Reviews*.
- Pelowski, M., Akiba, F., & Palacios, V. (2012). Satori, koan and aesthetic experience: Exploring the “realization of emptiness” in Buddhist enlightenment via an empirical study of modern art. *Psyke & Logos*, 33(2), 33.
- 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.
- Pinney, C., & Thomas, N. (Eds.). (2001). *Beyond Aesthetics*. Oxford: Berg.
- Poppe, A. (1907). *Alexander Gottlieb Baumgarten: Seine bedeutung und stellung in der Leibniz-wolffischen philosophie und seine beziehungen zu Kant*. Leipzig: Buchdruckerei Robert Noske.

- Martindale, C. (1988). "Aesthetics, psychobiology, and cognition," in *The Foundations of Aesthetics, Art, and Art Education*, eds F. H. Farley and R. W. Neperud (New York, NY: Praeger), 117–160.
- Mitias, M. H. (Ed.). (1986). *Possibility of the Aesthetic Experience*. Dordrecht: Martinus Nijhoff.
- Murdoch, I. (1977). *The fire and the sun; why Plato banished the artists*. Oxford: Oxford University Press.
- Osborne, H. (1979). Some theories of aesthetic judgment. *The Journal of Aesthetics and Art Criticism*, 38(2), 135-144.
- Ramachandran, V. S., & Hirstein, W. (1999). The science of art: A neurological theory of aesthetic experience. *Journal of consciousness Studies*, 6(6-7), 15-51.
- Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?. *Personality and social psychology review*, 8(4), 364-382.
- Robinson, E. (1928). *The behavior of the museum visitor*. Washington, DC: American Association of Museums.
- Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the Need for Closure Scale. *Personality and Individual Differences*, 50(1), 90-94.
- Rollins, M. (2001). The Invisible Content of Visual Art. *Journal of Aesthetics and Art Criticism*.
- Rothbaum, F., Weisz, J. R., & Snyder, S. S. (1982). Changing the world and changing the self: A two-process model of perceived control. *Journal of personality and social psychology*, 42(1), 5
- Russell, P. A., & Milne, S. (1997). Meaningfulness and hedonic value of paintings: Effects of titles. *Empirical Studies of the Arts*, 15(1), 61-73.
- Saito, M. (1996). Comparative studies on color preference in Japan and other Asian regions, with special emphasis on the preference for white. *Color Research & Application*, 21(1), 35-49
- Scherer, K. R. (2005). What are emotions? And how can they be measured?. *Social science information*, 44(4), 695-729
- Schwarz, N. (2011). Feelings-as-information theory. In P. Van Lange, A. Kruglanski, & E. T. Higgins (eds.), *Handbook of theories of social psychology* (pp. 289-308). Sage.
- Sibley, F. (1959). Aesthetic concepts. *The philosophical review*, 68(4), 421-450.
- Silvia, P. J. (2005). Emotional responses to art: From collation and arousal to cognition and emotion. *Review of general psychology*, 9(4), 342.
- Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, 3, 48–51.
- Silvia, P. J., & Nusbaum, E. C. (2011). On personality and piloerection: Individual differences in aesthetic chills and other unusual aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts*, 5(3), 208
- Spee, B., Ishizu, T., Leder, H., & Pelowski, M. (in press). Neuropsychopharmacological aesthetics: A theoretical consideration of pharmacological approaches to causative brain study in aesthetics and art. *Progress in Brain Research*.
- Speisman, J. C., Lazarus, R. S., Davison, L., & Mordkoff, A. M. (1964). Experimental analysis of a film used as a threatening stimulus. *Journal of Consulting Psychology*, 28(1), 23
- Summers, D. (1987). *The judgment of sense: Renaissance naturalism and the rise of aesthetics*. Cambridge: Cambridge University Press.
- Suzuki, M., Miyai, I., Ono, T., & Kubota, K. (2008). Activities in the frontal cortex and gait performance are modulated by preparation. An fNIRS study. *NeuroImage*, 39, 600-607.
- Swami, V., Stieger, S., Pietschnig, J., & Voracek, M. (2010). The disinterested play of thought: Individual differences and preference for surrealist motion pictures. *Personality and individual differences*, 48(7), 855-859.
- Tallis, R. (2008). The limitations of a neurological approach to art. *Lancet*, 372, 19–20.
- Tan, E. S. (2000). Emotion, art, and the humanities. *Handbook of emotions*, 2, 116-135.

- Tanaka, Y., Oyama, T., & Osgood, C. E. (1963). A cross-culture and cross-concept study of the generality of semantic spaces. *Journal of Memory and Language*, 2(5), 392.
- Tinio, P. P. L. (2013). From artistic creation to aesthetic reception: The Mirror model of art. *Psychology of Aesthetics, Creativity, and the Arts*, 7, 265–275.
- Tinio, P. P. L., Smith, J. K., & Smith, L. F. (2014). The walls do speak: Psychological aesthetics and the museum experience. In P. L. Tinio, J. K. Smith, & L. F. Smith (Eds.), *The Cambridge handbook of the psychology of aesthetics and the arts* (pp. 195–218). Cambridge, UK: Cambridge University Press.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*.
- Todorović, D. (2010). Context effects in visual perception and their explanations. *Review of Psychology*, 17(1), 17–32.
- Turner, A., & Penn, A. (2002). Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment. *Environment and planning B: Planning and Design*, 29(4), 473–490.
- Umiltà, M. A., Berchio, C., Sestito, M., Freedberg, D., & Gallese, V. (2012). Abstract art and cortical motor activation: An EEG study. *Frontiers in Human Neuroscience*, 6, 311.
- Van Den Abbeele, J., Penton-Voak, I. S., Attwood, A. S., Stephen, I. D., & Munafò, M. R. (2015). Increased facial attractiveness following moderate, but not high, alcohol consumption. *Alcohol and Alcoholism*, 50(3), 296–301.
- Wagner V, Menninghaus W, Hanich J, Jacobsen T (2014) Art schema effects on affective experience: the case of disgusting images. *Psychology of Aesthetics, Creativity, and the Arts* 8(2):120–129.
- Winston, A. S., & Cupchik, G. C. (1992). The evaluation of high art and popular art by naive and experienced viewers. *Visual Arts Research*, 1–14.
- Wolz, S. H., & Carbon, C. C. (2014). What's wrong with an art fake? Cognitive and emotional variables influenced by authenticity status of artworks. *Leonardo*, 47, 467–473.
- Zeglin Brand, P. (2000). Glaring omissions in traditional theories of art. in N. Carroll (Ed.), *Theories of Art Today* (p. 186), University of Wisconsin Press.

2.3.6

Tinio, P. P. L., & **Specker, E.** (in press). The method of observation. In M. Nadal, & O. Vartanian (Eds.), *The Oxford Handbook of Empirical Aesthetics*.

