

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Empirical Investigations on Art Viewing Behavior and Art Evaluation:
An Interdisciplinary and Cross-Cultural Study“

verfasst von / submitted by

Jan Mikuni

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Doktor der Philosophie (Dr.phil.)

Wien, 2021 / Vienna, 2021

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:

Doktoratsstudium
Sozialwissenschaften: Psychologie

Betreut von / Supervisor:

Univ. Prof. Dr. Helmut Leder
Univ. Prof. Dr. Raphael Rosenberg

博士論文 令和 3(2021)年度

Empirical Investigations on Art Viewing Behavior and Art Evaluation:
An Interdisciplinary and Cross-Cultural Study

慶應義塾大学大学院社会学研究科

三國 珠杏

Preface

This dissertation is approved as the first jointly supervised doctoral thesis (Cotutelle de Thèse) between Keio University (Tokyo, Japan) and the University of Vienna (Vienna, Austria). Consequently, my dissertation (“Empirical Investigations on Art Viewing Behavior and Art Evaluation: An Interdisciplinary and Cross-Cultural Study”) has been developed in the close collaboration between the two universities. In accordance with the agreement of the jointly supervised doctoral thesis, this project has been supervised by Univ.-Prof. Dr. Hideaki Kawabata (Keio University, the Faculty of Letters, Department of Psychology), Univ.-Prof. Dr. Helmut Leder (the University of Vienna, Faculty of Psychology, the Department of Cognition, Emotion, and Methods in Psychology), and Univ.-Prof. Dr. Raphael Rosenberg (the University of Vienna, Faculty of Historical and Cultural Studies, the Department of Art History).

This collaboration between two universities makes this dissertation an example of the joint research, wherein the experts of neuroaesthetics (Univ.-Prof. Dr. Hideaki Kawabata), of art psychological research and theories of aesthetics (Univ.-Prof. Dr. Helmut Leder), and of art history (Univ.-Prof. Dr. Raphael Rosenberg) collaborated with the Ph.D. candidate, Jan Mikuni, to answer questions relevant for ‘empirical aesthetics’, museum visitor research, and for art history.

It is important to note that the contents and the format of the dissertation must fulfil the requirements from both universities. Consequently, this dissertation is written as a monographic thesis, and only contains the scientific work where the Ph.D. candidate acted as (one of) the first author(s). Additional scientific work where the Ph.D. candidate acted as one of the co-authors, is included in *Additional Achievements* in a cumulative form to demonstrate her full research history during her Ph.D.

A lot of people have been involved to make this ever first collaboration possible. Here, I would like to thank all supervisors, my colleagues, and university staffs for their help and support.

Contents

1. Introduction	8
1.1 What is empirical aesthetics?	9
1.2. A gap in the experience of art between lab-based and real-life settings	11
1.3. The skewed participants' demographics included in the past research	15
1.4. An overview and the structure of the present dissertation.....	21
2. Study 1—Translating well-documented attention decrease phenomenon towards art from museum studies into lab-based study	26
2.1. Study 1-1: Cross-cultural replication study with an updated research design for examining the repeated art viewing effect in a lab.....	39
2.1.1. Method.....	40
2.1.2. Results	42
2.1.3. Discussion.....	49
2.2. Study 1-2	51
2.2.1. Method.....	51
2.2.2. Results	52
2.2.3. Discussion.....	61
2.3. General Discussion and Conclusion of Study 1	62
3. Study 2—Empirical investigation on the possible causes and how to counteract the attention decrease phenomenon in lab-based setting.....	67
3.1. Study 2-1: Examining impact of information about artworks on attention towards them	68

3.1.1. Method.....	72
3.1.2. Results	77
3.1.3. Discussion.....	83
3.2. Study 2-2: Examining one of the possible causes of the decreased attention towards artworks, ' <i>habituation</i> '	85
3.2.1. Method.....	92
3.2.2. Results	95
3.2.3 Discussion of the pilot study	100
3.3. Main Study	103
3.3.1. Method.....	105
3.3.2. Results	111
3.3.3. Discussion of Study 2-2.....	119
3.4. General discussion of Study 2	122
4. Study 3—Cross-cultural similarity/difference in aesthetic evaluation and art viewing.....	124
4.1. Study 3-1: the universality of the preference for symmetry	125
4.1.1. Method.....	130
4.1.2. Results	134
4.1.3. Discussion.....	139
4.2. Study 3-2: The differences in the concept of beauty across culture.....	142
4.2.1. Method.....	144
4.2.2. Results	148
4.2.3. Discussion.....	159
4.3. Study 3-3: the cross-cultural diversity in eye movement patterns while viewing art between two cultures	165

4.3.1. Method.....	168
4.3.2. Results	171
4.3.3. Discussion.....	182
4.4. General discussion for Study 3	185
5. General Discussion	187
6. References	195
7. English Abstract	212
8. German Abstract.....	213
9. Additional Achievements	215
9.1. 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, p. 47-53.	217
9.2. Pelowski, M., Cabbai, G., Brinkmann, H., Mikuni, J., Hegelmaier, L. M., Forster, M., Rosenberg, R., & Leder, H. (2020). The kitsch switch—or (when) do experts dislike Thomas Kinkade art? A study of time-based evaluation changes in top-down versus bottom-up assessment. <i>Psychology of Aesthetics, Creativity, and the Arts</i>	225

1. Introduction

This dissertation project, an interdisciplinary and cross-cultural joint-supervised dissertation between Keio University (Japan) and the University of Vienna (Austria), presents the scientific work performed by the Ph.D. candidate, Jan Mikuni. The dissertation constitutes three main empirical sections: (1) Translating the well-documented attention decrease phenomenon towards artworks from museum studies into lab-based settings, (2) Empirically investigating how to counteract the attention decrease phenomenon in lab-based settings, and (3) Analyzing cross-cultural similarities/differences in art viewing behavior and evaluation between native Japanese and German viewers.

Although the individual studies in this dissertation deal with/use quite diverse topics/methodologies, they share a common goal—to tackle the two limitations in empirical aesthetics, namely (1) A gap in the art experience between real-life and lab-based settings and (2) the skewed participant demographics in past research. In the following general introduction, I will first give a brief overview of the two general limitations in the field to provide proper context for this dissertation. The aim of the general introduction is to show how the present dissertation can contribute significantly to the field of empirical aesthetics. I will then present three main sections of the dissertation where the individual studies are discussed in detail. Since this dissertation does not deal with one specific phenomenon/topic, the detailed descriptions of the background/introduction of each study, its purpose, and hypothesis are presented respectively in each section. In the general discussion of each section, all results in the given section are discussed together. Finally, in the general discussion, it is discussed how the studies across three sections relate to/complete each other, the limitations of the studies in this dissertation, and possible avenues for future studies.

1.1 What is empirical aesthetics?

As a foundation to the empirical parts of this dissertation, I would like to clarify how empirical aesthetics is defined in the present thesis and to present a brief overview of the research field. The latest definition of empirical aesthetics can be found in *The Oxford Handbook of Empirical Aesthetics* published by Oxford University Press in 2021. In the handbook, empirical aesthetics is defined as “a scientific field that uses empirical methods to study aesthetics” (Nadal & Vartanian, 2021, p.3). Seeing this definition, one might question what *aesthetics* actually mean. The handbook states that “Aesthetics is a concept that means different things to different people” (Nadal & Vartanian, 2021, p.7). Here, it might be helpful to view the definitions of *aesthetics* from a philosophical perspective. Levinson (2003) stated in his article *Philosophical Aesthetics: An Overview*, “Aesthetics is the branch of philosophy devoted to the conceptual and theoretical inquiry into art and aesthetic experience (p.1)” that has three main foci, namely (1) “a certain kind of *practice* or *activity* or *object*—the practice of art or the activities of making and appreciating art,” (2) “a certain kind of *property*, *feature*, or *aspect* of things—one that is aesthetic, such as beauty or grace or dynamism” ,and (3) “a certain kind of *attitude*, *perception*, or *experience*—one that can also be labelled aesthetic.”

Back to the definition of empirical aesthetics, we can thus surmise that “Empirical aesthetics is the scientific field where researchers empirically investigate art, as well as perceptual features and experiences that bring aesthetic values.” While this definition does sound rather ubiquitous and ambiguous, empirical aesthetics *is* indeed a broad scientific field dealing with various types of stimuli, stages of neural/cognitive processes, and layers of behaviors. For example, art can be presented in various formats. Paintings, sculpture, music, dance, and cuisine can be all recognized as art. Furthermore, since what we perceive as aesthetic features is diverse, we can find aesthetic features/properties in faces, nature, sounds, textures, movements, tastes, and, of course, in visual artworks like paintings, sculptures, and designs. Finally, the level of *aesthetic* attitude/perception/experience also has multiple layers ranging from bottom-up information

processes like perception of hue, satiation, and contrast in one painting to more top-down information processes like peripheral information about the art such as if the artist is famous or not (Mastandrea & Crano, 2018) and/or social/peer bias when seeing the objects (e.g., Leder et al., 2004).

Considering the diversity in the field of empirical aesthetics, it might be helpful to clarify the focus of this dissertation. This dissertation is concerned with the interaction between visual stimuli (especially paintings/geometric images) and its perceivers. More specifically, I will focus on art viewing behavior and evaluation (especially the appraisal of beauty) of its viewers while they interact with art, i.e., how the attention of art viewers changes over time and what kind of features in paintings people find beautiful etc. Thus, faces, music, and/or natural scenes will not be used as stimuli in the main part of the dissertation. Furthermore, since this dissertation has been developed under the supervision of experts in the field of neuro/cognitive aesthetics, empirical aesthetics, museum studies, and art history in Japan and Austria, this dissertation contains interdisciplinary and cross-cultural components, such as rigorous investigation of past findings from museum visitor studies and examination of the impact of culture on art viewing behavior and evaluation. Based on these premises, in the following introduction, I will briefly introduce past findings and two general limitations in the field of empirical aesthetics.

Though the first investigations into *aesthetics* in the field of psychology can be traced back to 1876 when Gustav Theodor Fechner published the foundational text, *Vorschule der Aesthetik* (Fechner, 1876; see Nadal, 2021 for a recent review), empirical aesthetics showed a burgeoning growth after the year 2000 because, “owing to a number of drives: technological progress improved the conditions under which art could be studied in the laboratory, neuroscience discovered the arts as an area of interests, and new theories offered a more comprehensive look at aesthetic experiences” (Leder & Nadal, 2014, p.443). Now, we know that visual art provokes complex cognitive and physiological behaviors, and it requires the integration of memories, meaning making, diverse and potentially intense emotions, as well as general psychological processes such

as expectation, perception, and judgments. As suggested by representative art information processing models, these can be matched to numerous stages of cognitive processes (Pelowski et al., 2016 for review). As a result of the complex processes and reactions evoked by art, diverse research questions, methodological approaches, as well as further theoretical models and frameworks of art experience have been developed (e.g., Carbon, 2011; Chatterjee, 2003; Leder et al., 2004; Leder & Nadal, 2014; Locher et al., 2010; Pelowski et al., 2016; Pelowski & Akiba, 2011; Silvia, 2005; Tinio, 2013). However, while the field has shown rapid growth, scholars have started to point out crucial limitations in many past studies of empirical aesthetics: (1) a gap between empirical lab-based research and actual experiences of art in a real-life setting (e.g., Carbon, 2019; Makin, 2017; Pelowski et al., 2017), and (2) skewed participant samples—past studies in the field have been exclusively conducted either in North American or European countries (e.g., Che et al., 2018). In the following section, I will go through each limitation in detail.

1.2. A gap in the experience of art between lab-based and real-life settings

Most studies in the field of empirical aesthetics, regardless of the target phenomenon or metrics, have been conducted in the context of experimental laboratories (Carbon, 2020). This is because “laboratory settings provide ideal conditions for setting up experiments in a standardized and rigorous way” (Carbon, 2020, p.1). Investigations with well-controlled conditions in a laboratory might be especially crucial when researchers are trying to examine the interaction between specific artworks and its perceivers. As described above, since numerous factors/cognitive processes constitute the experience of art (e.g., Leder et al., 2004; Leder & Nadal, 2014), it is important to ensure that we assess the factors of interest while controlling for confounding factors, which is almost impossible to achieve in real-life settings like museums, galleries, or public spaces.

Nevertheless, past studies have demonstrated that people experience and appreciate artworks differently in a real-life setting from what they do in a laboratory. In other words, the context (Mastandrea et al., 2021), or the space in which artworks are presented in a broad sense (Brieber et al., 2014), can have a significant impact on our experience of art (Brieber et al., 2014; Brieber et al., 2015; Locher et al., 1999, 2001; Muth et al., 2017; Specker et al., 2017). Brieber et al. (2015) compared the appraisals of artworks (i.e., arousal, valence, liking, interest, and understanding) and the recall test results of the artworks between two distinct viewing contexts (museums vs. laboratory). All participants viewed the same artworks in the same order both in a museum and in a laboratory; which context the participants experienced first was counter-balanced across participants. The authors found that the physical context where participants viewed the artworks had a significant impact on their experiences of art. In the museum context, artworks were evaluated as more arousing, positive, interesting, and better liked than in the lab context. Furthermore, they found that participants who joined the museum context after the laboratory context could recall the artworks better and thus concluded that the ‘*real*’ art experience in an actual museum context enhanced the experience of art.

Following this pioneering work, scholars have shown an increasing awareness of the gap between experience of art in real-life and laboratory-based settings¹. To the best of my knowledge, Pelowski et al. (2017) conducted the first comprehensive review of key differences in the experiences of art in museums and laboratories, which might influence the experience of art. In museums where we typically interact with artworks, we make “a conscious decision to travel to such a setting, pay an entry fee, and proceed from room to room while perusing a collection of art

¹ Though I do acknowledge a more recent review by Kubovy (2020) criticizing the typical empirical aesthetic study’s approach, it mainly overviews and criticizes basic problems, especially in neuroaesthetics. Since this dissertation only includes behavioral experiments without using any method/metrics for neural activities such as EEGs, NIRs, and fMIRs, the detailed literature review of that article will not be included here.

and stopping and looking at those pieces that draw attention, all the while being surrounded by other visitors typically within a rather opulent building” (Pelowski et al., 2017, p. 246). However, the experience of art in a laboratory setting is quite different. “Participants [in a laboratory] ... may talk briefly with a researcher and are often told that they will see ‘some art’ or participate in an art-related task. They are then seated at a chair before a screen and are shown an image for a few (typically less than 5) seconds, followed by an intervening pause, cue, and then the next image. A task is often involved, such as making an evaluation. What results is a parade of digitalized art images presented without apparent reason or interconnection” (Pelowski et al., 2017, p. 246). In the same year, Makin (2017) also criticized the ‘psychological approach’ used in empirical aesthetic research, stating that since “the paradigmatic aesthetic experience involves strong emotions, fascination, and aesthetic rapture, it is nearly impossible to evoke these feelings in the lab, especially on multiple trials with well controlled stimuli.” (Makin, 2017, p.206).

Two years later, Carbon (2019) published a review article titled *Empirical Approaches to Studying Art Experience*. In this article, he recognized empirical testing as a powerful tool to examine the causality of the effect(s) of target factors, but criticized the decomposition approach in empirical studies which divided a whole phenomenon into measurable parts, for the following three reasons: First, this decomposition approach might measure something other than the actual experience i.e. the effect/experience from a whole phenomenon might differ from that of only part of the phenomenon. Second, this approach creates an illusion that *epiphanizing*—a concept which was introduced in (Pelowski & Akiba, 2011) to describe the elaborated experience of artworks—is a process which can be repeatedly/systematically varied. Last, rigorous decompositions in a lab cannot investigate the actual effects of the original artworks and contexts, e.g., the physical context (Brieber et al., 2014; Brieber et al., 2015; Locher et al., 2001; Muth et al., 2017; Specker et al., 2017), frame of the artworks (e.g., Ensor & Hamilton, 2014; Koutstaal, 1998), or with whom we view the artworks (e.g., Chang, 2006) are not the same in a laboratory and in a museum.

Carbon (2019) also summarized the three paths to gain knowledge from art experiences which most of the empirical studies follow, as well as the pros and cons of each path: Path #1 is “the ecological path to testing in the art gallery/museum”; Path #2 is “the path where gallery qualities are simulated to provide contextual embedment”; and Path #3 is “the context-free lab-oriented path” (Carbon, 2019, p. 3). Path #2 seems to be a way to bridge Path #1 and #3 by, though not fully, simulating the contextual information of a real-life setting: “The main aim is to simulate typical conditions in and with which artworks are epiphanized, while preserving the power and possibilities of an experimental approach” (Carbon, 2019, p.3). Presumably, studies which use recent technologies such as virtual reality (VR) and/or high-resolution images/monitors to mimic the high-quality visual images/information in a laboratory setting can be considered a Path #2 study. However, although it is possible that those technologies might be able to replicate the museum settings more accurately than a typical lab-based empirical setting, it is still difficult to perfectly replicate a museum environment. For example, the expectations/motivations for viewing art probably will still be largely different.

More recent individual empirical studies have discussed how to overcome the gap between ecologically valid and lab-based settings. For example, Weigand and Jacobsen (2021) recognized both the pros and cons of ecologically valid and lab-based settings, and further suggested that to capture, examine, and understand the complex experiences of art, “a multi-method approach of field study and laboratory studies where methodological strengths and limitations complement each other in order to see if results converge” (p. 12) might fill the gap in future studies. As reviewed here, the natural differences between the two distinct art viewing contexts (museum vs. laboratory) cannot be easily solved. Thus, as suggested by Weigand and Jacobsen (2021), a multi-method approach might be a possible way to fill the gap e.g., conducting a well-controlled empirical studies in a lab-based setting to test the effects of interesting factors, and then conducting a replication study in an ecologically valid setting like a museum, or checking if the similar effect is present in both museum and laboratory, and then conducting more rigorous investigations in a lab. Though

this approach will take extra time and effort, it will ensure that both the museum and lab context do not lose any of their strengths and compensate for each other's limitations.

In the present dissertation, through Study 1 and 2, I will present how the studies in the two distinct contexts can complement each other's findings. More specifically, I will focus on a well-documented phenomenon from museum studies, which is decreased attention over repeated art viewing. Though this phenomenon was first documented over a century ago (Gilman, 1916) in an actual museum exhibition setting, why it happens as well as how to counteract it still remains unclear due to the difficulties in controlling for confounding variables in museum exhibition settings. As such, in Study 1, I tested to see if I could recreate this phenomenon in a lab setting with multiple participants in both Japan and Austria (Study 1). Then I further examined the underlying mechanisms as well as how to counteract this phenomenon in a lab setting (Study 2), which needs to be assessed rigorously but has not yet been achieved in past museum studies. Here, by translating and examining the findings from museum research into a lab-based setting, I aim to overcome the shortcomings of past museum studies with lab-based studies.