Summary:

In this chapter we describe the observation method of research and its features and uses. Starting from a general description of the method we then discuss examples of research within empirical aesthetics that use different types of observational methods: direct observation, indirect observation, and augmented observation. Throughout the paper we focus on the use of observational methods within an ecological context (i.e. the museum). Observational methods at its core are similar to classic field methods in social sciences such as ethnographic research. Because the focus of these observations is to obtain a near-exhaustive account of a phenomenon, subject, or event, and because the process is typically open-ended and emergent, this process requires a significant amount of time and resources. Observational methods can both be a rich source that contributes to theory and hypothesis formulation but also be a means to test hypotheses, and, finally, to augment other research methods. In the last 20 years, the method has had a resurgence after decades of emphasis on laboratory studies. In fact, Martindale (2007) described psychologists as having “fled” into their laboratories. Though laboratory work is extremely valuable, we argue that there are certain aspects of aesthetic experiences that are best examined when art is presented in more natural contexts. In the direct observation section we discuss work (e.g. Carbon, 2017; Pelowski, Liu, Palacios, & Akiba, 2014; Pelowski, 2012; Smith & Smith, 2001; Smith, Smith, & Tinio, 2017) where researchers directly observed people experiencing art and recorded their behaviors such as the amount of time spend looking at art, which types of art they look at, how they move through spaces such as galleries and museums. In the indirect observation section we focus on work that uses methods of indirect observation through technological advances such as movement trackers as well as physiological measures (e.g. Brieber, Nadal, Leder, & Rosenberg, 2014; Pelowski, 2018 Trondle, 2014; Tschacher et al., 2012; Yoshimura et al. 2014). In the augmented observation section we discuss our own previous work (Specker, Tinio, & Van Elk, 2016) as an example that allows discussion of how observation can be complimented with other methods in order to address one limitation, namely, the inability to tap into people’s thoughts, emotions, and other internal processes. We conclude by providing guidelines on how to conduct observations while also addressing issues such as reliability and validity as well as the challenges and limitations of the method. We propose that observation is useful when measuring overt behaviors but becomes questionable otherwise. In addition, though perhaps seemingly obvious the fact that observations occur through the lens of the observer means that it is needed to take steps to ensure that the process of observation is rigorous, systematic, detailed, and accurate, since the quality and validity of the data will otherwise be compromised (Cotton, Stokes, & Cotton, 2010).

My Contribution:

In this case, PPLT was responsible for the core idea of the chapter. My contribution mainly concerns the sections Direct Observation, Indirect Observation Using Advanced Technologies, and Augmented Observation for which I did the main underlying literature review and wrote the section. PPLT then integrated this with the rest of the chapter.

Observation Method in Empirical Aesthetics

Pablo P. L. Tinio

Montclair State University

Eva Specker

University of Vienna

Observation Method in Empirical Aesthetics

A museum visitor strolls into a museum gallery wearing audio tour headphones and holding a brochure with information about the exhibition. He walks up to the wall text and reads about the exhibition on display and makes mental notes of the details surrounding the works including the artists who made them, when they were made, and the art style to which they belong. The visitor proceeds to browse the individual pieces, briefly glancing at some and slowing down to look closely at others. He might stop by a specific work, read the label that accompanies it, and use the audio guide device to listen to information about the piece.

Across the gallery, a museum researcher observes this visitor. She watches his every move, making notes of which artworks he looks at and for how long he looks at them as well as whether or not he reads the labels accompanying the works. She might note details about the artworks that the visitor spent most time on and those that seemed to engage the visitor the most. The researcher might come up with questions about the visitor's aesthetic experience of the artworks, questions such as how the information from the audio guide, brochure, and wall texts influence the visitor's perception of, and engagement with, the artworks? Which work did the visitor look at the longest and what is it about that particular work that might have garnered so much attention? What motivated the visitor to go to the museum that day and what were his goals for his visit?

In the vignette above, both the visitor and researcher are using the most fundamental method for experiencing our world and for gathering information: observation. In empirical aesthetics research, as in the above example, observation occupies a dual role. On the one hand, an aesthetic experience is fundamentally a type of

observation—looking at artworks in a museum (e.g., Smith & Smith, 2001), interacting with design objects at home (e.g., Hekkert & Leder, 2008), or perhaps appreciating modern architecture (e.g., Vartanian et. al., 2013). On the other hand, the study of such experiences—the focus of this chapter and volume—involves some form of observation by a researcher, whether through direct observation of people’s behaviors during aesthetic encounters, recording what people say in such situations, and noting features of the environments in which the aesthetic encounters take place; or through indirect observation, such as using advanced technologies to record what people look at when interacting with aesthetic objects and how museum visitors move around and physically interact with artworks in a gallery. All of these approaches—direct or indirect— fall under the purview of the observation method. It is important to note that empiricism itself, as a grand theory, is defined by the very act of making observations, and as Adler and Adler (1994) have stated, observation is “the fundamental base of all research methods” (p. 389).

In this chapter, we describe the observation method of research and its features and uses. We also provide examples of empirical aesthetics research that have used observation as the primary method for data collection. We subsequently use these examples as a foundation for providing guidelines on how to conduct observations while simultaneously considering issues related to reliability and validity as well as the challenges and limitations associated with the method. The chapter will focus on the method of observation employed for the purpose of gathering information about people’s aesthetic experiences of stimuli such as artworks. In addition, although observation could be conducted both in the highly-controlled context of the laboratory and in more natural

contexts such as a museum, emphasis is placed on the use of observation in the latter context. Much of the discussion below is applicable to areas beyond aesthetics. However, the discussion and examples provided will focus on empirical aesthetics.

Uses of Observations

Observation, as the most basic method for conducting empirical research, is used for several specific purposes. First, observation could be used to simply describe an aesthetics phenomenon or set of phenomena. This mode of observation bears some resemblance to classic fieldwork methods in the social sciences, such as case studies and ethnographic research, both of which are used to provide rich and thick descriptions of a phenomenon, subject, or event (Creswell, 2014; Geertz, 1973; Holloway, 1997; Lincoln & Guba, 1985). An example of this type of observation is research on people's behaviors in museums, as illustrated in the opening vignette. The researcher stands at a distance and makes notes of art perceivers' behaviors, while being careful not to be obtrusive so as to preserve, as much as possible, people's natural behaviors. Another example is research on the creative art making process in which the researcher directly observes an artist at work, typically in the artist's studio. The researcher is immersed in the artist's creative process, witnessing the different stages involved in the creation of an artwork from the initial conceptualization of an idea to its eventual realization in a tangible object (Locher, 2010). Because the focus of these observations is to obtain a near-exhaustive account of a phenomenon, subject, or event, and because the process is typically open-ended and emergent (i.e., divergent and deductive), the process requires a significant amount of time and resources. However, the amount and the quality of data gathered could be substantial

and the insights gained could enhance not only the understanding of the creative art making process, but also the art viewing experience (Tinio, 2013).

Another important use of observation is to characterize or refine a research question and to identify key phenomena related to the question. For example, a researcher who is seeking to understand what types of artworks people are most likely to engage with might observe people at an art exhibition and then record which artworks in the exhibition are being looked at the most. The researcher could also speak with some of these visitors as they exit the gallery, asking them which artworks they looked at, how long and why they looked at them, and what they thought about their experience looking at the works. The information collected could then help to determine the features that are common to the artworks that were viewed the most. Perhaps it is the style of the works, their size, the subject that they depict (or do not depict in case of abstract works) and their placements within the physical space of the gallery. An account of the relationships amongst these factors could then be produced and additional research questions and corresponding hypotheses could be developed for future study.

Researchers also rely on observations to test specific hypotheses, a process related to yet another purpose of observations in research, which is to confirm, refute, or revise theory. Theories are often inspired by informal observations from everyday life (e.g., seeing visitors' reactions to an installation artwork) or by data collected in a formal study (e.g., a novel finding that existing theories are unable to explain). Through these observations, patterns of behaviors or occurrences could be found, which in turn could lead to the development of a formal theory that could be tested, revised, and then perhaps tested again, and so on. The iterative process involved in the development of theory is

one of the hallmarks of science. For example, a classic theory in empirical aesthetics, one that might explain why observers find some artworks more engaging than others, is Berlyne's (1971) *psychobiological theory*. A significant aspect of the theory suggests that the things that people find most interesting and pleasant are those that have moderate levels of complexity. According to the theory, the reason is that such stimuli elicit the optimal amount of arousal in viewers—neither too simple, nor too complex, but just the right amount of complexity to be interesting and pleasant. Berlyne and other researchers (e.g., Nadal, Munar, Marty, & Cela-Conde, 2010; Silvia, 2005) have further developed, refined, and adapted the psychobiological theory in light of new findings about the art experience as well as the development of new research technologies, such as those that allow direct measurements of physical arousal.

In addition to using observation as a primary method for data collection, it could also be used to augment other research methods (Adler & Adler, 1994), which typically occurs naturally in most studies: a researcher conducting experimental aesthetics research in the laboratory might observe participants' unusual or extreme overt emotional reactions to the presented stimuli; when conducting interviews, the interviewer could note the interviewee's nonverbal reactions and facial expressions when answering questions; when implementing a novel research procedure, the researcher might closely look for unforeseen reactions of participants or unintended consequences of the new procedure. Observation, therefore, has a multitude of uses, both as primary and secondary methods for data collection as well as for the development of theory.

Empirical Aesthetics Research and the Observation Method

There are myriad examples in the history of empirical aesthetics of the use of observations—whether to test a hypothesis or theory, refine a research question, or characterize an aesthetics phenomenon. The field has its roots in philosophical aesthetics, which is associated with introspective and observational methods for studying phenomena (e.g., Baumgarten, 1735/1954). Towards the end of the 19th century, these methods gave way to more experimental approaches to research (Fechner, 1876), although observation remained an important empirical method. In the last 20 years, the method has had a resurgence after decades of emphasis on laboratory studies. In fact, Martindale (2007) has described psychologists as having “fled” into their laboratories in which the modus operandi has involved the following elements: recruit participants (typically college undergraduates who agree to participate in research for course credit or monetary compensation); present (in random order) to the participants a set of images of artworks on a computer screen; have participants use rating scales to evaluate each artwork for how much they like or prefer it, find it beautiful or interesting, or the extent to which they perceive it as pleasant, complex, symmetrical, novel, surprising, or dynamic; and compile and statistically analyze data across all participants. Laboratory studies that have employed these procedures for the purpose of increasing experimental control have reached important findings and have arguably advanced empirical aesthetics more than any other methodological approach. However, there are certain aspects of the aesthetic experience that are best examined when art is presented in more natural contexts. These are the circumstances in which observations conducted in galleries, museums, and other natural settings are useful (McKechnie, 2008).

Below, we describe empirical aesthetics research in which the observation method was used. Although observations could also be used in different research contexts, such as in the laboratory, we focus on observations that involve researchers watching people engage in the viewing of art in a natural setting such as a gallery or museum. It is in these settings where the observation method is typically used as the primary method for collecting data instead of being used to merely augment another method. Using the observation method, the researcher collects first-hand recordings of the key aspects of the aesthetic experience of art, such as the characteristics and behaviors of the perceivers and the attributes of the physical context of the aesthetic encounter, including information about the artwork (through wall text, brochures, or art labels) that is provided to the perceivers. It is important to note that this chapter is not meant to be an exhaustive review of all studies that have used observations; it is instead meant to describe a few key studies that illustrate what is possible with observations, in what contexts they could be optimally used, how observations may be used alongside other research methods, and how the type of observation used could be tailored to address specific research questions.

Direct Observation

The first set of studies involves researchers directly observing people experiencing art and recording their behaviors, which could include the amount of time that they spend looking at art, what types of art they look at, how they move through spaces such as galleries and museums, and other behavior tendencies during aesthetic experiences.

Time Spent on Art

In a now classic study that has used observation in its simplest form, Smith and Smith (2001) examined how much time people spend looking at art in a museum. Museum professionals had a notion that museum visitors do not spend a lot of time looking at each artwork, even those that are considered to be masterpieces. To test this notion empirically, Smith and Smith selected six artworks from the Metropolitan Museum of Art in New York. The artworks were known masterpieces by Rembrandt, Cezanne, Raphael, Bierstadt, Leutze, and Heade. A total of 150 museum visitors were observed. For each visitor, the researchers tracked the total time that he or she spent looking at the artwork and reading the accompanying label. They also recorded the visitor's gender, estimated age, and if the visitor was viewing art alone or with other people. Results showed that people, on average, spent 27.2 seconds looking at an artwork (including time spent reading the label) with the median time spent being 17 seconds. Some paintings received longer viewing times than others with a range of 13.2 to 44.6 seconds. Smith and Smith also found an effect of group size, noting that people who visited with others spent more time looking at individual artworks than people visiting alone. No effects of gender or age were observed.

This study was replicated by Smith, Smith, and Tinio (2017) who observed visitors to the Art Institute of Chicago. They found an average viewing time of 28.63 seconds with a median viewing time of 21 seconds. As with Smith and Smith (2001), Smith et al. (2017) found no effects of gender and age and a small effect of group size, with people visiting with others spending more time looking at artworks. Importantly, this study separated out the time that viewers spent looking at the artwork itself and the time viewers spent looking at the accompanying label. In the original study by Smith and

Smith, both of these behaviors were included in the time spent with the artworks. Smith et al. found that 246 out of a total sample of 456 museum visitors read the accompanying labels. The visitors who read the labels spent, on average, about as much time reading the label (15.79 seconds) as they did viewing the artworks (17.26). Finally, during the observations, the researchers were struck by museum visitors taking *selfie* pictures with the artworks, which the authors then appropriately named *arties*. Noting this as a particularly frequent behavior after the first day of observation the authors decided to track this behavior for the remaining observation period. They found that 35% of museum visitors snapped arties with the artworks on display.

The two studies above (Smith & Smith, 2001; Smith, Smith, & Tinio, 2017) demonstrate the potential of observations for uncovering behaviors fundamental to the aesthetic experience. These researchers used a very direct and straightforward technique: stand beside an artwork, wait for a visitor to engage with the work, and measure the duration of the engagement while recording perceiver characteristics as well as other notable behaviors such as reading labels and taking photographs.