1.3. The skewed participants' demographics included in the past research

Art creation and appreciation is universal. In this sense, art/aesthetics seem to be human universals (Anderson, 2004; Silver, 1979). In past studies of empirical aesthetics, it seems that researchers implicitly held one premise: "If all humans share a common capacity for aesthetic appreciation, it seems plausible that the underlying psychological and brain processes are also common to all humans" (Che et al., 2018, p.1). Presumably, this implicit assumption led to the current situation where past studies in the field almost exclusively investigated North American and European participants. The problem here is not that researchers did not conduct cross-cultural comparisons per se, but rather that they concluded and assumed that the findings from their skewed samplings (Western university students) could be generalized to all people universally without

empirical evidence examining the potential effect of cultural background. This problem was extensively discussed at length in the seminal article by Henrich et al. (2010). They described the participants in past social scientific studies, which are not limited to the empirical aesthetics, as WEIRD (Western, Educated, Industrialized, Rich, and Democratic) people and criticized the skewed demographics of their participants. In addition, they showed that culture does influence a variety of cognitive domains (e.g., in the perception of visual illusions) and thus universality cannot be assumed without empirical investigation. The situation also holds true in the field of empirical aesthetics. For example, recent models/frameworks of cognitive information processes in art experience are based on findings mainly derived from WEIRD people, yet they all implicitly claim universality (see Che et al., 2018 for a review). Consequently, examining the effects of culture on art experience is a key challenge against any implicit universality assumption in past studies.

It is important to note that I do acknowledge that there were several studies assessing the universality/cross-cultural differences in art appreciation especially from the 50s to 70s. However, the motivation of these attempts was “to vindicate and promote a normative view of aesthetic appreciation whereby people could be ranked according to their ability to correctly judge the beauty or artistic merit of objects... to classify people based on their innate abilities and to direct them toward careers suited to those abilities” (Che et al., 2018, p.2). Thus, cross-cultural studies that try to challenge the implicit assumption of universality in past studies with clear hypotheses based on cultural differences in certain aspects are still quite limited. In this dissertation, I recognize the implicit assumption of universality in past empirical aesthetic research as the second general limitation and aim to consider/assess, as a starting point, if the effect of culture is present and what kind of differences in culture might impact art experience. To highlight the importance of considering the effect of culture on aesthetics, I will present in the following section some examples of inter-cultural differences in everyday life/art as well as several theories and studies suggesting the potential impact of culture on aesthetics.

Contrary to the implicit assumption of scholars, as a person who grew up in an East Asian country (Japan), worked in a Western country (Austria) and is now presenting this dissertation to the reader as part of a cross-cultural collaborative effort, I find the small yet accumulated differences in aesthetics due to the differences in culture in everyday life as quite obvious. While I can list numerous examples from my personal experiences, the most surprising one is this: while “*your face is small and really white*” can be a huge compliment in Japan, Korea, and China, it is apparently not as big a compliment in Austria and presumably in other European countries, which really confused me at the beginning. This personal example shows a potential cultural influence in the aesthetics of faces, which can have an impact on such a biologically relevant aspect as facial perception (which can influence mate choice among others).

Not only in the facial components to be valued, but also visual art itself (e.g., the materials, subjects, forms, and development of art) obviously varies across cultures. Taking one of the most famous and traditional Ukiyo-e in Japan, “The Great Wave off Kanagawa” made by Katsushika Hokusai (Figure 1) as an example, we can see the many differences compared to the traditional paintings of other countries e.g., how the artist depicts scenes, the used colors, how the people are depicted etc. The waves are depicted with deep blues which are of a similar color as Mt. Fuji in this picture, which stands in striking contrast against the white foam of the water as well as the snow covering the mountain. Mt. Fuji and the people are depicted quite small, and the structure of the screen/faces of the people are drawn quite flat compared to the precise perspective method, which translates the fear of the waves in the people who are confronting it very well. The “flat” feature on the screen in Japanese art is quite unique, and it is the core concept of a modern art movement termed ‘*super flat*’, characterized by a flat structure on the screen with lots of empty space without using precise and fine perspective technique in traditional artworks, anime/manga/character culture, and pop-culture in Japan, by a Japanese artist, Takashi Murakami.



Figure 1. “The Great Wave off Kanagawa,” Katsushika Hokusai, 1831. The picture was retrieved from Wikiart (<https://www.wikiart.org/en/katsushika-hokusai/the-great-wave-of-kanagawa-1831>). , which has unlimited use e.g., copying, distributing, and displaying of the images in the public domain of artworks with permission from the website.

In fact, in the field of art history, the differences in art forms/material/historical development have been a major research topic, with some researchers arguing that the differences can also be connected to how we process our visual field. For example, Thausing (1983) argued that historic/stylistic changes in art might stem from the history of changes in perception. This idea—the interconnection between development of art and perception—was further explored in other theories in art history. For instance, the concept of ‘*period eye*’ was introduced by Baxandall (1988) in his book *Painting and Experience in Fifteenth-Century Italy*. Baxandall argued that though humans share innate skills to perceive the visual world, we all develop varieties of *cognitive*

styles. Here, *cognitive styles* are described as follows: “There are many other ways of perceiving it [the configuration in plate 13] as well. That which we tend toward will depend on many things—particularly on the configuration, which is suppressed here for the moment—but not least on the interpreting skills one happens to possess, the categories, the model patterns and the habits of inference and analogy: in short, what we may call one’s *cognitive styles*” (Baxandall, 1988, p. 29-30). Baxandall claimed that the differences in *cognitive styles* are derived from differences in experience such as education, social norms, language, and exposure to different visual forms implemented in each culture. Further, he also claimed that the cultural background might have influenced what visual features attract people during a particular time span, which thus implies that culture can indeed influence the perception/cognition (of art). Considering that what we see and are exposed to in our everyday lives like alphabets, sceneries, buildings, dishes, and artworks are culturally different, Baxandall’s assumption on the potential effect of culture on aesthetics is worth empirically investigating, since classic ‘*mere-exposure effect*’ (Zajonc, 1968) also showed that subjective experiences of visual stimuli can be modulated by the number of exposures—typically showing a positive correlations between the evaluation and the number of exposures.

Though the arguments from art historians remain theoretical, more recent psychological studies have provided some evidence to support their theories. The most important findings were presumably a series of studies reported by Richard Nisbett and his colleagues. In 2001, they published a model for cross-cultural differences in cognitive processing where they focused on two distinct cultures: East Asian and Western. They argued that East Asians tend to have holistic thoughts, while the Westerns have the more analytic ones (Nisbett et al., 2001). Note that although I used the term *East Asian and Western cultures* as in the original study, it is questionable and might paint a wrong picture of the results when we generalize the findings without testing each culture in *East Asian* and *Western* culture, since there might be a large difference in many aspects e.g., social system, language, sceneries obviously differ within *East Asian/Western* cultures. Thus,

in order to generalize the findings across a broad range in East Asian and/or Western culture, multiple replications in each culture are necessary to conclude generalizability.

Though they did not specify what differences in culture led to the differences in cognitive styles, they speculated that the social system in the distinct cultures were the cause of the differences. Later on, the research team empirically tested the model and accumulated evidence to support their hypothesis. For example, Masuda and Nisbett (2001) recruited participants from Japan and America and presented animated vignettes wherein a variety of fish were swimming in an aquarium. The participants were asked to watch the videos, and after they finished watching, they were further asked to describe “What did you see in the animation? Please describe it, taking as much as 2 min” (Masuda & Nisbett, 2001, p.925). Interestingly, the descriptions from the Japanese participants contained more contextual information compared to the American participants (e.g., color of water, background plants and corals etc.). A similar experiment was further conducted by Chua et al. (2005). In their study, they presented scenery pictures to their participants (25 European American students and 27 international Chinese students) at the University of Michigan and measured their eye movement patterns while they viewed the stimuli. The results showed that European American participants fixated on focal objects and the foreground information of the stimuli longer than the Chinese students. In contrast, the Chinese students fixated more on the contextual information of the image (e.g., background of the image) longer than European American students. These results clearly suggest that the cultural background does impact their perceptual/cognitive processes, and that there exists some perceptual/cognitive process patterns within a culture. Though the aforementioned studies did not use artwork as their stimuli, the results suggest a general impact of cultural background when people visually explore the environment, which can of course be applied to visual artworks as well.

In the present dissertation, the effect of culture on art experience was assessed in Study 1 and 3, but mainly in Study 3. It is important to note that, although it is hard to exactly identify culture as one single aspect, given that religions, languages, social norms, political systems, and

lifestyles can all be recognized as culture. Nevertheless, in the present dissertation, I will focus on the differences in *visual culture* i.e. what we see and are visually exposed to in our everyday lives, such as the form of artworks, alphabets, and/or reading/writing systems when constructing any hypotheses.

Study 1, where I translated the attention decrease phenomenon to a lab-based setting, was conducted both in Japan and Austria to examine if there are any inter-cultural differences in the attention decrease phenomenon. The cross-cultural component was motivated, considering that the actual population of museum visitors are not only from Western cultures, but also from more diverse cultures. This insight will be important to generalize the findings into museum practice as well. Consequently, the cross-cultural component was included without any clear hypothesis not to study the impact of culture on art experience, but rather to examine the generalizability of the specific effects across cultures.

In contrast, Study 3 focuses more on challenging the universality claim in aesthetics via three empirical studies. Specifically, I tested, firstly, the beauty appraisals of symmetry/asymmetry objects from native Japanese speakers, secondly, art attributes constituting beauty appraisals between native Japanese and German speakers, and lastly, eye movement patterns while native Japanese and German speakers viewed non-art and art images.

1.4. An overview and the structure of the present dissertation

The present dissertation consists of three main empirical sections (Study 1, 2, and 3) that tackle either one of the two limitations. In this section, I will briefly present the main aim of each study and how each section/study contributes to the field of empirical aesthetics. Do note that the detailed rationale of each study will again be presented separately in each section.

Study 1 and 2 are concerned with the gap in the experience of art between real-life and lab-based settings. In these two studies, I will focus on the well-documented attention decrease

phenomenon from repeated art viewing reported from past museum studies. Even though it has been more than 100 years since this phenomenon was reported for the first time (Gilman, 1916), only a limited number of studies have examined why this happens and how to counteract this unwanted phenomenon. A major obstacle for studying diminishing attention over repeated viewing of objects in museums can be found in the nature of the study environments in past studies. Specifically, almost all past studies on this topic have been conducted in actual museums where multiple confounding factors created difficult-to-control conditions (e.g., Bitgood et al., 1986; Falk et al., 1985; Gilman, 1916; Johnston, 1998; Marcellini & Jenssen, 1988; Matamoros, 1986; Melton, 1935; Porter, 1938; Robinson, 1928). Accordingly, scholars have insisted on the need for rigorous empirical tests to understand this phenomenon and to counteract it (Stephen Bitgood, 2009a; Gareth Davey, 2005; Pelowski, Forster, et al., 2017). This is in line with the arguments in Chapter 1.2 where I introduced the pros and cons of research in museum and laboratory settings. Through Study 1, I translated the past findings from museum studies into a laboratory setting to test the feasibility of studying this phenomenon in a lab-based setting with participants from both Japan and Austria. Through Study 2, I tried to identify what factor would cause this phenomenon and how to counteract it. It is quite important not only for accumulating knowledge on this phenomenon but also for museum/exhibition practices to consider how to counteract this phenomenon, since decreased attention over time might lead to some negative consequences. For example, past studies have pointed out that it may interrupt visitor satisfaction and the effective learning process (Bitgood, 2002; Falk & Storksdieck, 2005; Falk, 1983; Falk & Dierking, 2002; Hein, 1998; Serrell, 1997; see Davey, 2005 for a review). Thereby, the two foci, namely, to detect the cause and to develop countermeasures, were especially assessed in the present dissertation. The approach chosen in Study 1 and 2 might be key to bridging the gap between the museum and the lab while still complementing each other's findings. Although there have been several studies on both art viewing contexts (e.g., Brieber et al., 2014, 2015; Locher et al., 1999, 2001; Muth et al., 2017; Specker et al., 2017), they mostly aim to find differences in the art experience between two different art

viewing contexts i.e. how people experience art differently in different physical contexts. Thereby, studies where lab-based and museum findings complement each other were rarely conducted, e.g., examining the findings from laboratory studies in a real-life setting to validate the lab-based findings or empirically investigating the effect reported from museum studies in a laboratory to observe the factors of interest under a well-controlled condition. Study 1 and 2 in this dissertation can be seen as examples of the latter case. Specifically, the lab-based studies were designed to overcome the limitations of past museum research: we needed to control the confounding factors in a lab-based setting in order to assess the effect of factors of interest which might have caused the decreased attention phenomenon. This approach is expected to serve as an example for the field by showing how to combine the findings from different settings. Moreover, the findings of Study 1 and 2 are also expected to help improve actual museum designs/displays and the satisfactions of the visitors. In 2012, the most visited museum in the world (the Louvre in Paris) welcomed 9.7 million visitors and made 216 million Euros in revenue. More recently in 2018, the museum broke their own record with a record 10.2 million visitors that year (Stein, 2018). Given the number of people who visit museums and interact with art every single day, revealing the mechanism of the decreased attention phenomenon and a way to prevent it will not only contribute to the field of Empirical Aesthetics, but also to the museums as a whole, directly leading to improvements in exhibition design, visitor satisfaction, and return rates.

Study 1 and 3 will address the second limitation: The skewed participant demographics in the past research. As mentioned, past studies in Empirical Aesthetics have mostly focused on WEIRD people while assuming universality in art experience without enough empirical support. In Study 1, where I aimed to translate a well-documented phenomenon in museum studies into a lab-based setting, participants from both Japan and Austria were included to see if the art viewers from different cultures would behave similarly or not. Though the cross-cultural component in Study 1 is rather passive, i.e. though we did not have clear hypotheses on what different cultural aspects

would impact decreased attention, this attempt could be insightful for museums (considering the diversity in their visitors) interested in the inter-cultural generalizability of the results.

Study 3 further investigated the impact of culture on different types of aesthetics beyond the narrow focus on the decreased attention phenomenon. In Study 3, three experiments are going to be presented, all of which have different motivations and purposes. In the first study, the preference for symmetrical visual features—which are recognized as an important and rather universal feature of preference/aesthetic judgements and have frequently been observed in European participants—were assessed with Japanese participants. Since asymmetry is well-known as a beauty principle in Japanese aesthetics, we directly challenged the claim of universality regarding the preference/values of symmetry with Japanese participants. In the second study, the evaluations of beauty in artworks between native Japanese and German speaking participants were directly compared. Here, the art attributes which were commonly used in past studies on empirical aesthetics was listed and I assessed what kind of attributes were evaluated as more beautiful in each culture and if there were any systematic differences in the features that were attributed a judgement of beauty between the two cultures. Though this study is rather explorative, the second study is expected to paint a clearer picture of the systematic effect of culture on art evaluation vis-à-vis attributes and cultural differences. The third study will assess cultural diversity in art viewing behavior measured via eye movement patterns. In this study, we explicitly examine the effect of *visual culture* i.e., the reading/writing system. While the German reading/writing system in Austria reads and writes only from left to right, the Japanese reading/writing system has both horizontal and vertical reading/writing directions with different alphabets (e.g., Hiragana, Katakana, Kanji). This reading/writing system is recognized as unique even in East Asian countries. For example, while the vertical reading direction can only be found in old Chinese texts, Japanese still write/read vertically on a regular basis. We tested if these habitual practices in our everyday lives would have an impact on art viewing behavior or not. Our aims in Study 3 were to, firstly, directly challenge the existing assumed universality, secondly, explore the potential effect of culture on art evaluation,

and thirdly, investigate the effect of a specific aspect in visual culture. We believe this will shed further light on the effect of culture and inspire further and clearer hypotheses/assumptions on the impact of culture for future studies.

2. Study 1—Translating well-documented attention decrease phenomenon towards art from museum studies into lab-based study

From here on, the main empirical sections of this dissertation are presented. Note that the scientific work here includes collaborations between researchers in Tokyo/Vienna and the Ph.D. candidate. Consequently, I will use *we* as the subject in the following contents.

In Study 1, as a first step, we aim to transfer this well-documented effect of repeated art viewing on art viewers' attention reported from museum studies into a lab-based setting. If the effect could also be observed in a laboratory, then we could further test the underlying mechanisms with well-controlled lab-based setting. Since the effect of repeated art viewing is a core phenomenon across Study 1 and 2, in the following section, we will first present an elaborated overview of the past findings on this topic, the limitations in the past studies, and the purpose of Study 1. We will then present two empirical lab-based studies, investigating if the attention decrease phenomenon towards art from repeated art viewing could also be observed with an updated procedure in a laboratory.

To date, a series of studies in museums/art galleries have reported that repeated art viewing would bring a negative impact on the visitors' attention towards exhibits—the decreased attention towards exhibits as a function of repeated object exposure/of time in various exhibition settings (e.g., Bitgood et al., 1986; Falk et al., 1985; Gilman, 1916; Johnston, 1998; Marcellini & Jenssen, 1988; Matamoros, 1986; Melton, 1935; Porter, 1938; Robinson, 1928). However, due to the difficulty in controlling confounding variables in real-life settings, the underlying mechanism of the

phenomenon has not been revealed yet, leading to the lack of the knowledge on how to counteract this unwanted phenomenon.

One might argue that if the phenomenon is well-reported in museums/galleries, it would not bring new insights to perform replication studies in a laboratory. However, as pointed out, the experience of art viewers could largely differ between actual museum and laboratory setting (e.g., Brieber et al., 2014, 2015; Martindale, 2007; Muth et al., 2017; Specker et al., 2017). In fact, there are two studies on this topic, which were conducted in a laboratory setting (Bitgood et al., 2001; Robinson, 1928). However, importantly, they reported the opposite results. While Robinson (1928) did find the decrease in the art viewers' attention, Bitgood et al., (2001) did not find such a trend. Consequently, it is still unclear if the effect of repeated art viewing could be observed in a laboratory or not.

Further, though this effect in a laboratory has been ignored in the past, considering almost all past studies in empirical aesthetics have shown multiple artworks in a laboratory as well—in several cases, tens to even hundreds (e.g., Kirk et al., 2009; Leder et al., 2013; Van Paasschen et al., 2015), such a negative impact from repeated art viewing might also be a key factor in a laboratory and would have a direct relevance to all past findings in empirical aesthetics. Indeed, this was the main reason that many studies implemented randomization or counterbalancing of the presentation order of the stimuli. Nevertheless, given that the techniques like block-wise designs/within-participant testing are commonly used, the effect cannot always be fully eliminated. Hence, to study and qualify the effect of repeated art viewing in a laboratory would also bring important insights to lab-based empirical studies.

Past studies of the impact of repeated art viewing in museum/gallery exhibition settings

In various museum/art gallery exhibition settings, it has been well-reported that repeated art viewing brings negative impacts on the visitors' attention towards artworks. This attention decreased phenomenon was first found by Gilman (1916), known as the secretary of the Boston Museum of Fine Arts. In this study, he tracked and photographed one museum visitor's entire visit of the Boston Museum of Fine Arts. As a result of his observation, he found a pattern in the

visitor's behavior: "after a brief initial exertion he [the visitor] will resign himself to seeing practically everything imperfectly and by a passing glance" (Gilman, 1916, p. 62).