Recently, Carbon (2017) used a similar approach as Smith and Smith (2001) to study the behaviors of museum visitors to a Gerhard Richter exhibition at the Neues Museum in Nürnberg. In Smith and Smith as well as in Smith et al. (2017), only the viewing time of the first initial viewings of artworks were recorded. Carbon expanded on this method by examining repeated viewings of the same artworks by the same person during one visit. He found that people on average spent 32.9 seconds (median = 25.1 seconds) looking at individual artworks, a result that is relatively similar to those found by Smith and Smith. Furthermore, Carbon found that some paintings received longer

viewing times than others with a range of 25.7 to 41.0 seconds. Interestingly, more than half of the visitors (55.3%) returned to look at an artwork more than once, resulting in an average total viewing time of 50.5 seconds. It is worth noting that the initial viewing of artworks that were later viewed again were viewed for a shorter time (mean = 25.8 seconds) than artworks that were viewed only once (mean = 37.9 seconds). Carbon suggested that visitors may have realized during their visit that they looked at some artworks too briefly and therefore returned to look at them again. Another explanation may be that people initially wanted to spend more time looking at an artwork, but that their viewing may have been constrained by a number of factors. For example, people who are with others in a museum might feel pressured to stay with the group and forgo a closer, slower look at an artwork in order to move with the group to another artwork. Another possible explanation is that other visitors may block the path leading to, or line of sight towards, a particular work. In each of these situations, the amount of time that a visitor spends looking at an artwork would decrease and the motivation to want to return to a particular work later in the visit would increase. A carefully conducted observation, as Carbon had done, allows such interpretations to be made.

Physical Movement and Locomotion

The observation method also allows the measurement of people's location and movements in natural settings. Determining people's physical behaviors in a particular space is difficult, if not impossible, with other common methods of research such as surveys, questionnaires, and interviews because such methods depend heavily on memory of where one has been, of the exact path that one took to get to a location, of how long one spent at particular locations, and of how one has physically interacted with aesthetic

objects. Observations allow minute-to-minute, if not, second-to-second, recording of such behaviors. For example, in addition to measuring the amount of time people spent looking at artworks, Carbon (2017) also observed museum visitors' viewing distances from the artworks. Inspired by Locher, Smith, and Smith (2001) who reported that viewers in a museum had viewing distances from artworks ranging from 60cm to 120cm with the lower limit being the distance viewers had to the smallest artwork, Carbon's observations of viewing distances showed a range of 149cm to 212cm with a mean of 175cm. His observation data also showed a direct relationship between viewing distance and the size of an artwork: the bigger the artwork, the farther the distance. For tackling such a research question regarding viewing distance, observation is the optimal method.

Similarly, regarding physical location and movement in museums, Pelowski, Liu, Palacios, and Akiba (2014) used observation as a method to better understand previous findings (Pelowski, 2012) that found that people in the Rothko room in the Kawamura Memorial DIC Museum of Art in Japan had less engaged aesthetic experiences as compared to visitors to the Rothko Chapel in Houston, USA, and to a room of Rothko artworks at the Tate Modern in London, UK. The researchers postulated that aesthetic experience may have been inhibited by the architectural features of the Kawamura room in which the space made it more difficult to avoid standing in another visitor's line of sight. Pelowski et al. observed 65 visitors to the Rothko Room in Kawamura and found that people indeed explicitly tried to avoid being in another person's field of vision, a finding that suggests the impact of social aspects of art viewing on the aesthetic experience of art. Following their initial observations, the Kawamura room was remodeled in order to eliminate the cramped viewing conditions and to create a space

where physical movement and viewing were less restricted. Pelowski et al. returned to the Rothko room after it was remodeled and found evidence that art viewers had more profound aesthetic experiences than in the pre-remodeled room.

The Art-making Process

Another use of direct observation worth mentioning is the study of the creative art making process, an area that aesthetics scholars have also delved into as a means of gaining additional insight into the aesthetic experience (e.g., Arnheim, 1954/1974, 1962, 1969; Gombrich, 1960). Recently, the field of empirical aesthetics has witnessed a push towards characterizing the relationship between art making and aesthetic experiences (Specker, Tinio, & van Elk, 2017; Tinio, 2013), for example, by determining whether the emotional characteristics of artworks (i.e., their emotional affordance) can be experienced by the perceiver of the work. In other words, is there a correspondence between the emotional characteristics of a work that have resulted from specific steps that the artist took during the art making process and art perceivers' aesthetic experiences of the artwork (Tinio & Gartus, 2018). Observation is one method for gaining insight into the art making process. An important characteristic of this type of observation is direct observation of an artist at work in the natural context of a studio for the purpose of obtaining a detailed description of the art making process, which includes the materials and methods used by the artist as well as the making of photographic records of the artwork during different stages of its creation (Locher, 2010).

Indirect Observation Using Advanced Technologies

Various technologies could be used to conduct indirect observations. One of the strengths of this approach is that it allows for a more objective (i.e., a way of

circumventing human bias and observation error) and unobtrusive measure of behaviors during aesthetic encounters. This could be accomplished through the use of technologies that, for example, record where people look and where they move within a given space as well as physiological measures that could be used to further shed light into aesthetic experiences. The following section describes research that have used such technologies to good effect.

An example of indirect observation research was conducted by Tschacher et al. (2012) who observed 517 visitors to the Kunstmuseum St. Gallen, Switzerland, during a period of several months. The visitors wore electronic gloves that contained a transponder that enabled precise recordings of visitors' movement patterns within the museum. The gloves also measured skin conductance level, skin conductance variability, heart rate, and heart rate variability. This technology allowed for indirect observations of visitors' behaviors in an unobtrusive and natural manner with visitors able to determine how they spent their time at the museum and to freely choose what artworks they looked at and how long they looked at them. The research participants were visitors of the "11:1 (+3) = Eleven Collections for a Museum" exhibition that was curated specifically for the study. A total of 76 contemporary and modern paintings and sculptures comprised the exhibition. In addition to the objective measures, visitors also completed an aesthetics questionnaire that assessed the emotions conveyed by the artworks (e.g., joy, anger, sadness, and surprise) and the artworks' aesthetic qualities (e.g., beauty, artistic merit) as well as visitors' general aesthetic evaluations of the works (e.g., whether the works belonged in the gallery, how famous was the artist?). Tschacher et al. found that skin conductance variability, heart rate, and heart rate variability correlated highly with

participants' responses regarding the artworks' aesthetic quality, surprise/humor, dominance, and curatorial quality.

The same glove technology used by Tschacher et al. (2012) was used by Trondle (2014) to further observe people's movements within a museum space. Detailed recordings were made of the physical movements of 50 visitors to the Kunstmuseum St. Gallen. Trondle found that specific features of the museum space influenced visitors' movements within the museum galleries as well as which artworks they looked at and for how long they looked at them. Through the use of the transponder technology to unobtrusively observe art perceivers' movements, Trondle was able to discover what he termed *space-cells*, or particular areas within the museum that encouraged very specific responses, such as slowing down and stopping to look at an artwork and corresponding label. Furthermore, an *orientation point* marks the beginning of a space-cell by orienting visitors' attention to a salient feature, such as a wall text introducing the artworks within a particular gallery. The totality of space-cells in a museum represents the general pattern of behavior exhibited by the visitors. Indirect observation, such as that used by Trondle, allowed for a characterization of art perceivers' behaviors that would have been extremely difficult using other methods.