Before further proceeding the introduction, it is important to clarify two points to make the argument/purpose of the present study clearer. First, though we do acknowledge that Gilman (1916) named the decreased attention phenomenon as *museum fatigue*, following studies insisted that the term might bring potential misconceptions about the basis of the phenomenon since not only physical/mental fatigue, but several other factors and/or the combinations between them are attributing to attention decrease (e.g., see Bitgood, 2009a for a review). Hence, to prevent any confusions/misconceptions from the terminology, the term *museum fatigue* is only used in the literature review of self-described *museum fatigue* studies. When we describe the purposes or results of the studies in the present dissertation, we explicitly use phrases such as "decreased/diminished attention from repeated art viewing" or "the impact of repeated art viewing". Second, a brief note should be made on what *attention* means here. Typically, viewing time—the duration a person spent in front of one exhibited object (Bitgood et al., 2001; Bitgood et al., 1986; Johnston, 1998), a combination of viewing time and the percentage of stops in front of the exhibitions—calculated by dividing the number of the objects the visitors stopped to view by the total number of the objects on display (Matamoros, 1986; Melton, 1935; Porter, 1938a; Robinson, 1928), self-reported measures directly asking the level visitors' attention such as fatigue, interest in art, and/or desire to learn more (Bitgood et al., 2001; Cota-McKinley, 1999; Matamoros, 1986), or the combinations of the above variables are used as a proxies of visitors' *attention* towards the objects. According to Bitgood (2009a), though these metrics reflect visitors' *attention*, it is important to recognize that they are not equivalent, since "it is likely that these three measures may be sensitive to a different combination of phenomena [which might cause the diminished attention]" (Bitgood, 2009a, p.97). For example, viewing time is a proxy of *holding power* of an object which might be sensitive to the cause like satiation, while the percentage of stops is that of *attracting power* which might be sensitive to the cause like object competition. In the present

dissertation, we especially focus on the decrease in viewing time as a proxy of the holding power of an object caused by repeated art viewing. Note that elaborated descriptions on the possible causes of the diminished attention are shown in the following section, *Possible cause of decreased attention from repeated art viewing*.

Back to the literature review, following Gilman's study (1916), Robinson (1928) conducted systematic investigations on the behavior of museum visitors. In this study, he observed museum visitors' movements, stops, and viewing time for objects principally in three museums. Note that though he also observed the museum visitors in a fourth museum, this observation was rather unsatisfactory due to the change in the number of objects in the exhibition during the observation. One museum was located in a large city and had a big exhibition space, presenting roughly 1,000 paintings. The other two museums were located "in typical cities of the mid-west, cities having populations of a few hundred thousand" (Robinson, 1928, p. 15) with medium sized collections. Thus, the exhibited collections as well as the visitors' characteristics were different across three museums. As a result, he found a pattern in the behavior of museum visitors: when individuals confronted more artworks, the average time to view each artwork became shorter. Further, as the number of confronted artworks increased, individuals tended to skip over objects completely. These findings basically replicated the results of Gilman's work (1916), and again suggesting a diminished attention towards art after repeated art viewing. In a further subsequent analysis for each visitors' behavior pattern, he divided the seen artworks from each individual into tenths of the total number of viewed objects, since the total number of seen objects was varied across individuals ranged from 10 to 267. The results showed that the viewing time briefly increased or reached plateau at the begging of the art viewing for the first few tenths—he attribute such a rise to a *warming-up* effect. Afterwards, the viewing time showed decreased to the end of the visit. Importantly, the decline in the viewing time was observed across individuals regardless of the type of museums.

These pioneering studies were followed by a handful of researchers, generally showing that the viewing time as well as percentage of stops to see objects diminished over the progression of a visit in multiple settings such as aquariums (Matamoros, 1986), zoos (Bitgood et al., 1986; Marcellini & Jenssen, 1988), and history museums (Falk et al., 1985). Hence, the past studies repeatedly demonstrated that the attention decrease phenomenon from repeated art viewing could be observed regardless of the types of exhibitions as well as the characteristics of the viewers, suggesting that it is a general behavioral tendency in museum visitors.

Possible causes of decreased attention from repeated art viewing

Researchers have pointed out that the diminished attention from repeated art viewing may interrupt our satisfaction or effective learning of museum visits, regardless of the type of exhibition or stimuli themselves (Bitgood, 2002; Falk & Storksdieck, 2005; Falk, 1983; Falk & Dierking, 2002; Hein, 1998b; Serrell, 1997, see Davey, 2005a for a review). In other words, the effect of repeated art viewing has been recognized as a problem that needs to be solved in exhibition spaces. Consequently, the possible underlying mechanisms of decreased attention have been extensively discussed, which is eventually expected to build a concrete idea to counteract it.

Bitgood (2009a) especially focused on providing a comprehensive review of past studies on this topic as well as its possible causes. Again, as recognized by other researchers (Cota-Mckinley, 1999; Davey, 2005a; Melton, 1935; Robinson, 1928), he explicitly stated that “what has been called *museum fatigue*, is not a single phenomenon and may only partially be attributed to ‘physical or mental exhaustion’” (Bitgood, 2009a, p. 93) and “Several phenomena may play a role in museum fatigue” (Bitgood, 2009a, p. 94). He listed possible causes such as *fatigue* (tiredness, exhaustion), *satiation* defined as a decrease in viewing time after repeated exposure to a series of similar exhibits associated with boredom (Bitgood, 2009b), *stress*, *information overload* by being exposed to too much information, and/or *decision-making process* (see Bitgood, 2009a for the full review). However, though researchers have put considerable effort to summarize the possible causes of the

decreased attention from repeated object viewing (Bitgood, 2009a; Davey, 2005a), those ideas still remain rather conceptual—there have been less than a dozen empirical studies on this topic. Consequently, it is still an open question why attention towards objects decreases over the progression of visits and how to counteract it.

The reasons why there have not been many studies which empirically assess the cause of diminished engagement could be found in the nature of the study environments in the past studies. As aforementioned, almost all past studies have been conducted in actual museums exhibitions (Bitgood et al., 1986; Falk et al., 1985; Gilman, 1916; Johnston, 1998; Marcellini & Jenssen, 1988a; Matamoros, 1986; Melton, 1935; Porter, 1938; Robinson, 1928). Needless to say, we can gain a lot of benefits/knowledge by observing people in actual museum settings when we want to capture the natural experiences in ecologically valid settings, especially because art is always experienced in a specific context (Brieber et al., 2014). However, in museum settings, a large number of factors that might influence the diminished attention are involved with difficulties of controlling confounding factors. In other words, it is not always possible to examine the impact of one specific factor while controlling other confounding variables. For example, if we try to investigate the impacts of satiation on visitors' attention towards exhibits in an actual museum setting, depending on the time, date, and/or seasons it might be possible that some participants view a series of artworks during a very busy and crowded time, while others can do so during a composed and empty time. Furthermore, in museums, there is always a certain level of object competition—the presentation of the multiple objects at a time. Thus, a number of confounding variables such as physical fatigue, stress caused by the crowdedness and object competition can hardly be controlled, and thereby, it is difficult to evaluate the pure effect of satiation on visitors' attention. As such, though there are always limitations in lab-based studies, a better-controlled lab environment (despite its limitations) might provide important insights on how each phenomenon plays a role on the viewers' attention towards objects (Davey, 2005a), which has been almost impossible to achieve within the past museum studies.

Past studies of repeated art viewing in lab settings

In fact, there are two studies conducted in a laboratory on this topic, in order to empirically investigate possible causes of the decreased attention over repeated art viewing. The first lab-based study was a follow-up study in Robinson (1928) which extended his findings from museums into a lab setting. He was motivated to conduct another study in a lab in order to rule out the physical fatigue as a possible cause of the diminished attention in museum. In this study, the participants ($N = 27$, 14 females, “mostly undergraduate students of psychology having no special interest in painting or similar arts” (p. 36) were instructed to be seated at a table, and to view a series of art prints, presented on a board in front of them. One art print was presented at a time, and they could view each image as long as they wished. After seeing an image, it was replaced to another one by the experimenter. In total, 100 black and white art prints were used as the stimuli, and the participants repeated trials until they finished viewing all art prints. Further, the participants were told that there will be no additional tests after viewing the images e.g., memory/rating tasks in order to avoid any anticipation effects from the participants.

As a result, the viewing time for each art print decreased as a function of trial. Interestingly, the decrease in the viewing time was most marked in a lab-based setting compared to museum settings. Similar to the results from museums, the viewing time showed a brief increase at the beginning in a laboratory as well, and then steeply fell off. As expected by Robinson, the study design in a lab-setting seems to successfully control some confounding factors. For instance, the effect of physical fatigue was minimized, since participants were sitting in a lab. Further, the effect of object competition was also excluded, since one art print was presented at a time. Hence, Robinson (1928) could make an argument that the decrease in viewing time could be attributed to satiation, the effect of viewing homogenous stimuli for multiple times, and the lab-based study paradigm, where researchers can present one stimulus at a time and can control how many stimuli

are presented to the participants, could be considered suitable for assessing the effect of satiation on the decrease in viewing time.

The second lab-based study was performed by Bitgood et al. (2001) with a similar experimental paradigm as in Robinson (1928). However, interestingly, they did not find any decrease in viewing time. It is important to note that this study had specific aims which were different from Robinson (1928): “we were especially interested in assessing the ‘fatigue’ effect, as well as in determining the type of responses (idiosyncratic or predictable) given to the artwork” (Bitgood et al. 2001, p.26). In this study, 41 university students (“Approximately 50% were males” (p.26), age range between 18 and 36), who “...differed widely in terms of previous visitation to art museums, interest in art, and knowledge of art and artists” (p.26), participated. Though the art viewing task paradigm was quite similar to Robinson (1928), there were three distinct parts in the experiment: pre-session survey, art viewing task, and the post-session survey. All participants went through the three tasks in the same order. In the pre-session survey, they were asked to answer the following questions: ‘the frequency of museum visitation’, ‘interest in art’, and ‘the name of artists with whom they were familiar’. Following the pre-session survey, the experiment moved to the art viewing task. In this task, 40 color prints of portraits were shown to the participants. Again, one art print was presented at a time. The participants could view each art print as long as they wished. When they wanted to stop viewing one painting, they said ‘stop’. Then the print was replaced to another one by the experimenter. The viewing time for each art print, from the replacement of an art print to the participants’ verbal report of ‘stop’, was recorded as viewing time by the experimenter. Note that since their main aim was to assess ‘fatigue’ effect on the decrease in viewing time, they hypothesized that “if mental fatigue is an important factor in museum fatigue, then giving participants a more difficult task should hasten the effect” (Bitgood, 2009a, p. 102). Thus, they asked the participants to verbally report “whatever came to mind” (p.25) during the art viewing task. The number of verbal reports made by participants were also manually taped and

transcribed by the experimenter². The art viewing task was followed by the post-session survey, wherein the participants were asked to answer the following questions: ‘difficulty of the task’, ‘mental fatigue’, ‘interest in the task’, ‘desire for more information about the art they viewed’ and ‘interest in type of art’.

The results showed that the mean viewing time for each art print did not show a systematic decrease over time, formulated by the authors as “The ‘fatigue’ effect found by Robinson (1928) was not observed here” (p.27). Further, they found that the number of words made by the participants were varied with viewing time. Furthermore, they found some relationships between the ratings: firstly, ‘interest in art’ scores were higher in the post-session than in the pre-session, secondly, ‘interest in type of art’ in the post-session was positively correlated with ‘task interest’ ($r = .648$) as well as ‘desire for more information about the art they viewed’ ($r = .652$) in the post-session, thirdly, ‘task interest’ scores in the post-session were positively correlated with ‘desire for more information about the art they viewed’ ($r = .557$) in the post-session, finally, ‘interest in type of art’ in the post-session was negatively correlated with ‘mental fatigue’ ($r = -.511$) in the post-session.

Issues and questions for repeated art viewing studies in the laboratory

Though these two studies are pioneering examples that provided a starting point on how we could empirically investigate the attention decrease phenomenon as well as its causes in a laboratory setting, they left several questions open.

First, since the two lab-based studies on this topic showed incompatible results with different procedures, it still remains unclear if the viewing time as a proxy of viewers’ attention decrease in a lab-based setting or not. While Robinson (1928) only asked the participants to view a

² Note, in Bitgood et al. (2001), the analysis of the content of verbal reports was also included. However, the contents analysis is out of the scope of the present dissertation, and thereby, I only report the results of pre-/post- session survey, of the viewing time, and of the number of the words made by participants.

series of artworks without any secondary task—a typical assessment method on this topic in past museum studies, Bitgood et al. (2001) did include an obvious secondary task—verbally describing whatever came to mind while viewing the artworks, meaning that the duration when the participants made their verbal responses were also included in the viewing time analysis. This analysis method leaves another question open: Did Bitgood et al. (2001) measure a pure index for *attention* towards artworks? The authors' discussions also alludes to this aspect, “In general, a ‘practice effect’ across successive art prints was characterized by closer matching of time and words with increased experience on the task” (p.27). Further, as shown in the results of the pre- and post-session, the scores of interests in art were higher in the post-session than in the pre-session. This result demonstrated that the task itself increased the participants’ interest to art, and “participants obviously enjoyed the task, even when they reported considerable difficulty in describing the art works” (Bitgood et al., 2001, p.30). Thus, it might have been possible that the participants did not show any decrease in the viewing time, since the task itself made the art viewing more enjoyable, or they became more fluent in making verbal responses over the course of the experimental task. Hence, the procedure seems to assess the effect of making verbal responses on the viewing time, rather than to assess the effect of pure repeated art viewing effect. We cannot see the impact of making verbal reports during art viewing per se, since there was no control group included in Bitgood et al. (2001), wherein the participants did not provide any verbal reports.

Second, both studies (Bitgood et al., 2001; Robinson, 1928) had another limitation that might have influenced the results. In both studies, the participants were allowed to view presented art prints as long as they wished, and there were no time restrictions in the art viewing task. This procedure were typical in the past museum studies and could be considered as replicating natural art viewing settings in museums/art galleries, wherein we do not typically have any time limits. However, in this procedure, the duration for the entire experiment—the total amount of time to finish the task—differs across participants, completely depending on the participants’ response time for artworks. Say one participant noticed during the experimental task that s/he could finish the task

faster if s/he makes each response quickly. What we see as a result is a gradual or even steep decrease in viewing time. Considering that the participants normally participate in psychological studies for monetary compensations or course credit, it is the most effect way to finish the task with minimum effort to get the reward. Thus, this procedure may show the decrease in viewing time, since the participants wanted to finish the study as fast as they could, raising a question if the results of the past studies reflect *the attention decrease* or the participants' wanting to finish the study for their reward. Further, this point raises another question when we transfer this topic to typical laboratory-based paradigms in empirical aesthetics. Lab-based studies in the field do sometimes fix the stimulus presentation duration (e.g., Augustin et al., 2008; Leder et al., 2006; Zangemeister et al., 1995). Hence, it is another question if such a study design would show similar effect as well. In sum, it still remains unclear if the decreased attention phenomenon could be observed in a laboratory setting due to the incompatible results and methodological limitations in the past two laboratory studies. Accordingly, further replications with updated procedure are necessary to see the effect of repeated art viewing in a lab-based setting.

Lastly, while the effect of repeated art viewing on the attention of the viewers has been extensively studied in the past, its impacts on art ratings/appraisals towards art have been little assessed. As the past psychological studies have shown, the attention for visual stimuli often correlates with hedonic experience of the stimuli. For instance, interesting/emotionally arousing stimuli receive longer viewing time (Lang et al., 1993; Smets, 1975) and attractive faces (Kwart et al., 2012; Leder et al., 2010) and artworks (Goller et al., 2019) also receive longer view. Further, Brieber et al., (2014) also found that viewing time for artworks and the appreciation scores (compounding liking and interest scores towards artworks) was positively correlated both in a museum and in a laboratory. At the same time, as shown in other research examining phenomena like *mere-exposure effect* (Zajonc, 1968), the number of exposures of visual stimuli might have impacts on subjective experiences of the stimuli. Generally, hedonic ratings towards stimuli are enhanced after the repeated exposures to the stimuli—a classic *mere exposure effect*, but Zajonc et

al. (1974) further reported that extreme numbers of exposures might result in an inverted U-shaped curve of hedonic ratings. As even more direct implication was reported by Leventhal et al. (2007), showing that individuals evaluated one stimulus less pleasurable after the 20times exposures of the stimulus. Considering the relationship between the viewing time and ratings for stimuli as well as the impact of extreme exposure of stimuli on hedonic subjective experiences, it seems plausible to assume that hedonic experiences towards artworks could also decrease as a function of repeated art viewing (see also Davey, 2005a for the argument in line with museum studies). If repeated art viewing also has negative impacts on hedonic subjective experience of art, it not only offsets pleasant experiences and benefits provoked by artworks, but also converts art appreciation into less pleasant experience (Davey, 2005b). Considering also that virtually all studies in empirical aesthetic ask individuals to evaluate their subjective experiences towards art, this aspect needs to be further examined and would actually be an important question for both museum visitor research as well as for lab-based studies.

The purpose of Study 1

Study 1 aims to transfer the well-documented attention decrease phenomenon from museum studies into laboratory setting and examine its impact on viewers' attention and appraisal towards art through two psychological experiments. In the first experiment, we aim to assess the general impact of repeated art viewing on the viewing time as a proxy of their attention towards presented objects as well as appraisals towards art measured by beauty rating with an updated procedure. Specifically, given the shortcomings in the methodologies in the past two lab-based studies (Bitgood et al., 2001; Robinson, 1928), we invented another experimental procedure wherein the participants cannot intentionally shorten the overall duration of the experimental task. The detailed procedure will be explained in the following section. In the second experiment, we introduced another variant of the repeated art viewing question: if taking a rest could mitigate the typical reduction in viewing time. Though the effect of taking rests in between has never been tested in a

lab-based setting, for instance, Robinson (1928; see also Davey, 2005a) pointed out that whether the viewers can take a rest or not is one of the potentially important differences between lab-based and museum studies. In actual museum/gallery exhibition settings, there are some rooms, spaces, and/or benches where the visitors can take a break during their visit. Presumably, these practices are implemented, assuming that such a break might be an important for the visitors to prolong their enjoyments. Thus, in the second study, we try to empirically test this implicit assumption, whether such a pause during art viewing could recover/mitigate the decrease in viewing time as well as in appraisals for the art.

Finally, past studies on this topic have assumed that the decrease attention phenomenon is rather a general tendency of art viewers and most studies recruited participants only from European or North American countries. However, in fact, researchers have suggested that the impact of culture on the effect of repeated art viewing should be investigated (Davey, 2005; Falk et al., 1985). To the best of our knowledge, Davey (2005b) is the only study, considering different cultures other than WEIRD people (Henrich et al., 2010). Given that culture might have an impact of our cognitive styles and how we explore the visual field (e.g., Chua et al., 2005; Masuda & Nisbett, 2001) and museums typically deal with visitors from various cultural backgrounds, cross-cultural components on this topic might contribute to accumulate the knowledge on the universality of the effect. Consequently, though the cross-cultural components is rather explorative, Study 1 recruited participants from Japan and Austria, and directly compared the results between the two cultures.