Another technology, the Bluetooth proximity sensor, was used by Yoshimura et al. (2014) to indirectly observe visitor movements at the Louvre Museum. The researchers examined the flow of visitors from the entrance of the museum through the gallery spaces with specific attention to the volume of visitors within areas of the galleries and their sequence of movements as well as the length of time that they spent at the museum. Unlike the studies described above, such as those by Trondle (2014) and

Tschacher et al. (2012), Yoshimura et al. focused entirely on visitor movements and did not gather other, more cognitive data, an approach that truly characterizes indirect observation. The Bluetooth technology also did not require visitors to carry or wear a tracking device, but instead, their movements were recorded by sensors placed in key locations around the museum. This method was highly unobtrusive with visitors not being aware that their physical behaviors were being recorded. In total, data from over 20,000 visitors during a period of one month were analyzed.

Yoshimura et al. (2014) were able to determine, with extreme detail, the amount of time that each visitor spent at the museum, which ranged from less than one hour to 15 hours, with the majority of visitors staying between 4-6 hours. Regarding visitors' movements in the museum, the researchers found something surprising: the amount of time that visitors spent at the museum did not make a significant difference on how much of the museum these visitors explored. Both short-stay and long-stay visitors tended to visit the same areas of the museum that are typically popular to visitors. This is a notable finding given the vastness of the Louvre Museum as well as the size of its collection of art. Yoshimura et al.'s use of Bluetooth technology was an optimal method for indirectly observing the physical behaviors of museum visitors. The Bluetooth method is objective, precise, and unobtrusive as well as efficient in the sense that it enabled data gathering at a large scale for an extended period of time.

Another method for conducting indirect observations makes use of technology that records where people look (e.g, Brieber, Nadal, Leder, & Rosenberg, 2014). This allows the measurement of what objects people look at, how and how long they look at them, and, with the use of video recording of their visual attention, their movements and

other physical behaviors within a given space. An example of research that used this technology was conducted by Pelowski et al. (2018) who explored art perceivers' experience of installation art. The study was the first to systematically and empirically examine the aesthetic experience of this type of art. The two works under examination were part of the exhibition, "Baroque Baroque!" by artist Olafur Eliasson and were installed at the Belvedere Museum, Austria. By their very nature, installation artworks involve an interplay between the installed features and the environment in which they are installed. Pelowski et al. examined the interplay between elements of Eliasson's works and the Baroque interior of the Museum's galleries: whether visitors engaged with the installed elements independent of their surroundings or whether they engaged with both aspects. The researchers found evidence for the latter with people visually attending to both artworks and the gallery space in which they were installed. Interestingly, for both artworks, people looked more at (i.e., more and longer eye fixations) aspects of the rooms than elements of the artworks. One of the artworks, *Wishes versus Wonders*, which consisted of a tall mirror that subdivided a room that had landscape and battlefield scenes painted on its walls. Pelowski et al. found evidence that people were highly engaged with the paintings and that visitors might have integrated the painted scenes into their experience of the art. Indirect observation using eye tracking and video recording technologies, as Pelowski et al. had done, allows for rigorous and detailed accounts of people's behavior during aesthetic encounters with art.

Augmented Observation

The observation method could be augmented with other methods in order to address one of its limitations, namely, the ability to tap into people's thoughts, emotions,

and other internal processes. An example of a study that made use of augmented observations was conducted by Specker et al. (2017). They observed people interact with artworks in a museum and in a lab. In both contexts, they used a think-aloud procedure where research participants verbalized their thoughts aloud while viewing William Gropper's *the Senate* (c.1950), a painting on exhibit at the Queens Museum in New York. Participants used an audio recording device to record their verbalized thoughts. The researchers also took notes of other relevant behaviors such as how visitors physically looked at the painting as well as if they read the museum label or spoke to other visitors while looking at the artwork. The use of the think-aloud procedure allowed the researchers to gain knowledge about what the visitors were thinking about the painting. Observations, combined with the think-aloud procedure, allowed access to visitors' thoughts about the artwork while being less obtrusive to visitors' aesthetic experience of the artwork. Specker et al. performed a similar study in the lab by presenting a reproduction of the same Gropper painting.

Specker et al.'s (2017) use of augmented observation led to important findings, such as evidence that art perceivers' aesthetic engagement with the Gropper painting largely followed the information processing stages described in models of the aesthetic experience (Chatterjee, 2003; Leder, Belke, Oeberst, & Augustin, 2004; Leder & Nadal, 2014; Pelowski, Markey, Forster, Gerger, & Leder, 2017; Tinio, 2013). The study also found that art perceivers generally had a strong motivation to derive meaning from the work, and they often described very specific emotional reactions to the piece, stating that the painting made them feel depressed or that it had a sense of eeriness. Augmented

observation, in this sense, was used to empirically test various theories of the aesthetic experience of art as well as to characterize the nuances associated with this experience.

Conducting Observations

The studies described in the previous sections illustrate the strengths of the observation method for empirically examining aesthetic experiences. As with all empirical methods, observation has practical, logistical, and psychometric properties—such as validity and reliability—that must be considered for the method to be effective (see Fendler, 2016 for a comprehensive overview of these properties). The following issues must be taken into account when using the observation method: *appropriateness*—is the method appropriate for the research question being asked and is the research designed in a way that allows the benefits of the method to come through?; *comprehensiveness* and *reliability*—are the observations, whether performed solely by a researcher and/or augmented by advanced technology, performed in a comprehensive, objective, systematic, rigorous, detailed, and accurate manner?; *validity*—is the interpretation (and where appropriate, the *application*) of the results consistent with the data that are collected (Creswell, 2014)? The next section provides guidelines for the use of the observation method, and the studies above will be used to illustrate important concepts.

Appropriateness

As fundamental and powerful as the observation method is, care must be taken to ensure that it can effectively capture the data needed to answer the research question or to test a specific hypothesis (Silverman, 1993). Observation is optimal for measuring overt behaviors to answer research questions, such as which artworks within an exhibition

appear to engage people the most as indicated by the number of people who approach the works and look at them as well as whether they read labels accompanying the works and discuss the works with others. Observation would also be appropriate for determining which of a number of design objects—smartphones, for instance—would people interact with the most when given the simple task of using the objects. It is important to note that these observations involve overt behaviors and directly address the research question of aesthetic engagement with visual stimuli of art or design objects.

If used beyond measuring overt behaviors, the appropriateness of the observation method becomes questionable. Observation is less effective at determining people's cognitions (how much they find a painting interesting, beautiful, or intellectually challenging), emotional reactions (if an installation artwork made them feel anxious, uncomfortable, or nostalgic), and knowledge and experiences (how much formal art education people have received and to what extent this influences their aesthetic experiences). Although observations could be augmented with other methods to gain insight into internal processes, such as Specker et al.'s (2017) use of the think-aloud procedure, the ultimate strength of observations involve what can be directly seen.

Comprehensiveness and Reliability

The primary purpose of the observation method is to describe, and according to Patton (2002), must have the following:

“Observational data must have depth and detail. The data must be descriptive—sufficiently descriptive that the reader can understand what occurred and how it occurred. The observer's notes become the eyes, ears, and perceptual senses for

the reader. The descriptions must be factual, accurate, and thorough without being cluttered by irrelevant minutiae and trivia” (Patton, 2002, p. 23).