2.1. Study 1-1: Cross-cultural replication study with an updated research design for examining the repeated art viewing effect in a lab

Published work

Mikuni, J., Specker, E., Pelowski, M., Leder, H., & Kawabata, H. (2021). Is there a general “art fatigue” effect? A cross-paradigm, cross-cultural study of repeated art viewing in the laboratory.

Psychology of Aesthetics, Creativity, and the Arts. 88, p37-51. <https://doi.org/10.1037/aca0000396>

My Contribution:

I acted as the first author of the paper and contributed to it with various aspects: with HK, we developed the core concept of the paper. I made an experimental program file, collected the participants with ES, and analyzed the results. I was the main writer of this article, and elaborated the contents with feedback from ES, MP, HL, and HK. Note that Study 1-1 in the present dissertation corresponds to Study 2 in the published article. Further, Study 1-2 in the present dissertation corresponds to Study 3 in the published work.

2.1.1. Method

Participants

In Study 1-1, 36 native Japanese psychology students (24 female, $M_{age} = 20.19$, $SD = 1.48$) and 38 psychology students from the University of Vienna (30 female, $M_{age} = 20.79$, $SD = 3.39$, all native German speaking and of European cultures) were participated. They all had normal or corrected-to-normal vision. All participants did not have an academic background in art/art history and were not professional artists (verbally assessed after the testing session). They all received course credits from their own universities for the participations. The study was approved by the local ethical committee of Keio University (Japan) and followed the guideline of the University of Vienna, Austria. Note that the final sample was reduced from an original collection of 42 participants from Japan and 50 participants from the University of Vienna. Hence, six participants from Japan and 12 from the University of Vienna were excluded due the data collection issues (i.e., the responses of the participants were not recorded correctly).

Stimuli and Apparatus

120 paintings from WikiArt (<http://www.wikiart.org>, the online artwork archive) were used as the stimuli. The stimuli are intended to represent a wide range of art styles and periods in order to reduce the effects from specific artists, styles, and to mimic a typical art viewing setting as could be encountered and as were assessed previously in museums and in a typical lab-based setting. Half of the stimuli were landscape, and the other half were still life paintings. The 120 paintings were further divided into four blocks. Thus, one block contained 30 paintings with equal numbers of landscape and still life. This division was explicitly made to make sure that the mean beauty rating scores between the four block did not differ ($F(3, 116) = 0.19$, $\eta^2=0.01$, $p = .90$), suggesting that the beauty ratings in the stimulus sequence was constant from the beginning to the end. The beauty rating scores were acquired from the pre-test with 20 native Japanese participants (13 females, $M_{age} = 22.20$) who made beauty ratings ranged from 1 (*not at all beautiful*) to 7 (*very beautiful*) for 100

landscape and still life paintings (200 paintings in total). Note that the presentation order of the stimuli within each blocks randomized, and the order of four blocks was counter balanced across participants. A CRT monitor (Trinitron CPD-G420, SONY) was used for the presentation of the stimuli. One artwork was presented at a time in the center of the screen with a gray background. The horizontal dimension of all paintings was fixed at 600 pixels. The experiment was controlled by PsychoPy 3.0 (Peirce, 2007) on a MacBook Pro (Apple, USA). Participant sat in front of the monitor at a distance of about 80 cm away.

Procedure

We used similar procedure as in Robinson (1928). However, importantly, we also asked participants to rate the beauty of the presented painting after viewing each painting. The beauty assessment was added in order to better capture the typical lab-based paradigm in empirical aesthetics, and to be able to analyze art appraisals as a function of trials. Beauty appraisal was explicitly selected as it acts as a general metric of art appreciation especially for the art lay people and as it is one of the most commonly used rating scale in previous research (Jacobsen et al., 2004).

After we welcomed the participants, they got instructions saying that they would be shown a series of artworks presented in the screen monitor. Further, they were also asked to make a beauty rating after viewing each painting. In order to start each trial, the participants pressed a key, followed by a painting presented on the screen monitor. When they wanted to finish viewing one painting, they pressed the same key again. Importantly, in contrast to Bitgood et al. (2001) and Robinson (1928), there was a time limit in each study. Each viewing trial duration was always set at 15s, regardless of whether the participants pressed the key to stop viewing or not. More specifically, when the participants pressed the key to stop viewing each painting, the screen would be altered by a blank gray screen for the rest of time. After 15s passed, the screen moved onto the rating screen, and the participants were asked to rate the beauty of the painting with a Likert-type beauty scale ranged from 1 (*not at all beautiful*) to 9 (*very beautiful*). Further, we also set a

minimum required time and maximum viewing time for the painting to be appeared on the screen of 3 s (in that case, followed by 12s of blank screen) and 12s (in this case, followed by 3s blank screen). In other words, the participants viewed all paintings at least 3s and could view one painting for 12s maximum. Thus, for example, if a participant did not press the key to finish viewing in a given time period, the painting remained on the screen for 12s and then replaced by a blank gray screen for the additional three seconds. The minimum required time and the maximum viewing duration were implemented to create more equal conditions across participants to assure that all participants did not accidentally skipped some paintings as well as to minimize potential effects for memory-related impacts on art ratings. The 12s for the maximum viewing time was chosen based on previous studies both in museums and laboratories, reporting that the viewing time is around 11s when individuals are allowed to view artworks freely (Smith & Smith, 2001; see Pelowski et al., 2017 for a review).

Though the exact number of paintings used in the experiment was not told to the participants in order to avoid any anticipation effects, they were told that the experiment would last a full 30minutes and they would not be able to leave the testing room until the experiment would be completed. Before the participants started the main experimental task, they completed five practice trials to become familiar with the procedure. Experimental conditions, the study design as well as the instructions included in the experiment were identical in Japan and Austria.

2.1.2. Results

Overall, the average artwork viewing time for each individual ranged from 3.49 to 12.00s for the Japanese and from 5.09 to 11.80s for the participants in Austria; the *SDs* ranged from 0.00 to 3.72 for the Japanese and 0.79 to 3.62 for the participants in Austria, respectively. The average viewing time across participants was 6.45s (*SD* = 3.23) for Japanese and 7.63s (*SD* = 3.31) for the participants in Austria. Here, we can see that participants did press the key to stop viewing, although there was forced waiting period with a blank gray screen for the rest of the trial duration.

In total, in 14.95% of trials for Japanese and in 26.54% for the participants in Austria, the participants spent the full 12s without pressing the key to stop viewing with within participant range between 0.00% and 100.00% for the Japanese and 0.01% and 99.17% for participants in Austria. Note that one Japanese participant had never pressed the key to stop viewing and always viewed paintings for 12s. Considering that the data from this participant might skew the data structure, the person's data was excluded from the following analysis of the viewing time.

The average beauty rating scores for each individual ranged from 3.75 to 6.42 for the Japanese and from 3.61 to 7.70 for the participants in Austria; the *SDs* ranged from 0.83 to 2.67 for the Japanese and 1.04 to 2.81 for the participants in Austria. The average beauty rating scores across participants were 5.23 (*SD* = 1.98) for the Japanese and 5.01 (*SD* = 2.12) for the participants in Austria. Before diving into the main part of the results section, we computed the correlation score between viewing time and beauty rating scores for individual participant. The correlation scores ranged from 0.24 to 0.80 for the Japanese and -0.12 to 0.56 for the participants in Austria. Further we conducted one-sample *t*-tests comparing the average correlation scores of all participants to zero, separately for Japan and Austria groups. The results showed that, in both groups, the average correlation scores differed significantly from zero (Japan: $M_r = 0.14$, $t(34) = 3.29$, $p < .01$, $d = 0.56$ Austria: $M_r = 0.22$, $t(37) = 7.56$, $p < .01$, $d = 1.23$), suggesting that the viewing time and beauty rating scores were positively correlated in both culture groups.

Impact of repeated art viewing on viewing times

Figure 2 shows the average viewing times for each trial in both culture groups. As shown in Figure 2, the Austrian group generally viewed paintings longer, especially at the beginning of the experiment. Nevertheless, the viewing time in both culture groups showed decrease as a function of trials, with both groups ending at about six to seven seconds during the last few trials.

To assess the impact of repeated art viewing on viewing time, we conducted a GLMM (Generalized Linear Mixed Model). The LMM was carried out in the R environment (Version

1.1.423. R Development Core Team, Bates et al., 2017) with lme4 package (Version 1.1-13; Bates, Mächler, Bolker, & Walker, 2015) applying Satterthwaite approximation for the estimations of p values. In the following analysis in Study 1, we used the same software and packages for analyzing the data. We set viewing time for each trial as our dependent variable, and the number of trials (from 0 to 119, coded continuous), a treatment coded contrast comparing the participants' *Culture*, with Austrian (baseline) against Japanese, as well as the interaction between the two variables as the fixed effects. For random effects, random intercepts for participants and images were included. Note that the random by-participant slope was not included in the random effect structure, since otherwise the model did not converge with this component.

All the fixed effects and the results are shown in the upper part of Table 1.

As a result, we found a significant main effect of trial, meaning that the viewing time for each artwork decreased as a function of the trial number progressed. Hence, we found that the attention decrease could be found in a laboratory with updated procedure where the participants could not intentionally shorten the duration of the experiment. Further, the main effect of cultures was also significant, suggesting that the viewing time of the participants in Austria was generally longer compared to those of Japanese participants. Also, the interaction between the trial number and culture was significant, reflecting that the slope of the viewing time was significantly steeper in Austrian group than that in the Japanese one. Given this significant interaction, we further performed a post-hoc GLMM only using the viewing time data from the Japanese group to make sure that viewing time in Japanese group also decreased as a function of the trials. Since we only used the data from Japanese participants, in this model, the fixed effect of Culture was removed from the model structure, and thereby, only the main effect of trial was included as the fixed effect. The structure for the random effects was identical to the model presented before. The results are shown in the middle part of Table 1. As a result, the main effect of trial was significant in this model as well, showing that the viewing time were significantly decreased over trials in Japanese group.

As shown in Figure 2, the data showed that the viewing times especially for the first 10 trials in the Austrian group seemed quite long. The similar trend could also be seen in the first 10 trials in the Japanese group. This trend might suggest that the participants in Austria showed a truly pronounced point of inflection after a few exposures, or that they might have taken the first trials to practice each trial flow before settling into a standardized viewing attitude. To further check especially the latter possibility, we performed another GLMM without the viewing times in the first 10 trials for both culture groups. The structure of the model was identical to the first model with trial, culture, and the interaction between them as the independent variables. Again, the structure of the random effects were identical to the other models. All fixed effects and the results are shown in the lower part of Table 1. The results again showed that there was a significant main effects of trial number and culture, and the significant interaction between the two variables. Hence, as suggested in the first model, the viewing time in Austrian group were significantly longer than in Japanese group after the first 10 trials as well. Further the significant interaction suggests that the slope of the viewing time was still steeper in the Austrian group after the first 10 trials compared to the Japanese group.

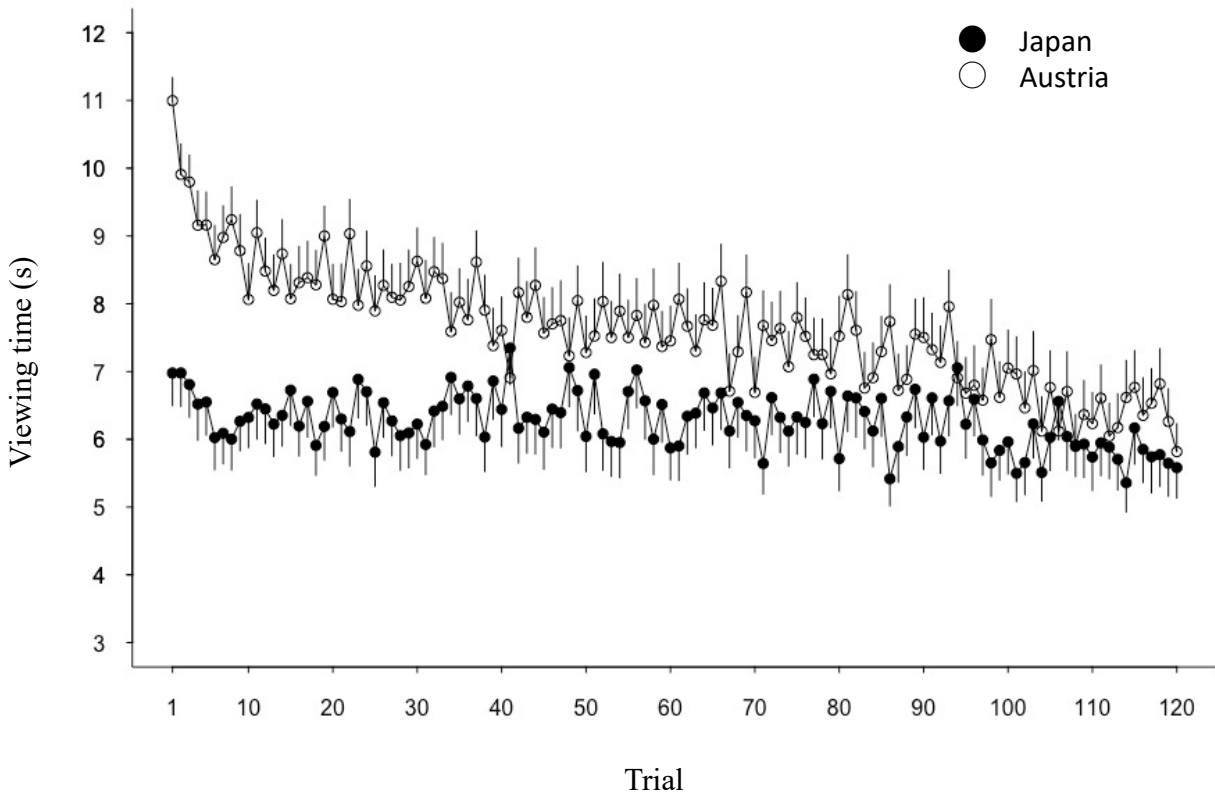


Figure 2. Average artwork viewing time for each trial in both culture groups. Error bars indicate the Standard Error (SE) for each point. The black points represent the result of Japanese participants, and the white points represent the result of participants in Austria. Note that the figure content is partially identical to the *Figure 2* in Mikuni, Specker, Pelowski, Leder, & Kawabata. (2021).

Table 1. Fixed effects and coefficients in the GLMMs predicting the viewing time.

Fixed effect	Estimate	95 % CI		t value	Pr (> t)
		Lower	Upper		
General impact of repeated exposure and culture					
(Intercept)	1.99	1.88	2.10	35.02	$p < .01$
Trial	-0.12	-0.13	-0.10	15.72	$p < .01$
Culture	-0.23	-0.38	-0.07	2.87	$p < .01$
Trial * Culture	0.08	0.07	0.09	11.09	$p < .01$
General impact of repeated exposure and culture: Japan					
(Intercept)	1.84	1.82	1.85	259.85	$p < .01$
Trial	-0.03	-0.04	-0.01	3.94	$p < .01$
General impact of repeated exposures and culture, without the first 10 trials					
(Intercept)	1.97	1.86	2.09	33.61	$p < .01$
Trial	-0.10	-0.11	-0.08	12.69	$p < .01$
Culture	-0.21	-0.37	-0.05	2.60	$p < .01$
Trial * Culture	0.07	0.05	0.08	8.82	$p < .01$

Note: The top part of the table shows the fixed effects and the results of the GLMM including the number of trials, culture as independent variables. In the middle, the result of the GLMM using Japanese participants data is shown. The bottom shows the results of the GLMM excluding first 10 trials are shown. Note that the table content is identical to the Table 2 in Mikuni, Specker, Pelowski, Leder, & Kawabata. (2021).

Impact of repeated art viewing on beauty rating score

Figure 3 presents the average beauty ratings for each trial in Austrian and Japanese groups. Though the decrease is not so salient as in the viewing time, we can see that there is at least a slight decrease in beauty rating scores as well. In order to statistically assess the effect of repeated art viewing on rating scores, we performed a similar LMM (Linear Mixed Model). For this model, we set beauty rating scores for each trial as the dependent variable, and the trail number (from 0 to 119,

coded continuous), participants' *Culture* with Austrian set as the baseline, and the interaction between the two variables as the fixed effects. Note again that the structure of the random effects was identical to the models for the viewing times. All fixed effects and the results are shown in Table 2. The results of LMM again showed that there was a significant main effect of trial, suggesting that not only the viewing time but also the beauty rating scores significantly decreased over trials. However, in contrast to the results for the viewing time, the main effect of culture and the interaction between the trial numbers and culture were not significant.

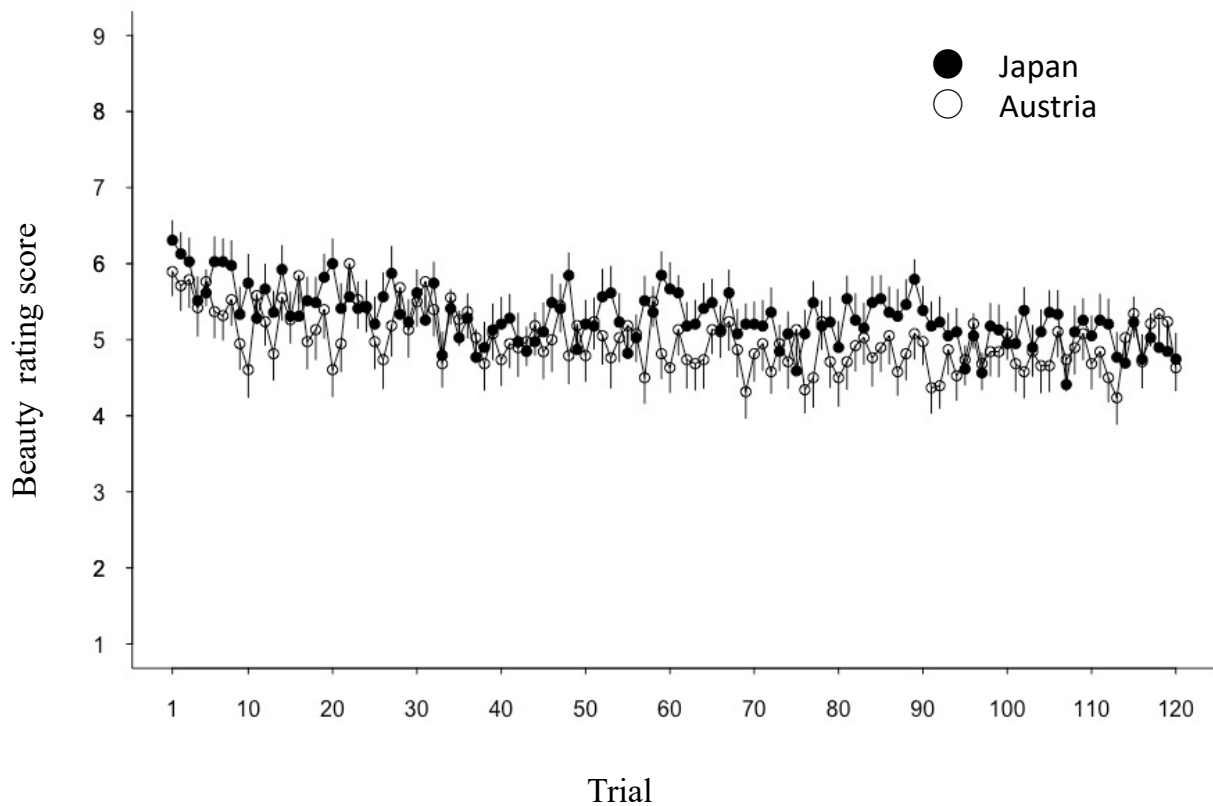


Figure 3. Average beauty rating scores for each trial in both culture groups. Error bars indicate the Standard Error (SE) for each point. The black points represent the result of Japanese participants, and the white points represent the result of participants in Austria. Note that the

figure content is partially identical to the *Figure 2* in Mikuni, Specker, Pelowski, Leder, & Kawabata. (2021).

Table 2. Fixed effects and coefficients in the LMMs predicting the beauty rating score.

Fixed effect	Estimate	95 % CI		t value	Pr (> t)
		Lower	Upper		
General impact of repeated artwork viewings and culture					
(Intercept)	5.01	4.71	5.30	33.39	$p < .01$
Trial	-0.15	-0.26	-0.03	2.56	$p = 0.011$
Culture	0.28	0.12	0.81	1.56	0.12
Trial * Culture	-0.01	-0.08	0.06	0.20	0.84

Note: The table content is identical to the Table 3 in Mikuni, Specker, Pelowski, Leder, & Kawabata. (2021).

2.1.3. Discussion

Study 1-1 presented an updated procedure to assess the effect of repeated art viewing on the viewing time and beauty rating scores in a laboratory. The results showed that, first, the viewing time decreased over trials regardless of the culture groups (Austria and Japan), replicating the results of Robinson (1928). Thus, to conclude, the attention towards artworks did decrease throughout a typical lab-based study. Second, not only the viewing time, but also beauty appraisals towards art showed decrease as a function of the trial numbers again regardless of the culture. Last, the results of correlation scores between the viewing times and beauty rating scores and one sample *t*-tests showed that there was a positive correlation between the two variables—more beautiful paintings got longer view.

Interestingly, Study 1-1 showed that the viewing times and beauty ratings did decrease with an updated experimental procedure, where the participants could not intentionally shorten the total duration to finish the experiment and had to stay in one trial for 15s. In other words, the participants chose to remove the paintings from the screen and to wait for the rest of time to the next trial in

front of a blanc gray screen. This result is quite surprising, since, at least from a physical movement perspective, the most *efficient strategy* for this paradigm is to wait until the screen changed automatically to the rating task—viewing the artworks for the full 12s, so that they do not have to put additional effort without keypress. Nevertheless, there was only one Japanese participant who took this strategy, and all other participants explicitly chose to remove the paintings from the screen. These results might suggest that, for the majority of the participants, viewing and paying attention towards a series of artworks might have been cognitively taxing, leading to the explicit actions to reduce/stop the exposures to the artworks.

Though the past studies on this topic assumed that decreased attention from the repeated art viewing is rather a general tendency in art viewers, they did not explicitly compare the results from art viewers from different cultures. Our results showed that main effect of trials on viewing time could be found regardless of cultures, providing another evidence that there is an inter-cultural consensus of this effect. However, we did find some cross-cultural differences. For example, the viewing time in the Austrian group was generally longer than in the Japanese group, and the slope of the viewing time was significantly steeper in the Austrian group compared to the Japanese group. Hence, it might be possible that, though the effect of repeated art viewing could be found regardless of culture, how and how much it impacts the viewing times could differ across culture. This point will be further discussed in the general discussion for Study 1.

2.2. Study 1-2

Study 1-2 provides another replication of the effect of repeated art viewing with the updated experimental procedure. However, this time, we also tested the impact of a rest period during the repeated art viewing, which is commonly implemented in art viewing settings in museums/galleries and lab-based experiments. As aforementioned in the general introduction for Study 1, in natural art viewing settings, individuals might not engage in art viewing as one continuous march without breaks, but rather, they might take some breaks in between, sitting on a bench or perhaps in a museum café. Perhaps, this common practice might be expected to refresh physical/mental fatigue accumulated from the previous art viewing (e.g., Bitgood, 2015). At the same time, a lot of lab-based studies in empirical aesthetics have used block-wise experimental design with short resting period in-between the blocks to provide some rests for the participants or to balance tasks and/or stimuli. This practice raises another question if such a rest during the repeated art viewing could mitigate the decrease in viewing times and beauty rating scores. Therefore, in Study 1-2, we used the same art viewing paradigm as in Study 1-1 with a resting period in the middle of the experiment, which allows us to compare the dependent variables before and after the break. As with Study 1-1, we again recruited our participants in both Japan and Austria.

2.2.1. Method

Participants

In Study 1-2, 38 native Japanese psychology students (24 female, $M_{age} = 19.61$, $SD = 0.63$) and 38 psychology students from the University of Vienna in Austria (32 female, $M_{age} = 20.03$, $SD = 1.63$) participated. The participants participated in Study 1-1 were not include in Study 1-2. All participants had normal vision or corrected-to-normal vision and had no academic backgrounds in art history/art (verbally assessed after the testing session). They all received course credit for their participations. The study was approved by the local ethical committee of Keio University, Japan

and followed the guideline of the University of Vienna, Austria. Note that the final samples were reduced from an initial sample of 43 Japanese participants and 41 participants in Austria, due to the recording issues (i.e., some responses from the participants were not correctly recorded).

Stimuli and Procedure

The presented stimuli, apparatus to present the stimuli and control the experiment, as well as the basic procedure of the experiment were the same as in Study 1-1. However, there was one change in Study 1-2. After half of the stimuli were presented (after the 60th trial), an instruction, saying *please take a rest for 90seconds*, was presented on the screen monitor. During the resting period, they could relax in the testing room or walk around the room. However, they could not go outside of the testing room and eat/drink anything. After the 90s passed, the screen was automatically going back to the art viewing task. The participants could restart the trial by pressing a key. Again, in order to avoid anticipation effect, the exact number of the stimuli used in the experiment as well as the presence of the resting period were not announced to the participants before they started the experiment.

2.2.2. Results

Overall, the average viewing times for individuals ranged from 3.47 to 11.92s (*SDs* ranged from 0.64 to 3.27) for the Japanese and from 3.67 to 11.88s (*SDs* ranged from 0.75 to 3.69) for the Austrian group. The average viewing time across all participants was 6.47s (*SD* = 3.22) for the Japanese and 7.06s (*SD* = 3.26) for the Austrian group. In total, in 17.60% of trials for the Japanese and 20.94% for the participants in Austria, the participants spent the full 12s without pressing the key to stop viewing with within participant rang between 0.00% and 98.33% for the Japanese and 0.00% and 98.33% for the participants in Austria. Thus, again, the participants in both culture groups did chose to remove the artworks from the screen.

The average beauty rating scores ranged from 4.38 to 6.75 (*SDs* ranged from 1.16 to 2.67) for the Japanese and from 3.27 to 6.43 (*SDs* ranged from 1.34 to 2.84) for the Austrian group. The average beauty rating score across participants was 5.48 (*SD* = 1.89) for the Japanese and 4.64 (*SD* = 2.19) for the Austrian group. As with Study 1-1, we computed the correlation scores between viewing times and beauty rating scores. The correlation scores for individual participants ranged from -0.21 to 0.62 for the Japanese and -0.10 to 0.52 for the Austrian group. Further, we conducted one-sample *t*-tests comparing the individual correlation scores from each participant to zero, separately for two culture groups. As a result, the mean correlation scores were significantly different from zero in both culture group (Japan: $M_r = 0.10$, $t(37) = 3.08$, $p < .01$, $d = 0.50$, Austria: $M_r = 0.11$, $t(37) = 4.64$, $p < .01$, $d = 0.75$), suggesting that the participants viewed more beautiful paintings longer.

Impact of repeated art viewing on viewing times and beauty rating scores

Figure 4 shows the average viewing times for each trial and Figure 5 shows the average beauty rating scores for each trial for the two culture groups. In the two figures, the resting periods were highlighted with dashed lines. As shown in Figure 4 and 5, we could see that both variables decreased over trials in both culture groups. Again, in order to statistically assess the impact of repeated art viewing, culture, and the interaction between the two variables on the viewing times, we performed another GLMM. The structure of the model was identical to the one in the first model in Study 1-1. All fixed effects and results are shown in the upper part of Table 3. As a result, we found a significant main effect of trial number, meaning that the viewing times decreased as a function of trial numbers. Further, as with Study 1-1, we again found a significant interaction between the trial number and the participants' culture. The direction of the interaction was also identical to the results of Study 1-1, suggesting that the slope of the viewing time over the trials was steeper in the Austrian group compared to the Japanese group. In contrast to Study 1-1, there was no significant main effect of participants' culture, and thus, the viewing time in general did not

differ between the two culture groups. Since there was a significant interaction, suggesting more gradual decrease in the viewing times for the Japanese group, we performed another GLMM only with the viewing time data from the Japanese participants. The model structure was identical to the one in the second model in Study 1-1. The results are shown in the lower part of Table 3. As a result, the main effect of trial numbers was again significant, showing that the viewing time in the Japanese group did also decrease over trials.

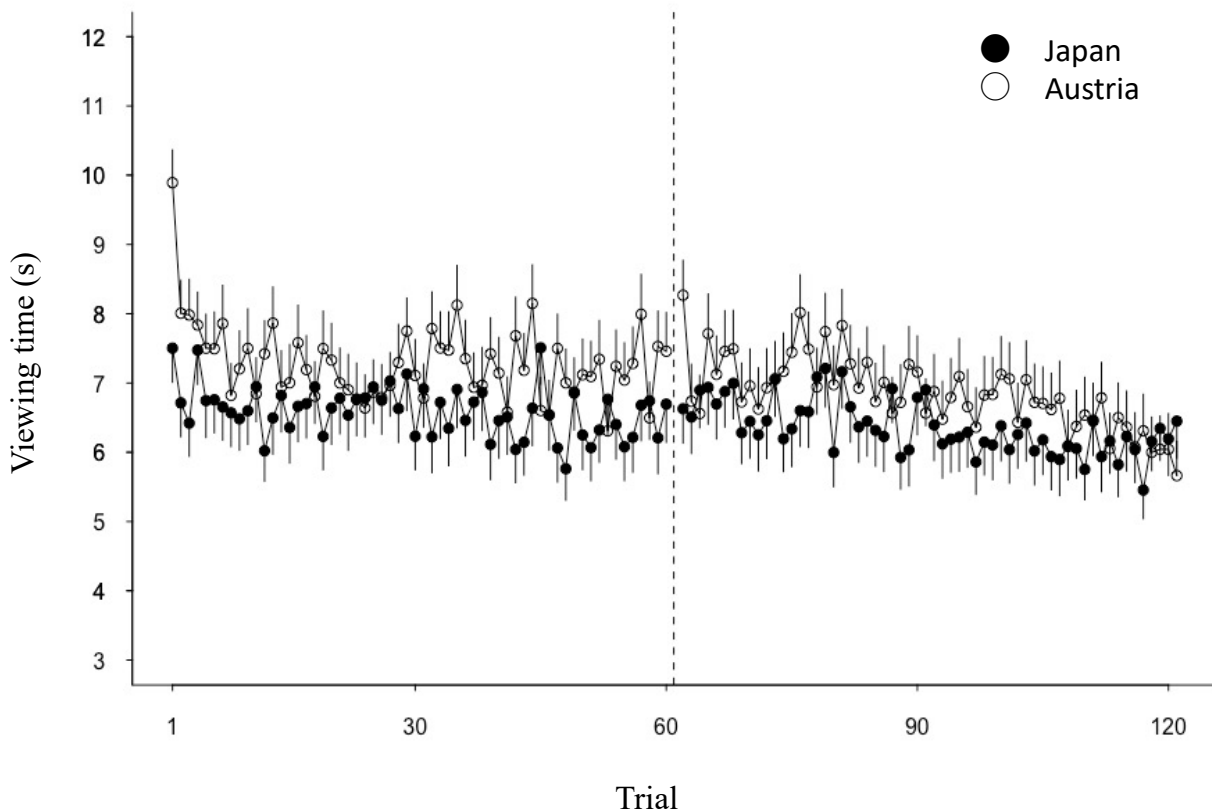


Figure 4. Average viewing time for each trial. Error bars present the Standard Error (SE) for each point. The black points represent the result of the Japanese participants, and the white points represent that of the Austrian group. Vertical dashed line represents the resting period. Note that the figure content is identical to the *Figure 3* in Mikuni, Specker, Pelowski, Leder, & Kawabata. (2021).

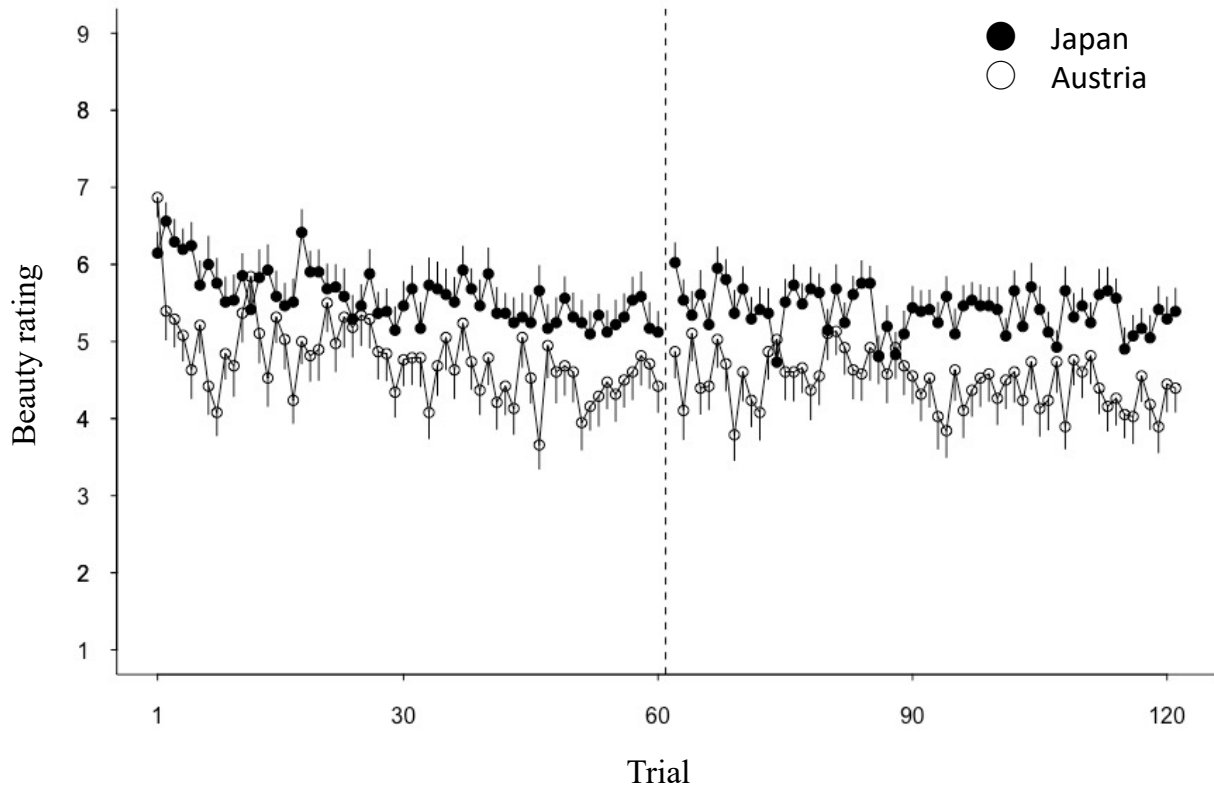


Figure 5. Average beauty rating scores for each trial. Error bars present the Standard Error (SE) for each point. The black points represent the result of the Japanese participants, and the white points represent that of the Austrian group. Vertical dashed line represents the resting period. Note that the figure content is identical to the *Figure 3* in Mikuni, Specker, Pelowski, Leder, & Kawabata. (2021).

Table 3. Fixed effects and coefficients in the GLMMs predicting the viewing time.

Fixed effect	Estimate	95 % CI		t value	Pr (> t)
		Lower	Upper		
General impact of repeated exposure and culture					
(Intercept)	1.90	1.78	2.02	30.61	<i>p</i> < .01
Trial	-0.07	-0.08	-0.05	8.90	<i>p</i> < .01
Culture	-0.11	-0.28	0.06	1.24	0.215
Trial * Culture	0.03	0.01	0.04	3.80	<i>p</i> < .01
General impact of repeated exposure and culture: Japan					
(Intercept)	1.79	1.66	1.92	26.73	<i>p</i> < .01
Trial	26.73	-0.05	-0.02	4.12	<i>p</i> < .01

Note: The top part of the table shows the fixed effects and the results of the GLMM including the number of trials, culture, and block as independent variables. In the bottom, the results of the GLMM using only Japanese participants data are shown. Note that the figure content is identical to the Table 4 in Mikuni, Specker, Pelowski, Leder, & Kawabata. (2021).

Further, to assess the impacts of repeated art viewing, culture, and the interaction between the two variables on beauty rating scores, we ran another LMM using beauty rating scores as the dependent variable. The fixed effects and results are shown in Table 4. Again, we found a significant main effect of trial numbers, suggesting the beauty rating scores did decrease over trials. However, in contrast to Study 1-1, we also found a significant main effect of culture, suggesting that the Japanese participants evaluated paintings as more beautiful compared to the participants in Austria. The significant interaction between the trial number and culture was absent.

Table 4. Fixed effects in the LMM predicting beauty rating scores.

Fixed effect	Estimate	95 % CI		t value	Pr (> t)
		Lower	Upper		
General impact of repeated exposure and culture					
(Intercept)	4.64	4.38	4.91	34.38	<i>p</i> < .01
Trial	-0.32	-0.44	-0.21	5.60	<i>p</i> < .01
Culture	0.85	0.56	1.14	5.78	<i>p</i> < .01
Trial * Culture	0.06	-0.01	0.13	1.65	0.099

Note: The table content is identical to the Table 5 in Mikuni, Specker, Pelowski, Leder, & Kawabata. (2021).

Impact of the resting period on the relation of repeated art viewing and viewing time

To further examine the impact of the resting period in the middle of the experiment, we performed another set of GLMMs. First, as with the previous models, we set viewing time for each trial as the dependent variable and the number of trials (from 0 to 119, coded continuous) and the participants' *Culture* (Austrian as the baseline) as the independent variables. In this model, to see the effect of the resting period more in detail, we further included a treatment coded contrast for *block*, setting 1st to the 60th trials as 'before the rest' (baseline) and 61st to 120th trials as 'after the rest'. The interaction between the three variables and the combinations of the two variables were also included in the model. The structure for the random effects were the same as the previous models. Note that the random slopes were not included again, otherwise the model did not converge. All fix effects and the results are shown in the upper part of Table 5.