Observations are conducted through the lens of the observer. This may sound like a truism, but if steps are not taken to ensure that the process of observing is rigorous, systematic, detailed, and accurate, the quality and validity of the data will be compromised (Cotton, Stokes, & Cotton, 2010). The observer’s own lens may lead to bias in what he or she actually observes, which would limit a study’s comprehensiveness—the thorough sampling of the range of behaviors of interest including the minutiae of what people do and say. It could also include observing, in detail, the context in which the behaviors occur. Comprehensiveness also entails including the appropriate range of time during which the sampling of behavior occurs. It could involve seconds and minutes as in Smith and Smith’s (2001) study of the amount of time people spend looking at paintings, or weeks and months as in Yoshimura et al.’s (2014) study of people’s movements in a museum.

In addition to the need for observations to be comprehensive, they must also be reliable. There must be little variability in how observations are made and recorded. Accuracy is key. The exact behaviors of participants, their expression, and other features that are associated with them must be described in accurate detail (e.g., Creswell, 1998; Geertz, 1973; Holloway, 1997; Tjora, 2006), and such accuracy should apply to each and every research participant. Observations must also be consistent across different observers who are tasked with observing the same aesthetic phenomena; in other words, inter-rater reliability must be high (Creswell, 2014).

Reliability and objectivity are critical to ensuring the validity of observations, yet they are two of the most challenging issues to address. Observers are fallible. They are prone to mistakes and observer bias—seeing things in a manner that is consistent with their perspectives and beliefs and that may confirm their expectations, or hypotheses, for the study (e.g., Connor-Greene, 2007). Observations are typically recorded using checklists, rubrics, observational protocols, and often, rough field notes and short narratives that are quickly written as the events and behaviors are taking place. None of these recording methods is immune to inaccuracies. A rigorous approach to recording observations is therefore needed to minimize errors and ensure the quality of data (Creswell, 1998). Because of the complexities associated with observations, it is important for observers to be trained in the following skills (adapted from Patton, 2002):

- Noticing what there is to see and hear
- Knowing what to pay attention to and what to ignore
- Recording observations systematically, accurately, and descriptively
- Recognizing and acknowledging the limitations of one's observations and potential biases
- Minimizing any influence that one might have on the research context and participants' behaviors

One way to address issues of reliability is through the use of technology. Observations made by a person could be augmented with technology, or as in studies such as that conducted by Yoshimura et al. (2014), observations could be solely technology-based. Whether it is video recording people engaging in a certain behavior or performing a task or using wearable devices that track locations and movements within a

space, appropriate technology use could significantly increase the objectivity and reliability of observation data, which in turn would help increase the validity of the data.

Validity

Validity, within the context of observations, refers to the extent to which the observations made are appropriately interpreted and applied. Therefore, valid interpretations of observations are only possible when the observations themselves are highly objective and reliable. Once the conditions of objectivity and reliability are established at a sufficient level (which together signify the quality of the data and the strength of evidence to support predictions or to address the research questions), the next step would be to ensure that the interpretation of the results is consistent with the observation data and that any generalizations drawn from the study are appropriate, which requires careful determination of the representativeness of the data (Booth, Colomb, & Williams, 2008).

One important point to consider when interpreting and generalizing the results is the possibility that the observer's mere presence in the research context has influenced the events and behaviors of the participants under investigation (Silverman, 1985). Although the negative impact resulting from the presence of the observer could be mitigated by taking the necessary steps to be as inconspicuous as possible (e.g., blending in), such impact can never be fully ruled out. In presenting the results of an observation study, at the minimum, one must acknowledge the potential impact of the observer and to consider such impact when generalizing the results.

Conclusion

The observation method is arguably the most fundamental of all research methods (Adler & Adler, 1994). It could be used as a sole approach for data collection or to augment other methods (e.g., interviews and laboratory experiments) as well as to develop and refine theory. With these uses come some costs, such as the challenges associated with designing observation research. For the observation method to be effective, it must be appropriate for the research question being addressed. The observations themselves must be rigorously and systematically performed as well as comprehensive and reliable. The interpretation of the data must also be valid. Empirical aesthetics has depended on observations since its inception as a field of study (Fechner, 1876). Even with the continuous development and use of advanced technologies for conducting aesthetics research, researchers will always depend on the observation method as the most basic, yet one of the most powerful, methods for empirically examining aesthetics phenomena.

References

- Adler, P. A., & Adler, P. (1994). Observational techniques. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 377-392). Thousand Oaks, CA: Sage.
- Arnheim, R. (1954/1974). *Art and visual perception*. Berkeley, CA: University of California Press.
- Arnheim, R. (1962). *The genesis of a painting. Picasso's Guernica*. Berkeley, CA: University of California Press.
- Arnheim, R. (1969). *Visual thinking*. Berkeley, CA: University of California Press.
- Baumgarten, A. G. (1735/1954). *Reflections on poetry*. Berkeley, CA: University of California Press.
- Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.
- Booth, W. C., Colomb, G. G., Williams, J. M. (2008). *The craft of research*. Chicago, IL: University of Chicago Press.
- Brieber, D., Nadal, M., Leder, H., & Rosenberg, R. (2014). Art in time and space: Context modulates the relation between art experience and viewing time. *PLoS ONE*, 9, e99019.
- Carbon, C.-C. (2017). Art perception in the museum: How we spend time and space in art exhibitions. *I-Perception*, 8, 204166951769418.
- Chatterjee, A. (2003). Prospects for a cognitive neuroscience of visual aesthetics. *Bulletin of Psychology and the Arts*, 4, 55-60.

- Connor-Greene, P. A. (2007). Observation or interpretation? Demonstrating unintentional subjectivity and interpretive variance. *Teaching of Psychology*, 34, 167-171.
- Cotton, D. R. E., Stokes, A., & Cotton, P. A. (2010). Using observational methods to research the student experience. *Journal of Geography in Higher Education*, 34, 463-473.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed Methods approaches*. Thousand Oaks, CA: Sage.
- Fechner, G. T. (1876). *Vorschule der Ästhetik [Primary School of Aesthetics]*. Leipzig: Breitkopf & Härtel.
- Fendler, L. (2016). Ethical implications of validity-vs.-reliability trade-offs in educational research. *Ethics and Education*, 11, 214-229.
- Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.
- Gombrich, E. H. (1960). *Art and illusion: A study in the psychology of pictorial representation*. Princeton, NJ: Princeton University Press.
- Hekkert, P. & Leder, H. (2008). Product aesthetics. In R. Schifferstein & P. Hekkert (Eds.). *Product experience* (pp. 259-286). San Diego, CA: Elsevier.
- Holloway, I. (1997). *Basic concepts for qualitative research*. London: Blackwell Science.
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95, 489-508.
- 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.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Locher, P. J. (2010). How does a visual artist create an artwork? In J. Kaufman & R. Sternberg (Eds.), *Cambridge handbook of creativity* (pp. 131-144). Cambridge, UK: Cambridge University Press.
- Locher, P., Smith, J. K., & Smith, L. F. (2001). The influence of presentation format and viewer training in the visual arts on the perception of pictorial and aesthetic qualities of paintings. *Perception*, 30, 449–465.
- Martindale, C. (2007). Recent trends in the psychological study of aesthetics, creativity, and the arts. *Empirical Studies of the Arts*, 25, 121–141.
- McKechnie, L. E. F. (2008). Observational research. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 574–577). Thousand Oaks, CA: Sage.
- Nadal, M., Munar, E., Marty, G., & Cela-conde, C. J. (2010). Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, 28, 173-191.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.
- Pelowski, M. (2012). (Not) myspace: Social interaction as detriment to cognitive processing and aesthetic experience in the museum of art. In *Intelligent interactive multimedia: Systems and Services* (pp. 399-410). Berlin, Heidelberg: Springer.