As a result, other than the significant main effects of the number of trials, there was a significant main effect *block*, suggesting that, in the Austrian group, the viewing times after the resting period were significantly shorter than those before the rest period. Further, the interaction between *block* and *culture* was also significant. This result indicates that the viewing times after the resting period in the Japanese group were significantly longer than those before the resting period in the Austrian group. To further compare the viewing times between the two level of *block* i.e.,

before and after the resting period in the Japanese group, we performed another GLMM with only the viewing time data from the Japanese participants. Since this model only includes the data from the Japanese participants, the *Culture* was removed from the model structure. Consequently, in this model, the number trials, the *block* (before vs. after), and the interaction of the two variables were included. The results are shown in the lower part of Table 5.

The results showed again the significant main effect of block, suggesting that the viewing times after the resting period in the Japanese group were shorter than before the resting period. Hence, there was an inter-cultural consensus in this trend. Notably, the interaction between the number of trials and block was not significant. Thus, the slopes of the viewing time before and after the resting period were not significantly different.

Additionally, as shown in Figure 5, the data suggested brief recoveries in the viewing times when the participants restarted the viewing task after the resting period. Accordingly, in order to statistically examine if the viewing times recovered after the resting period or not, we performed pair-wise *t*-tests comparing the viewing times between the five trials before and after the resting period for each culture group. The result showed that there were no viewing time recoveries in both culture groups (Japan: $t(37) = 1.157, p = .254, d = 0.19$, Austria: $t(37) = .040, p = .968, d = 0.006$). Hence, the resting period did not have impacts on the viewing times.

Table 5. Fixed effects and coefficients in the GLMMs predicting the viewing time.

Fixed effect	Estimate	95% CI		t value	Pr (> t)
		Lower	Upper		
General impact of repeated exposure, culture and rest					
(Intercept)	1.95	1.82	2.07	31.55	<i>p</i> < .01
Trial	-0.04	-0.06	-0.02	4.50	<i>p</i> < .01
Block	-0.09	-0.12	-0.06	5.80	<i>p</i> < .01
Culture	-0.13	-0.29	0.04	1.52	0.13
Trial * Block	-0.01	-0.04	0.01	0.93	0.35
Trial * Culture	0.00	-0.02	0.02	0.45	0.65
Block * Culture	0.05	0.02	0.08	3.72	<i>p</i> < .01
Trial * Block * Culture	0.02	-0.01	0.05	1.31	0.19
General impact of repeated exposure, culture and rest: Japan					
(Intercept)	1.82	1.81	1.82	28.9	<i>p</i> < .01
Trial	-0.04	-0.03	-0.02	8.69	<i>p</i> < .01
Block	-0.09	-0.04	-0.03	10.68	<i>p</i> < .01
Trial * Block	-0.01	-0.01	0.01	0.02	0.99

Note: The top part of the table shows the fixed effects and the results of the GLMM including the number of trials, culture, and block as fixed effects. In the bottom, the results of the GLMM using only Japanese participants data without the fixed effect *Culture* are shown. Note that the table content is partially identical to the Table 6 in Mikuni, Specker, Pelowski, Leder, & Kawabata. (2021).

Impact of the resting period on the relation of repeated art viewing and beauty rating scores

To assess the impact of the resting period on rating scores, we ran a series of LMMs for beauty rating scores. The model structure was identical to the one in the previous section, using the viewing times as the dependent variable. The fixed effects and results are shown in the upper part of Table 6. Other than the significant main effects of trials and culture, we found a significant

interaction between the number of trials and block. Importantly, this interaction suggests that, in the Austrian group, the slope of beauty rating scores after the resting period was significantly more gradual compared to that before the resting period. In order to further assess if the beauty rating scores in the Austrian group still decreased after the resting period, we performed another LMM only using beauty rating scores after the resting period in the Austrian group. The model structure was identical to the previous model. The fixed effects and results are shown in the lower part of Table 6. The results showed that the main effect of the trial numbers was *not* significant, suggesting that there was no decrease in beauty rating scores in the Austrian group after the resting period. Thus, here, we found that the resting period did counteract the decrease in beauty rating scores in the Austrian group. However, the interaction between the number of trials, block and culture was not significant, meaning that there was not such a trend in the Japanese group.

As with the viewing time analysis, we conducted pair-wise *t*-tests comparing the beauty rating scores between five trials before and after the resting period for each culture group to see if there were any recoveries in beauty rating scores after taking the resting period. As a result, there were no significant recoveries in beauty rating scores in both groups (Japan: $t(37) = 1.20, p = .24, d = 0.20$, Austrian: $t(37) = .36, p = .72, d = 0.05$).

To conclude, resting period did have an impact on beauty rating scores only in the Austrian group.

Table 6. Fixed effects in the LMMs predicting beauty rating scores.

Fixed effect	Estimate	CI		t value	Pr (> t)
		Lower	Upper		
General impact of repeated exposure, culture and rest					
(Intercept)	4.79	4.46	5.10	29.93	$p < .01$
Trial	-0.31	-0.41	-0.20	5.73	$p < .01$
Block	-0.29	-0.63	0.05	1.70	0.09
Culture	0.80	0.50	1.10	5.29	$p < .01$
Trial * Block	0.22	0.07	0.37	2.92	$p < .01$
Trial * Culture	0.01	-0.08	0.11	0.29	0.77
Block * Culture	0.10	-0.03	0.24	1.48	0.14
Trial * Block * Culture	0.02	-0.12	0.16	0.31	0.75
General impact of repeated exposure using only the data after rest: Austria					
(Intercept)	4.50	4.12	4.86	24.06	$p < .01$
Trial	-0.01	-0.14	0.13	0.10	$p = 0.92$

Note: The upper part shows the fixed effects coefficients for each fix effect of the LMM setting the number of trials, culture, and block as the fixed effects. In the lower part, the results of the LMM using only the Austrian participants data are shown. Note that the table content is identical to the Table 7 in Mikuni, Specker, Pelowski, Leder, & Kawabata. (2021).

2.2.3. Discussion

Study 1-2 provided another replication, showing the general decrease in the viewing times and beauty appraisals from repeated art viewing with the updated procedure in a laboratory. Hence, throughout Study 1-1 and 1-2, we repeatedly confirmed that the attention and appraisal towards artworks decreased over trials, even when the participants could note intentionally shorten the total duration of the experiment. Further, as with Study 1-1, the decrease in the viewing times and

beauty appraisals were observed regardless of culture. Hence, indeed, this phenomenon could be inter-individually found. However, note that we found several differences across culture. First, as with the results in Study 1-1, the slope of the viewing times was steeper in the Austrian group compared to that in the Japanese group. Second, the Japanese participants evaluated artworks as more beautiful in general compared to the participants in Austria. Note that this difference was only found in Study 1-2.

Regarding the effect of the resting period in Study 1-2, contrary to the assumptions in the past studies (e.g., Davey, 2005a), it did not counteract/mitigate the decrease in the viewing times in both culture groups. However, it did stop the decrease in beauty rating scores only in the Austrian group. The difference in the effect of resting period for the two culture groups might suggest that the effect of taking a resting period during the repeated art viewing might have different impacts depending on the types of the variables and the individual characteristics of the art viewers. However, note that this difference might be stem to the floor effect. This point is further discussed in the general discussion for Study 1.

Importantly, the constant decrease in viewing time as well as in beauty rating scores even with a resting period might suggest that the typical manipulations such as block-wise experimental design would not bring significant impacts on the behavior and appraisals towards artworks, and thereby, what should be concerned seems to be the general impacts of repeated art viewing.

2.3. General Discussion and Conclusion of Study 1

In Study 1, throughout the two psychological experiments, we aimed to replicate a well-documented phenomenon from museum studies, decrease attention towards presented objects from repeated art viewing to a laboratory setting. Also, we also assessed its impacts on subjective beauty appraisals towards artworks in order to provide further insights to a typical lab-based studies.

Regarding our main research question, *if viewing times/beauty rating scores towards artworks decrease over trials in a lab as well?*, our results consistently showed the decrease in both

variables over trials. Hence, across the two studies, we replicated the results of Robinson (1928). Further, the decrease in the two variables was observed regardless of the culture, supporting an implicit assumption in the past museum studies on this topic—inter-cultural agreements of this phenomenon.

In Study 1, we updated the typical experimental procedure to measure decrease in the viewing time, where the participants could freely view a series of the stimuli without any time restrictions. We tried to assess if the decrease in the viewing times/beauty rating scores is an effect of viewing a series of artworks repeatedly and is not only the by-products of the participants finding efficient strategy to finish the task as fast as possible by fixing the total duration of the experiment. As a results, both Study 1-1 and 1-2 confirmed that the attention/appraisals truly decreased with the updated procedure. Study 1-2 further suggested that the resting period during the repeated art viewing did not refresh especially the decrease in the viewing times. Thus, a common practice to take a short rest(s) in actual museum/gallery exhibitions or the set the mid-rest in the middle of the experimental task in a typical lab-based empirical study seem not to be efficient to refresh/mitigate the decrease in the responses. Accordingly, especially for the empirical studies in a laboratory, not only to set a resting period in an experimental procedure, but also to consider this potential confounding factors in statistical analysis is highly recommended. The results of Study 1 were particularly interesting, since it showed that the participants chose to wait for the next trials in front of a blanc gray screen, rather than to present artworks on the screen monitor. Considering that recent studies also reported that aesthetic experiences seem to require cognitive resources (e.g., Che et al., 2021; Weigand & Jacobsen, 2021), our results speaks to us that, though we enjoy artworks and they bring pleasurable experiences (e.g., Leder et al., 2004), repeated art viewing could be a cognitively-taxing activity. Finally, the two studies consistently found a positive correlation between the viewing times and beauty ratings. These results are in line with the past studies (e.g., Goller et al., 2019; Leder et al., 2016), showing that people tend to look more beautiful artworks longer.

A brief note should be made that, though the past studies reported a *warm up effect* at the beginning of the experimental task—a brief increase in the viewing times (Bitgood et al., 2001; Robinson, 1928), we found the decrease in both viewing times and beauty ratings from the very beginning of the experimental task. The difference in the results might be attributed to the five practice trials in Study 1, where the participants were more familiar with the procedure and prepped for the participations. Nevertheless, considering that these conditions are quite popular for many lab-based empirical studies, researchers should be mindful that even a quite small numbers of the presentations of artworks might also produce such a decrease in responses.

It might also be worth mentioning that, throughout the two studies in Study 1 regardless of culture, the viewing times at the very end of the experimental task were around four to six seconds. Though past studies measuring average viewing times especially in museums reported different duration (Smith & Smith, 2001; Smith et al., 2017; Tschacher et al., 2012), other researchers suggest that basic art assessment e.g., evaluating the pleasantness, liking, art style/contents categorizations and basic emotional could be done in three to four seconds or even less amount of time (e.g., Geday et al., 2007; Locher et al., 2010). Hence, our results are in line with these studies.

We found several cross-cultural differences in our results. There were two general findings regarding this aspect. First, the slope of the viewing times in the Austrian group was significantly steeper compared to that in the Japanese group. Further, the viewing times especially at the very beginning of the experimental task in the Austrian group were longer than those in the Japanese group. Though speculative, this difference could be attributed to the difference in familiarity for the presented stimuli across culture. Since the presented paintings in Study 1 were all Western paintings, it might have been possible that the participants in Austria were more familiar to the presented paintings. Hence, it could have been possible that the Austrian participants made a longer look at the very beginning, since it might have been hedonic due to the past exposure experiences e.g., *mere exposure effect* (Zajonc, 1968). However, the viewing times steeply decreased due to the extreme exposure to the similar stimuli. Second, in Study 1-2, the Japanese participants evaluated

artworks as more beautiful in general than the participants in Austria. Further, importantly, the resting period did mitigate the decrease in beauty appraisal only in the Austrian group. However, such a trend was not observed in the Japanese participants. Regarding the first point, there might be two possible explanations, firstly, the participants in Austria could have made harsh evaluations for the presented artworks compared to the Japanese participants, secondly, there might be a response bias across cultures. For example Marković et al. (2016) recruited the participants from Japan and Serbia, and compared the evaluations towards Japanese and Serbian architectural objects between the two cultures. As a result, they found that the Japanese participants evaluated the presented objects as more beautiful and pleasant in general compared to the Serbian participants. Thus, for the future studies, it would be worth examining such a response bias across cultures. Regarding the second point, there might be also two possible explanations, firstly, the resting period did have an impact on the decrease in beauty rating scores only in the Austrian participants, and secondly, beauty rating scores after the resting period in the Austrian group met a floor-effect. Considering especially that, the beauty rating scores in the Japanese group were generally higher than in the Austrian group, it could have been possible that beauty rating scores in the Austrian groups had become low enough, resulting in no further decrease after the resting period. In order to examine which explanation is plausible, future studies are necessary. For example, not only including one single resting period, but multiple ones in the experimental procedure would make it possible to see if the resting period has power to stop the decrease in the Austrian group or not. Nevertheless, we hope that our results, showing the inter-cultural consensus in the decrease in viewing time and appraisals towards art, inspire an avenue for the future studies and add knowledge on the art viewers behavior.

Finally, there were of course several limitations in Study 1. Though two psychological studies in Study 1 repeatedly showed the general decrease in the viewing times and beauty rating scores as a function of repeated art viewing, considering that there have been less than a dozen of empirical studies on this topic (see also Bitgood, 2009a), further replications are necessary in the

future. Importantly, not only performing multiple replications, but conducting a meta-analysis including the past findings on this topic is highly expected, hence, only by doing so, we could see a *real* effect of the repeated art viewing. Further, so far, only Study 1 in the present dissertation measured art appraisals over extreme exposure towards artworks, and we only asked to evaluate beauty of the paintings to the participants. Though beauty is known as a core appraisal especially for art laypersons and also as a metric which shows strong positive correlations with other appraisals such as liking, interest (e.g., Jacobsen et al., 2004), it might be interesting to employ other ratings and to see what kind of art assessment are immune to the massive exposures. Importantly, the main aim of Study 1 was to translate the impact of repeated art viewing on viewers' attention and evaluation towards art from museum studies to a lab-based setting. Hence, we did not examine the effect of specific cause of this effect. Though we do acknowledge that it is especially important to examine the cause of this effect, we believe that the multiple replications of this phenomenon in a laboratory is an important first step, since if the effect is not observed in a laboratory, any further examination of its causes in a laboratory seems not to be warranted. Nevertheless, in Study 1, some factors were controlled such as physical fatigue and object competition, since the participants were sitting in a laboratory and only one painting was presented at a time. Furthermore, as discussed by Robinson (1928), the most likely reason of the decreased viewing time and appraisals towards art in Study 1 would be satiation, considering that the participants only viewed a series of paintings, which could be seen as the homogeneous stimuli presentations. In the following Study 2, two studies explicitly investigating the way to counteract this phenomenon as well as its possible causes are presented.

3. Study 2—Empirical investigation on the possible causes and how to counteract the attention decrease phenomenon in lab-based setting

In Study 2, two experiments in a laboratory setting are presented. Through this section, we aim to empirically examine a possible cause of the decreased attention from repeated art viewing as well as how to counteract it. Specifically, in Study 2-1, we tested if the ever-presented way to counteract this phenomenon, providing information about the artworks, could actually stop the decrease in viewing time with an updated procedure. In Study 2-2, we tested if one of the factors that might cause the decreased attention phenomenon towards objects, *habituation*, is an actual cause as well as if the *novelty effect* in the stimulus sequence, which is known as a way to counteract habituation process, can stop/mitigate the decrease in viewing time over repeated art viewing. The detailed introduction, purpose, methods, and results of each study will be respectively presented in the following sections.

3.1. Study 2-1: Examining impact of information about artworks on attention towards them

Published work

Mikuni, J., & Kawabata, H. (2019). Effect of artwork information on behavior and evaluation of art: the viewer perspective. *Studies in sociology, psychology, and education: inquiries into humans and societies*. 88, p37-51.

https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN0006957X-00000088-0037

My Contribution:

I acted as the first author of the paper. With HK, we developed the core concept of the study idea. I built the experiment, collected the participants, and analyzed the results. I was the main writer of this paper and elaborated the manuscript with HK.

Though little knowledge has been accumulated on how to counteract the decreased attention phenomenon from repeated art viewing, in fact, there have been a few studies, showing that decreased attention towards artwork can be avoided by presenting information about the exhibits (e.g., Bitgood et al., 1992; Porter, 1938; Robinson, 1928).

For example, in the file study of Robinson (1928), he examined the impact of providing information—the effect of pamphlets—on the decreased attention towards art over time in an actual museum exhibition setting. He measured the total amount of time the visitors spent in the exhibition space, the total number of exhibits they stopped, and the viewing time for each exhibit as a proxy of their attention towards the exhibits. He randomly assigned participants into two conditions i.e., the control and experimental condition. In the control condition, participants freely viewed exhibits as they wished. In contrast, in the experimental condition, though the participants could also freely view the exhibits, they were provided a pamphlet which describes the locations, titles and a brief description for 20 artworks. As a result, the participants in the control condition stayed in the exhibition for 17 minutes on average, while those in the experimental condition did for 28 minutes, showing that the participants in the experimental condition stayed longer in the exhibition space. Further, the average number of exhibits seen by the participants was 30 in the control condition, and 46 in the experimental condition, suggesting that providing information about the exhibits also increased the total number of exhibits seen. Importantly, Robinson compared the transition of the viewing time between the two conditions and found that participants in the control condition showed decrease in viewing times as their visits progressed, while those in the experimental condition did not show any decrease. Further, the participants in the experimental condition viewed each artwork longer in the last half of their visit compared to the first half of their visit. Thus, this study showed that providing information of the presented artworks is a way to counteract the decrease in the viewing times. Assuming that the explanations and/or information of artworks will help the art viewers to process more deep assessment for the artworks such as meaning making and/or making a relation between the artworks and the person etc. (Leder et al.,

2004), it seems plausible that this manipulations might made the art experience more enjoyable, which might have resulted in no decrease in the viewing times.

Other studies which examined the impact of giving pamphlets/handouts also found that they can maintain the attention towards objects (Bitgood et al., 1992; Porter, 1938). However, the results showed different trend than in Robinson (1928). Bitgood et al. (1992), for example, provided handouts to the visitors in the Predator Building of the Birmingham Zoo and the Alabama Cave exhibit at the Anniston Museum on Natural History. In the handouts, 10 questions were included to stimulate the visitors to read the labels of the exhibits. In this study, four metrics were measured, namely, if the visitors took the handout or not, the time duration when the visitors read the label, the viewing times for the presented objects, and if the visitors actually used the handouts during their visits or not. In total, they observed more than 200 visitors in the Birmingham Zoo and 95 visitors in the Anniston Museum on Natural History. In the Birmingham Zoo, the visitors who actually used the handouts during their visit made longer looks to the label in the exhibition compared to those who did not use them. However, this was especially salient at the very beginning of the exhibition, and the visitors did not take a fine look at the labels anymore at the end of the exhibition space. Importantly, they also found the differences in the viewing times towards the exhibits in the two groups. The visitors who used the handouts did view each exhibit longer compared to those who did not use it. However, this trend was found only at the very beginning of the exhibition, and both group did show the decrease in the viewing times. Note that Porter (1938) also found the same trend as in Bitgood et al. (2012).

In sum, while Robinson (1928) found that the manipulation of providing information on the exhibits can be effective to maintain the viewers' attention towards exhibits over the course of exhibitions, other studies only found this effect at the beginning of the exhibition space (Stephen Bitgood et al., 1992; Porter, 1938b). Hence, due to the incompatible results from the past studies, the impacts of providing the information about the exhibits on the decreased attention remains still unclear. Of course, the difference in the results could also be attributed to the differences in the

quality of the information presented in the exhibition places. Nevertheless, to further assess the general impact of providing information could be meaningful, considering that most of the museums provide their visitors a (free) pamphlet as well as an audio guide, which might be based on an implicit assumption that it can maintain the attention/interests towards the exhibitions.

However, the aforementioned studies had some methodological limitations. First, especially in Robinson (1928), the duration when the participants in the experimental condition read the information in the pamphlet seemed to be also included in the viewing time. Thus, it was possible that the participants in the experimental condition stayed longer since they simply spent additional time to read the texts in the pamphlet. Further, it could be possible that participants, in the experimental condition, actually did not pay attention towards the exhibits at the end of the exhibition, but only read the texts written in the pamphlet provided in front of an object, which might have resulted in observing no decrease in viewing time. Accordingly, it is still unclear if this manipulation had an actual impact on the decreased attention of the viewers or not. In order to investigate the effect of providing information of the exhibits in detail, it is necessary to separate the time when the participants read the information about the artworks and when they view each artwork. Second, though Bitgood et al. (1992) did separate the viewing times towards exhibits and label, the participants could freely decide whether to use the handouts or not as well as which labels they read. Thus, each participant was exposed to different numbers and types of texts, which makes it difficult to see the true effect of presented information.

In the present study, given the methodological limitations in the past studies, we updated the experimental procedure used in past research in order to examine the impact of providing information of the artworks on the decreased attention over repeated art viewing. Specifically, in the present study, half of the participants were assigned to the control group where they read texts which are irrelevant to the presented artworks before they viewed a series of artworks. On the other hand, the other half of the participants were assigned to the experimental group where they read texts which are explaining the historical/cultural background of the presented artworks before they

viewed the same artworks as in the control group. With this procedure, it is possible to separate the time when participants read the texts and when they view artworks and to provide the same amount of information to all participants, which makes it possible to systematically assess the impact of providing information about the artworks on viewing time as a proxy of the attention towards the stimuli.

Further, in the present study, we were more interested in the general impact of providing information about the presented artworks, rather than the effect of a specific type of information/texts. Consequently, we used paintings and captions from two different exhibitions with different set of artworks and exhibition structures as the stimuli. Specifically, we chose two exhibitions which were held in the same museum, but with very different structures and contents—a period-type exhibition (period condition) where the chronological history of the paintings was presented, and a category-type exhibition (category condition) where paintings including different types of art category from one art-school were presented. The participants viewed paintings from either one of the exhibition conditions (i.e., period condition and category condition) and the participants in each condition were further randomly assigned either to the control/experimental group. If the manipulation of providing information about artworks is truly effective on the decreased attention phenomenon towards artworks over repeated art viewing, it is expected that the viewing time in the control condition would decrease as a function of the trial, while the ones in the experimental condition would not show any decrease regardless of the exhibition conditions.

3.1.1. Method

Participants

In the present study, participants were first randomly assigned to either the period condition or the category condition. In the period condition, participants viewed a series of paintings from an exhibition “400 Years of European Masterpieces from the State Hermitage Museum” (see *Stimuli and Apparatus* for the detailed description). In contrast, in the category condition, participants

viewed a series of paintings from another exhibition “The Birth of Impressionism – Freedom in Painting: Masterpieces from the Musée d’Orsay” (see also the detailed information in *Stimuli and Apparatus*). Both exhibitions were held at The National Art Center (Tokyo) in 2012 and 2014 respectively. In the period condition, 46 native Japanese psychology students participated (28 female, $M_{age} = 24.02$, $SD = 8.62$), and in the category condition, 46 native Japanese psychology students participated (25 female, $M_{age} = 25.84$, $SD = 11.75$). In each condition, participants were further divided into the control group, where they read irrelevant texts to the artworks to be presented, and the experiment group, where they read relative texts to the artworks to be presented. All participants had normal vision or corrected-to-normal vision, no academic backgrounds in art, and received 1,000 Japanese yen, approximately \$ 10 US) for their participations. The study was approved by the local ethical committee of Keio University, Japan.

Stimuli and Apparatus

The stimuli consisted of 160 images of paintings exhibited in two museum exhibitions held in Japan. Half of the stimuli (80 images) were from a museum exhibition “400 Years of European Masterpieces from the State Hermitage Museum” for the period condition, and another half of stimuli from “The Birth of Impressionism – Freedom in Painting: Masterpieces from the Musée d’Orsay” for the category condition. The paintings as well as the texts presented in the exhibition from Hermitage museum (Saint Petersburg, Russia) were only shown to the participants in the period condition, and those presented in the exhibition from Orsay museum (Paris, France) were only shown to the participants in the category condition.

The exhibition from the State Hermitage Museum, especially focused on chronologically presenting the historical development of the Western paintings. The small sections in the exhibition were divided into four different chronological periods (i.e., 16th, 17th, 18th, and 19th century), which consist of four stimulus-blocks in the period condition. Each stimulus block contained 20 artworks, and thus, in total, 80 paintings were included. On the other hand, the exhibition from the

Musée d'Orsay, focused on introducing the paintings from one specific art school/style:

Impressionism. The small sections within the exhibition were divided into four different categories of paintings (i.e., history paintings, landscape, portrait, and genre paintings), which consist of four stimulus-blocks in the category condition as well. Each stimulus block contained 20 artworks, and thus, in total, 80 paintings were included. The horizontal dimensions of all images used in the present study was fixed at 600 pixels.

As mentioned, in both period and category conditions, the paintings were divided into four chronological periods/categories with 20 paintings for each. In the control group, before the participants started viewing 20 artworks included in each stimulus block, they read texts irrelevant to the artworks. In contrast, in the experimental group, they read texts related to the artworks. Note that in the experimental group, the provided texts were different depending on which condition (either the period or category conditions) the participants were assigned. For example, if a participant was assigned to the experimental group in the period condition, s/he read the texts related to the paintings from the Hermitage museum, while that in the category condition read the texts related to the painting from the Orsay museum. Further the texts used in the control group were the same across the two conditions. Hence, in the present study, a total of 12 texts stimuli were used: four irrelevant texts in the control condition, four captions in the experimental group in the period condition, four captions in the experimental group in the category condition. The irrelevant texts used in the control group were collected from the Aozora Bunko. None of the texts contained descriptions of art, and they were reviews of either medicine, literature, or history. The caption texts in the experimental group were retrieved from the official catalogues from two distinct exhibitions respectively. The four captions used in the experimental group of the period condition contained the descriptions about the historical/cultural background of the presented artworks in each period, famous artist and common art styles used, and popular contents to be drawn in that period. In contrast, the four captions used in the experimental group of the category condition explained the importance and uniqueness of the category of the paintings in each block.

For example, in the second stimulus block in the category condition, 20 landscapes were presented. The caption presented before the second stimulus block was mainly explaining the importance as well as the historical background of the landscape paintings i.e., landscapes were just the part of background in traditional paintings but emerged as a major content of paintings in the 19th century on. All text stimuli were printed and afterwards laminated, so that the experimenter can provide the texts stimuli to the participants.

The experiment was controlled via PsychoPy 3.0 (Peirce, 2007). Prints of paintings were presented individually on the center of cathode ray tube monitor (Trinitron, CPD-G420, SONY) with a screen resolution of 1280×960 pixels. All stimuli were presented with a gray background. Each participant sat in front of the monitor at a distance of about 80 cm.

Procedure

Each participant viewed 80 paintings in total. The total 80 paintings were divided into four blocks with 20 artworks for each. In both period and category conditions, the presentation order of the paintings was fixed as the same order in actual museum exhibitions.

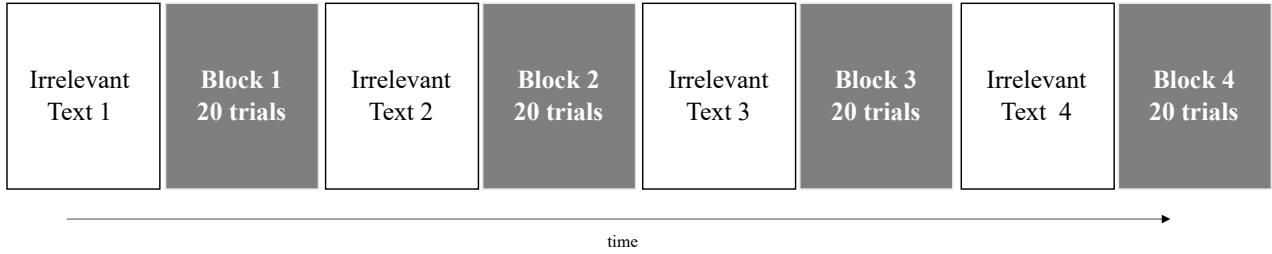
The experiment was conducted individually in a sound-proof room in Keio University. After the experimenter welcomed the participants in the laboratory, the experimenter explained that there are going to be several blocks in the experiment where a series of artworks are going to be presented, and before starting each block, the participants would read texts provided by the experimenter. They were also instructed that there are no time-limits for reading the texts as well as for viewing a series of artworks. Further, the participants were told that they do not have to read the sentences aloud but silently and were asked to call the experimenter after they finish reading the texts. In order to avoid any anticipation effects from the participants, the exact number of blocks as well as paintings included in the block/experiment was not told to the participants.

Before each block started, the experimenter provided a laminated text to the participants—in the control group, an irrelevant text was provided, while in the experimental group, a caption

was provided. After the participants finished reading all texts, the participants called the experimenter and gave the text back. When the experiment was out from the testing room, the participants started the viewing task. After the participants finished viewing all 20 paintings included in one block, a gray screen was presented on the screen monitor. When the participants reached the gray screen, they were told to call the experimenter again in order to get another text. They repeated this flow until they finished the fourth stimulus block. Figure 6 shows the task flow in the control and experimental group in the present study.

In each stimulus block, participants were asked to press a key to start each trial. Following the key press, a painting was presented at a time on the center of the screen monitor, and they could view each artwork as long as they wished. When they pressed the same key again, the painting was replaced by a 9-point Likert-type scale, asking ‘How beautiful the painting was’ (1, *not at all beautiful* to 9, *very beautiful*). Though the appraisals towards the artworks were not the main dependent variable in the present study, in order to set the motivation to view the artworks the same across participants, we explicitly asked the participants to rate the beauty of each artwork. Further, after the participants rated the beauty of the paintings, they were asked to rate the familiarity for the given painting in the same manner as the beauty rating (1, *not at all familiar* to 9, *very familiar*). After the familiarity rating, the screen went back to the screen to start the new trial, and they repeated this flow until all 20 paintings were presented. Note that the presentation order of the paintings was identical to the one in the original exhibition.

Control condition



Experimental condition

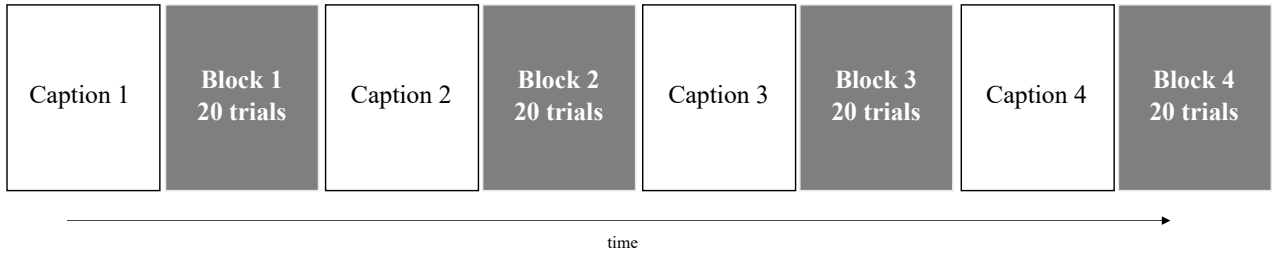


Figure 6. Task flow of the control and experimental condition

3.1.2. Results

In the present study, as in art viewing in actual museum exhibitions, there was no time-limit for viewing each artwork. Thereby, though the experimenter told the participants not to take rest while a painting was shown on the screen monitor, it was still possible that the participants took a rest without paying attention towards the artworks or slept in the experiment room. In order to avoid conflating the results by those data, we first calculated the mean viewing time across participants, set the viewing time ± 3 SD from the mean viewing time of each participant as the outliers, and excluded those data from the analysis below. As a result, in the control group, 2.45% of the data was excluded, and in the experimental group, 2.07% of the data was excluded from the analysis.

The analysis of the viewing time

In order to assess the transition of the viewing time in both period and category condition, we performed the same analysis separately for the two conditions. Thus, in the following section, we report the results of the viewing times separately for the period and category condition.

The transition of the viewing time for the control group and the experimental group in the period condition

The mean artwork viewing time ranged between 2.43 and 18.31s for the control group, and from 2.65 and 41.29s in the experimental group; the *SDs* ranged from 1.15 to 7.44 for the control group, and 1.22 to 20.66 for the experiment group. The mean viewing time across all participants was 8.70s (*SD* = 6.57) for the control group, and 13.17s (*SD* = 11.50) for the experiment group.

In order to assess the general trend of viewing time for the control and the experimental group as a function of the trials, first, we calculated the mean viewing time for each trial in the control and the experimental group. The mean viewing times for each trial in the two groups are presented in Figure 7. As can be seen in the figure, mean viewing times in both groups seem to decrease over the trials. Further, in order to statistically assess the trend of viewing time, I computed the Kendall's rank correlation coefficient using the mean viewing time for each trial in the control/experimental groups and the number of trials. As a result, the correlation scores showed that the viewing times in both groups significantly decreased as a function of trials (control group: $\tau = -0.34, p < .001$, experimental group: $\tau = -0.31, p < .001$). Further, in order to investigate if there is a significant difference in the mean viewing time for the control and experimental condition, we conducted a two-sample *t*-tests using the mean viewing time for each trial from all participants in both groups. As a result, the mean viewing time in the control group was significantly shorter than those in the experimental group ($t(44) = 2.11, p < .05$, 95% CI[0.071s, 8.920s], Cohen's $d = 0.61$).

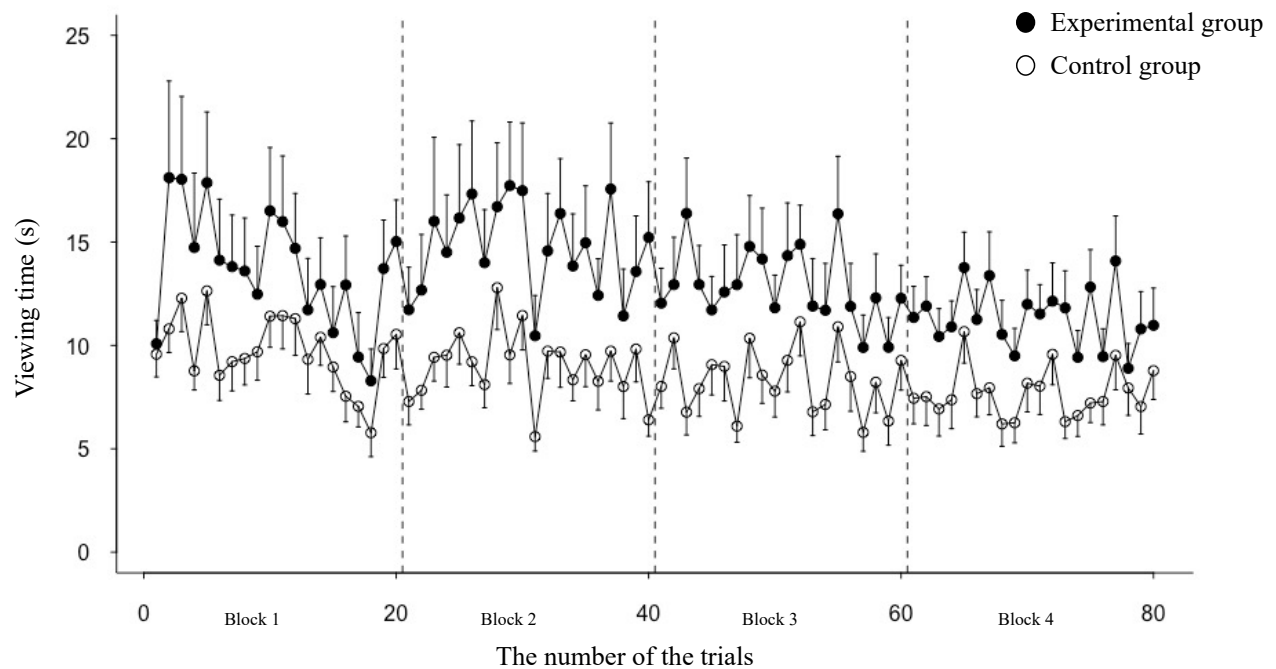


Figure 7. Mean viewing time for the control and experimental groups in the period condition. Error bars indicate the Standard Error (SE) for each point. The black points represent the result of the experimental group, and the white points represent that of the control group. Vertical lines highlight the points where the stimulus block changes. Note that the figure contents are identical to Figure 3 in Mikuni & Kawabata (2019).

In the present study, participants first read the texts before they started viewing 20 paintings, and they repeated this procedure for four times. Hence, in order to assess the transition of the viewing time from a block to another more in detail, I computed the Tau-U scores (Parker et al., 2011) using the viewing time included in the two blocks next to each other, in order to compare the viewing time between e.g., block 1 vs. 2, block 2 vs. 3. The Tau-U scores comparing all possible pairs of the blocks next to each other are shown in the upper part of Table 7. As shown in Table 7, in the control group, the viewing time included in the blocks next to each other were not significantly different in all possible combinations of the stimulus block. However, considering the

Kendall's Tau-U score for the control group did show the decrease in the viewing time in general, viewing time in the control condition seems to decrease slowly over the repeated art viewing. In contrast, in the experimental group, viewing time first increased from block 1 and 2, and started decreasing over the following blocks.

Table 7. Tau-U scores for the control and the experimental groups in both period and category condition.