- Pelowski, M., Liu, T., Palacios, V., & Akiba, F. (2014). When a body meets a body: An exploration of the negative impact of social interactions on museum experiences of art. *International Journal of Education & the Arts*, 15. Retrieved from <https://eric.ed.gov/?id=EJ1045933>
- Pelowski, M., Leder, H., Mitschke, V., Specker, E., Tinio, P. P. L., Gerger, G., Vaporova, E., Bieg, T., Husslein-Arco, A. (2018). Capturing aesthetic experiences with Installation Art: An empirical assessment of emotion, valuations, and mobile eye tracking in Olafur Eliasson's 'Baroque, Baroque!' *Frontiers in Psychology*, 9.
- 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.
- Silverman D (1985) *Qualitative methodology and sociology: Describing the social world*. Aldershot: Gower Publishing Co.
- Silverman D (1993) *Interpreting qualitative data: Methods for analysing talk, text and interaction*. London: SAGE.
- Silvia, P. J. (2005). Emotional responses to art: From collation and arousal to cognition and emotion. *Review of General Psychology*, 9, 342–357.
- Smith, J. K., & Smith, L. F. (2001). Spending time on art. *Empirical Studies of the Arts*, 19, 229–236.
- Smith, L. F., Smith, J. K., & Tinio, P. P. L. (2017). Time spent viewing art and reading labels. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 77–85.

- Specker, E., Tinio, P. P. L., & van Elk, M. (2017). Do you see what I see? An investigation of the aesthetic experience in the laboratory and museum. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 265–275.
- Tinio, P. P. L. (2013). From artistic creation to aesthetic reception: The mirror model of art. *Psychology of Aesthetics, Creativity, and the Arts*, 7, 265–275.
- Tinio, P. P. L., & Gartus, A. (2018). Characterizing the emotional response to art beyond pleasure: Correspondence between the emotional characteristics of artworks and viewers' emotional responses. *Progress in Brain Research*, 237, 319-342.
- Tjora, A. H. (2006). Writing small discoveries: an exploration of fresh observers' observations. *Qualitative Research*, 6, 429-451.
- Trondle, M. (2014). Space, movement and attention: Affordances of the museum environment. *International Journal of Arts Management*, 17, 4-17.
- Tschacher, W., S. Greenwood, V. Kirchberg, S. Wintzerith, K. van den Berg, and M.Trondle. (2012). Physiological correlates of aesthetic perception of artworks in a museum. *Psychology of Aesthetics, Creativity, and the Arts*, 6, 96-103.
- Vartanian, O., Navarrete, G., Chatterjee, A., Fich, L. B., Leder, H., Modroño, C., Nadal, M., Rostrup, N., & Skov, M. (2013). Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture. *PNAS*, 110, 1-8.
- Yoshimura, Y., Sobolevsky, S., Ratti, C., Girardin, F., Carrascal, J. P., Blat, J., & Sinatra, R. (2014). An analysis of visitors' behavior in the Louvre Museum: A study using Bluetooth data. *Environment and Planning B: Planning and Design*, 41, 1113-1131.

3. Discussion

My PhD dissertation (*“Psychological Aesthetic Effects of Lines and Colors: An Interdisciplinary Inquiry”*) consists of three main sections: (1) Basic Psychological Research, (2) Implementation within the project, (3) Field Studies. Since all papers include a specific general discussion of the findings I do not reiterate these here, instead I aim to illustrate what can be learned from the thesis as a whole.

The articles belonging to section (1) laid the groundwork to be further explored in section (2) and (3). The first paper, Specker et al. (2018; section 2.1.1), shows the development of the VAIK to adequately measure art interest and art knowledge. This paper made significant contributions to the field of empirical aesthetics by not only providing clear theoretical arguments for disentangling art interest and art knowledge and thereby adding to the discussion of our conceptualization of “expertise” but also providing methodological advancement by providing a new and validated tool to measure art interest and art knowledge. This played an essential role in our future studies by allowing us to assess these constructs adequately which we used in Specker et al. (*under review*, section 2.2.1), Grüner, Specker, and Leder (2019, section 2.3.3) and Specker, Stamkou, Pelowski, and Leder (*under review*, section 2.3.4). In addition, Grüner, Specker, and Leder (2019, section 2.3.3), and Specker et al. (*under review*, section 2.3.4) provided important data on the validity of the questionnaire outside of the lab. Even though the time frame of the PhD did not allow for further exploration of this in the form of a paper, follow-up research on the validity of the VAIK within field settings (and other conditions influencing measurement such as language) is ongoing. Thus these studies provide data that will be essential to further establishing the output of the project even if this was not the main goal of the individual studies. Finally, Specker et al. (2018; section 2.1.1) provides a psychological and measurement perspective towards the notion of expertise in art or sensitivity to art. This served as a basis to also

investigate this notion from an art historical perspective in the forthcoming (Boddy et al., *in prep*; section 2.2.4). The other two papers in this section Specker et al. (2018, section 2.1.2) in Specker and Leder (2018, section 2.1.3) showed evidence for an association between brightness and positivity that holds implicitly and cross-culturally. These findings became essential to our forthcoming paper Specker, Fried, Rosenberg, and Leder (*under review*; section 2.2.2.). Within this paper, the results obtained from this previous work could be used to establish a suitable manipulation to test our statistical models (see section 1.2 and 2.2.2.). As such they could provide a link between elementary research on one association with a complex network system explored in Specker et al. (*under review*; section 2.2.2.).

The second section presents the basic core of the project and aims to assess the universality of aesthetic effects. The paper by Specker et al. (*under review*; section 2.2.1) can be seen as the core publication within the project. This paper relied on Specker et al. (2018; section 2.1.1) in order to assess boundary conditions of universality (in this case art knowledge/interest). In order to do this empirical work it was needed to provide the necessary art historical background which was achieved by Brinkmann et al. (2018, section 2.2.3). This paper provided not only a theoretical basis on which we could base our choice of terms (i.e. aesthetic effects) to be used as dependent variables but also preliminary experimental research on which we could base our question format (bipolar scale). In Specker et al. (*under review*; section 2.2.1) we found significantly lower agreement in the effects of artworks and their single elements than expected. For the whole artworks, participants agreed on the effects of warm-cold, heavy-light, and happy-sad, but not on the 11 other scales. Further, we found that the image type (artwork or its constituting elements) was a major factor influencing agreement. Art expertise did not play a significant role and agreement was low on dimensions usually of interest in empirical aesthetics (e.g., like-dislike). These results pose a challenge to the practice of interpreting artworks based on their aesthetic effects, as these effects may not be as universal as previously thought. This work thus provided the basis for a discussion of

these findings from an art historical perspective in Boddy et al. (*in prep*; section 2.2.4). In addition, the Boddy et al. (*in prep*; section 2.2.4) paper represents an art historical follow-up on the Brinkmann et al. (2018, section 2.2.3) paper as well as an extension by investigating the notion of expertise in or sensitivity to art from a historical perspective in correspondence to Specker et al. (2018; section 2.1.1) which provides a psychological and measurement perspective. These results then inspired for the Specker et al. (*under review*; section 2.2.2.) paper which presents an extension on theory of aesthetic effects from a methodological perspective. It takes aesthetic effects and the experience of them as a specific cognitive process present in aesthetic experience. It proposes an aesthetic effect network (AEN) where the experience of aesthetic effects is seen as an associative process where having one association leads to the next association. Through a network of associations, this generates an overall aesthetic experience. Conceptualizing aesthetic experience (at least in part) as an associative process provides an explicit account of a specific cognitive process involved in aesthetic experience and that extends beyond research on aesthetic effects, that can be applied to aesthetic experience in general, and provides a basis for future studies. This facilitates our understanding of why aesthetic experience is both somehow different and also somehow the same for all of us and is thus intimately paired with the differences found in experience in Specker et al. (*under review*; section 2.2.1). The AEN model does not exclude network structures differing among people which potentially could explain the findings of Specker et al. (*under review*; section 2.2.1). Furthermore, it might be that network structures are shared among people, but that at the same time, how an artwork influences the network differs among people which would also fit with the findings of Specker et al. (*under review*; section 2.2.1). Though further research would be needed to explore these theoretical ideas. In addition, Specker et al. (*under review*; section 2.2.2.), as noted, built on the results of Specker et al. (2018, section 2.1.2) in Specker and Leder (2018, section 2.1.3) to create a suitable manipulation which allowed us to differentiate our AEN model from other potential models.

Finally, the perspective laid out in this paper applies to research in aesthetic experience in general including field research. As such, I have already conducted some exploratory work with application of networks of aesthetic experience within a field setting (Pelowski et al., 2018; section 2.3.1) which showed potentially different structures for different artworks. Since the application of networks to aesthetics is in its infancy, the Specker et al. (*under review*; section 2.2.2.) paper can provide a strong theoretical basis to guide future studies.

The third section, though not a main focus of the project, addresses the need to assess people in ecologically valid settings to understand their aesthetic experiences that arises from doing research from an interdisciplinary perspective. As noted, Pelowski et al. (2018; section 2.3.1) includes an exploratory network analysis which is related to my theoretical work in Specker et al. (*under review*; section 2.2.2.). Showing both similarities and differences in network structure that may shape our aesthetic experiences and needs to be explored further in future research. Grüner, Specker, and Leder (2019) allowed for the testing of the VAIK (Specker et al., 2018; section 2.1.1) within a field setting (the same goes for Specker et al, *under review*; section 2.3.4), see above. Finally, Pelowski and Specker (*in press*) takes a wide perspective on context and also considers context in the form of learning or education making it correspond to Specker et al. (2018, section 2.1.1.) and through its inclusion of a historical perspective also connects to Boddy et al. (*in prep*, section 2.2.4). Together, they represent a comprehensive account of the role of art knowledge and art interest within aesthetic experience incorporating theoretical, methodological, historical, and art historical perspectives. Grüner, Specker, and Leder (2019; section 2.3.3) and Pelowski, Specker, Gerger, Leder, and Weingarten (2018; section 2.3.2) allowed for an extension of previous work together with Tinio & Specker (*in press*). Tinio and Specker (*in press*) is based both on own previous work (Specker, Tinio, & Van Elk, 2017) as well as work done within the thesis notably Pelowski et al. (2018; section 2.3.1). Showing a variety of approaches to studying art in the museum.

To summarize, the three sections of my thesis include various research projects with different foci. Nonetheless, all parts can be used to build on each other – and can complement each other. As such, the thesis represents a comprehensive mixed-method approach to empirical aesthetics with a broad perspective, it is truly interdisciplinary, and, I hope, highlights the need to study art from different perspectives and with a variety of methods since only then can we come to a more comprehensive understanding.

4. English Abstract

My PhD dissertation (“Psychological Aesthetic Effects of Lines and Colors: An Interdisciplinary Inquiry”) consists of three main sections: (1) Basic Psychological Research, (2) Implementation within the project, (3) Field Studies. The studies in (1) laid the groundwork for (2) and (3), and concerned testing associations between brightness and positivity as well as the development of a questionnaire to measure art interest and art knowledge. The studies in (2) represent two art historical papers on the history of aesthetic effects and two empirical papers. Within the empirical papers we tested the notion of universality of aesthetic effects while accounting for possible differences in sensitivity of viewers (in terms of art interest and knowledge), and we provided a new theoretical account of aesthetic effects. To substantiate this new theoretical model we varied the brightness of artworks to change artwork ratings which allowed us to differentiate between our models and potential other accounts. The studies in (3) concerned ecologically valid tests of aesthetic experiences within a museum context. The three sections include various research projects with different foci. Nonetheless, all parts can be used to build on each other – and can complement each other. As such, the thesis represents a comprehensive mixed-method approach to empirical aesthetics with a broad perspective, it is truly interdisciplinary, and aims to highlight the need to study art from different perspectives and with a variety of methods since only then can we come to a more comprehensive understanding.

5. German Abstract

Diese Dissertation (“Psychological Aesthetic Effects of Lines and Colors: An Interdisciplinary Inquiry”) ist in drei zentrale Abschnitte unterteilt: (1) Psychologische Grundlagenforschung, (2) Implementierung innerhalb des Projekts und (3) Feldstudien. Die Studien in (1) legten den Grundstein für (2) und (3), und bezogen sich einerseits auf die Testung von Assoziationen zwischen den Konzepten Helligkeit und Positivität und andererseits auf die Entwicklung eines Fragebogens zur Messung von Kunstinteresse und

Kunstwissen. Die Studien in (2) beinhalten zwei kunstgeschichtliche Artikel zur Geschichte von ästhetischen Effekten und zwei empirische Studien. In den empirischen Studien untersuchten wir einerseits die Idee einer Universalität von ästhetischen Effekten bei gleichzeitiger Beachtung von möglichen Unterschieden der Betrachtenden in ihrer Sensitivität (in Bezug auf Kunstinteresse und –wissen), und andererseits lieferten wir ein neues theoretisches Modell zum Verständnis von ästhetischen Effekten. Zur Untermauerung dieses theoretischen Modells variierten wir die Helligkeit der Kunstwerke, um dadurch die Bewertungen zu beeinflussen. Dies erlaubte uns zwischen unserem Modell und anderen Erklärungsalternativen zu unterscheiden. Die Studien in (3) beschäftigten sich mit der ökologisch validen Testung von ästhetischen Erfahrungen in einem Museumskontext. Alle drei Abschnitte beinhalten eine Anzahl an Forschungsprojekten mit verschiedenen Schwerpunkten. Dennoch bauen die Abschnitte aufeinander auf und ergänzen sich gegenseitig. Als solches stellt die vorliegende Dissertation einen umfassenden mixed-methods Zugang zur empirischen Ästhetik mit einer breiten Perspektive dar. Sie ist klar interdisziplinär und zielt darauf ab das Bedürfnis nach der Untersuchung von Kunst aus mehreren Perspektiven hervor zu streichen und dies auch anhand verschiedenster methodischer Zugänge zu zeigen. Letztendlich können wir nur dadurch zu einem tieferen Verständnis gelangen.