Condition/group	Block comparison	Tau-U	90% CI		Z value	p
			Lower	Upper		
Period condition						
Control group	1 vs. 2	-0.13	-0.430	0.18	-0.68	0.50
	2 vs. 3	-0.29	-0.59	0.02	-1.54	0.12
	3 vs. 4	-0.21	-0.51	0.09	-1.14	0.26
Experimental group	1 vs. 2	0.38	0.08	0.68	2.06	0.04
	2 vs. 3	-0.47	-0.77	-0.17	-2.54	0.01
	3 vs. 4	-0.43	-0.73	-0.12	-2.30	0.02
Category condition						
Control group	1 vs. 2	-0.63	-0.93	-0.32	-3.38	0.001
	2 vs. 3	-0.55	-0.85	-0.24	-2.95	0.003
	3 vs. 4	0.34	0.03	0.64	1.81	0.07
Experimental group	1 vs. 2	-0.56	-0.86	-0.25	-3.00	0.003
	2 vs. 3	-0.68	-0.98	-0.37	-3.65	0.00
	3 vs. 4	0.27	-0.04	0.57	1.43	0.15

Note. The upper part of the results represents the Tau-U scores for the period condition, and the lower part of the results represents the Tau-U scores for the category condition. Note that the contents of the table is identical to Table 2 in Mikuni and Kawabata (2019).

Transition of the viewing time for the control group and the experimental group in the category condition

As with the period condition, we first calculated the mean viewing time for the control and experimental groups. The mean artwork viewing time ranged between 2.13 and 25.00s for the control group, and from 3.87 and 32.93s in the experimental group; the *SDs* ranged from 1.01 to 10.33 for the control group, and 1.07 to 11.45 for the experiment group. The mean viewing time across all participants was 10.50s (*SD* = 8.12) for the control group, and 11.81s (*SD* = 9.91) for the experiment group.

Again, to investigate the general trend of the viewing time in both groups, we calculated the mean viewing time for each trial in the control and the experimental group. The mean viewing times for each trial for the two groups are presented in Figure 8. As can be seen in the Figure 8, similar to the period condition, the viewing times in both groups seem to decrease over the trials. Further, we also computed the Kendall's rank correlation coefficient using the mean viewing time for each trial in the control/experimental groups and the number of the trials. As a result, the viewing times in both groups significantly decreased as a function of trials (control group: $\tau = -0.49, p < .01$, experimental group: $\tau = -0.41, p < .01$). In addition, in order to examine if there is a significant difference in the mean viewing time for the control and experimental condition, I conducted a two-sample *t*-test using the mean viewing time for each trial from all participants in both groups. In contrast to the result from the period condition, the result showed that the mean viewing time did not differ across the two groups ($t(44) = 0.58, p = .57$, Cohen's $d = 0.18$). Further, as shown in Figure 8, the data suggests that the viewing times of the first block in the experimental group are longer than those in the control group. Consequently, we performed a two-sample *t*-test, using the mean 20 viewing times included in both groups as the dependent variable. As a result, the mean viewing time of the experimental group in the first block was significantly longer than that in the control group ($t(34.67) = -4.24, p < .001$, Cohen's $d = 1.34$).

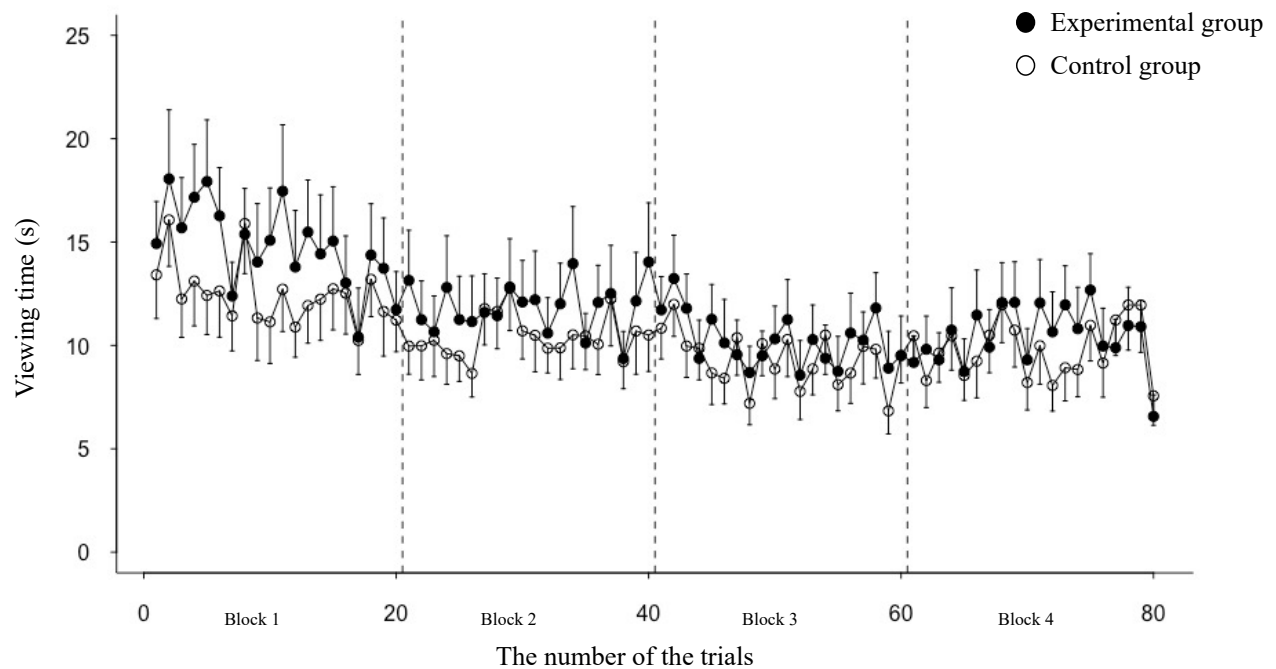


Figure 8. The mean viewing time for the control and experimental groups in the category condition. Error bars indicate the Standard Error (SE) for each point. The black points represent the result of experimental group, and the white points represent that of control group. Vertical lines highlight the points where the stimulus block changes.

Further, again, to assess the transition of the viewing time from a block to the following block, we calculated the Tau-U scores using the viewing time included in the two blocks next to each other. The Tau-U scores comparing all possible pairs of the blocks next to each other are shown in the lower part of Table 7. As can be seen in Table 7, in the category condition, there was a similar tendency in the control and experimental groups. In both groups, the viewing time significantly decreased from block 1 to block 2, and from block 2 to block 3. However, the viewing time stopped decreasing from block 3 to block 4. Considering that the viewing times in both groups stopped decreasing at the very end of the experiment, it might be possible that the viewing times met the floor effect in both groups.

3.1.3. Discussion

In Study 2-1, we examined if the ever-presented way to counteract the decreased attention towards artworks over time, namely, the effect of providing the information about the presented artworks could stop/mitigate the decrease in the viewing times. Overall, in both groups the viewing times decreased as a function of the trials. Hence, the results were replicated the results of Bitgood et al. (1992) and Porter (1938), but did not support the findings from Robinson (1928).

The incompatible results could be attributed to the methodologies. Given that, especially in Robinson (1928), the viewing times when the participants read the information about the exhibits was included in the viewing time analysis, it could be possible that the viewing time in the experimental condition did not show any decrease, since the participants in the study might just have spent extra time to read the written information especially at the end of the observation, which could result in no decrease in the viewing time. Consequently, together with the results from the present study, we can conclude that the manipulation to provide information about the presented objects might not function as a way to counteract/mitigate the decreased attention phenomenon.

Nevertheless, giving the information of artworks did increase the attention in the period condition and category condition. Note that this effect was present in the category condition at the very beginning of the experimental task. As shown in Figure 8, in the category condition, though the viewing times at the beginning of the task in the experimental group was higher than in the control group, the two-sample *t*-test using the mean viewing time for 80 trials showed that there was no difference in the viewing times between the two groups. This might suggest that the decrease in the viewing times in the experimental group was even larger than in the control group. Considering that providing information of the artworks could enhance more deeper processes e.g., meaning making, self-reflection, this manipulation could also increase the amount of information process during art viewing. Given that Bitgood (2009a) also listed *information overload* defined as “inability to process the number of simultaneous inputs or successive inputs presented too quickly” (Bitgood, 2009, p.106), presenting the information about the artworks on top of a series of artworks

could require additional cognitive process, which could even hasten the decrease in the viewing times. In the present study, the amount of cognitive load between the control and experimental condition could be assumed to be equal, since in both conditions, participants read similar number of texts before viewing the artworks. In the future studies, it might be interesting to empirically test further if the *information overload* could modulate the decrease attention. For example, one might test this effect with a similar procedure as in the present study with an additional condition where the participants spend some time without loading additional cognitive effort e.g., doing nothing in a room while the other two groups reading texts, and compare the slope of the viewing time between the three conditions. If the slope of the viewing time in this additional condition is more gradual compared to the other two condition where they read similar number of texts before viewing artworks, this might, though indirect, suggest that higher cognitive load would hasten the decrease of the viewing time.

A brief note should be made on the different results in the period and category conditions. As a result, both the control and experimental group did show the decrease in viewing time. However, there was a slight difference between the two conditions. In the period condition, viewing time in the experimental group was significantly longer than those in the control group, while in the control condition, there was not such a tendency. Considering the contents of the captions between the two conditions were different, it might be possible that depending on the quality/content of the text information, the effect of providing information about the artworks might be different. However, given that the longest mean viewing time in the experimental group in the period condition was 41.29s, while that in the control group was 18.61s, it could be possible that there were some participants who generally looked at artworks longer, which resulted in the significant longer viewing time in the experimental group in the period condition.

3.2. Study 2-2: Examining one of the possible causes of the decreased attention towards artworks, ‘*habituation*’

Manuscript in preparation

Mikuni, J., Specker, E., Leder, H., Rosenberg, R., & Kawabata, H. (*in prep*). The Power of Something New: The Novelty Effect on Decreased Engagement towards Artworks.

My Contribution:

I acted as the first author of this work. With HK, we developed the core concept of the paper. I built the experiment, collected the participants with ES, and analyzed the results. I am the main writer of this draft, and elaborating the contents with feedback from ES, HL, RR, and HK.

Study 2-1 showed that one of the ever-presented ways to counteract the decreased attention phenomenon—providing the information about the presented artworks—could not mitigate the decrease in the viewing times. Thus, it still remains unclear how to counteract decreased attention from repeated art viewing.

In Study 2-2, I aim to examine, firstly, if another suggested cause of decreased attention phenomenon—*habituation (satiation)*—attributes to the decreased viewing time and secondly, if introducing novelty—known as a way to counteract the habituation process—in the presented stimulus sequence could counteract the decreased attention. In order to introduce the theoretical background of Study 2-2, in the following section, we briefly summarize the past findings regarding habituation as well as the way to empirically detect the habituation process and how to counteract it.

Before starting the introduction of Study 2-2, it might be helpful to make a note on the terminology in order to make the arguments clearer. Past museum studies on the decreased attention phenomenon as well as reviews which summaries its possible causes used the term *satiation* rather than *habituation*. In the literature, satiation is defined as the attention decrease due to the homogenous stimuli presentation for multiple times. However, in past psychological studies—especially in the field of psychology of learning—, satiation is defined as a decrease in the consumption of ingestive stimuli such as food/water over the repeated consumption of that stimulus, while habituation is defined as a decrease in responsiveness to a stimulus or stimuli that is/are presented repeatedly or for a prolonged time (McSweeney & Roll, 1998; Thompson & Spencer, 1966b). Accordingly, the cause of the decreased attention towards art is more accurately described as habituation than satiation. thus, in the following section, we use the term habituation as a key term to cause the attention decrease phenomenon in the present study. However, when reviewing the past literature on attention decreased phenomenon, the term satiation is kept in order to avoid any confusions on the terminology. To the best of our knowledge, none of the past studies

or reviews on this topic mentioned the studies of habituation, and thereby, using the term habituation in the review may paint a confusing picture.

Past studies on habituation

In Bitgood (2009a), satiation is listed as one of the possible causes of the attention decrease phenomenon towards artworks. The lab-based study in Robinson (1928) is argued as a study where he successfully investigated the effect of satiation, since the participants were sitting in the lab—no physical fatigue, and only one painting was presented at a time—no object competition. Thus, the decrease in the viewing time could be attributed to the repeated presentation of the homogenous stimuli.

In fact, in the field of psychology, a quite similar concept, habituation, has been extensively studied especially in the field of psychology of learning. Habituation is documented from single-celled animals to primates for a variety of responses (Harris, 1943), and is considered to be one of the most elementary forms of behavioral changes (Humphrey, 2013; Thompson & Spencer, 1966). For example, Gardner (1968) showed that when a vibration was presented to earthworms, a strong reflective response to this stimulus was observed at first. However, after several presentations of the same stimulus, responses became weaker, and eventually, they did not show any responses to the stimulus. Hence, habituation can be defined as the decrease in response after the repeated presentation of a homogenous stimulus. This result was not from human participants, and one might question if similar results can be obtained for humans or not. Interestingly, a decrease in responses towards the same or similar stimulus/stimuli has also been reported when human participants evaluate visual stimuli, known as ‘*affective habituation*’ (Klorman, 1974; Leventhal et al., 2007; Mangelsdorff & Zuckerman, 1975; Morii et al., 2017). As can be seen in those examples, even though the terminology in the past museum and psychological studies differed, both fields have examined a quite similar effect, the effect of repeated presentation of homogenous stimulus/stimuli.

Regarding habituation, knowledge has accumulated, and, importantly, it is known that, firstly, habituation can be distinguished from other decrementing processes such as physical fatigue, as it can be interrupted by a change in the presented stimulus, which is known as “*dishabituation*”, and secondly, habituation process could be mitigated/delayed by introducing novelty in the presented stimulus sequences. Regarding the first point, although past lab-based studies (Bitgood et al., 2001; Robinson, 1928) as well as the studies in the present dissertation seem to investigate the effect of habituation by ruling out some confounding factors, such as physical fatigue, stress, and objective competition, it is still unclear if the viewing time for each picture decreased because of the repeated presentation of homogenous stimuli, or because of the accumulated physical tiredness during the experiment. Especially in the computer-based laboratory studies in this dissertation, physical fatigue could be also accumulated by looking at a computer screen for a while. Further, none of the studies directly measured the metrics representing the level of fatigue and stress. Thus, the conclusion is rather speculative. If we could observe habituation-dishabituation process in the viewing time towards artworks, this could be the empirical evidence, suggesting that habituation is at least one of the causes of the decreased viewing time phenomenon. Regarding the second point, if habituation process is one of the causes of the decreased viewing time, we could apply the knowledge on how to mitigate/delay habituation process, which might suggest a way to counteract decreased viewing time towards artworks.

In the following section, we further review how to observe habituation-dishabituation process, and the past studies showing the novelty effect as a way to counteract habituation.

Methodology to observe habituation-dishabituation processes

Groves and Thompson (1970) studied how to distinguish habituation from other decrementing process by causing dishabituation. In their experiment, they presented a habituation stimulus (i.e., loud auditory stimuli) to infant rats for multiple times. In the experimental condition, a novel stimulus (i.e., flashing light) was presented in the latter stage of the experiment, while in

the control condition, the novel stimulus was not presented. As a result, the response to the habituation stimulus (i.e., startle response) decreased as a function of the presentation time in both conditions. However, the novel stimulus produced an increase in response to the immediately following habituating stimulus only in the experimental condition. The response recovery is so-called *dishabituation*, and it is important to observe the dishabituation process for differentiating habituation from other decrementing processes such as physical fatigue—if the rats were physically fatigued, they would have not shown the recoveries in their responses. In sum, when an organism is exposed to novel stimuli in the environment after they are exposed to the habituation stimulus multiple times, the habituated responses might be recovered, and this recovery indicates the presence of the habituation process.

Novelty effect as a way to mitigate/delay the habituation process

Leventhal et al. (2007) showed their human participants one picture from the IAPS (International Affective Picture Set: Bradley & Lang, 2017) 20 times and asked them to rate how pleasurable the picture was. As a result, the pleasure rating score towards the one image decreased over the repeated exposure to the stimulus, so-called *affective habituation process*. In their follow-up study, they used four different IAPS pictures, and presented each of those four pictures 5 times—to reach the same total of 20 presentations. Note that in the second experiment, the images were always presented in the same sequence (1,2,3,4) rather than a random sequence of 20 presentations. Hence, in the follow-up study, the novelty within the stimulus sequences was kept higher compared to the first experiment. As a result, though they found the decrease in pleasure rating scores in both studies, the slope of the pleasure rating scores was significantly more gradual in the follow-up study compared to the first study. They concluded that higher novelty in the presented stimulus sequences can counteracted affective habituation process. In other words, the stimulus sequences with high stimulus variability should mitigate/delay habituation process due to this *novelty effect*.

Present study

In Study 2-2, we will apply those findings to detect the underlying mechanism as well as the way to counteract decreased attention from repeated art viewing. Specifically, we aimed to assess two main questions: 1. Does habituation process cause the decreased attention from repeated art viewing? 2. Does higher stimulus novelty in the stimulus sequence counteract/delay the decreased attention over repeated art viewing? In the following section, we introduce how to assess each question as well as hypotheses corresponding to each question.

Question 1: Does habituation process cause the decreased attention from repeated art viewing?

First, we assessed whether the habituation process is a *real* cause of the decreased attention phenomenon from repeated art viewing or not. To this aim, we examined if there would be recoveries in the viewing time after the introduction of novel stimuli—causing dishabituation—to detect the presence of the habituation process.

Specifically, we prepared two between-subject conditions wherein the participants viewed the same images of paintings in different orders, namely Periodic Categorization (described as PC condition in the following texts) condition and Categorical Categorization (described as CC condition in the following texts) condition. In the PC condition, the paintings were presented in a chronological order (see also the *Stimuli* section for the detailed descriptions of the stimulus/stimulus selection criteria). For example, the paintings from early period (i.e., the 16th century) were presented at the beginning, following the next period (i.e., the 17/18th century), and later period (i.e., the 19th century) were presented at the end of the stimulus sequence. The stimulus sequence of the PC condition had a variety of contents/styles, since when ordering by period, each period contained a variety of contents/styles which lead to a high novelty among stimuli. In contrast, in the CC condition, the paintings were presented along with the categories of the paintings, where participants viewed the same category of the paintings (i.e., portraits) in a row, then proceeded to the different category of the paintings (i.e., landscape). The stimulus sequence of

the CC condition had lower variety of contents, since when ordering by category, each category contained the same/similar contents e.g., viewing a series of landscape in a row, which lead to a lower variability among stimuli. Thus, within one category, the stimuli might appear to be similar compared to the PC condition. However, importantly, when there was a change of the category from one to another (i.e., from portraits to landscape) in the CC condition, the change itself could appear more salient compared to the PC condition, since in the PC condition, there was always high variability in the stimulus sequence, and there would be no clear changing point of the stimuli. In sum, although the paintings presented in the two different conditions were the same, the PC condition would have a much higher amount of novelty within the sequences, while the CC condition would have only a high novelty when the category changed from one category to another.

These two distinct presentation orders made it possible to examine the viewing time recoveries after the changes in stimuli—the observation of dishabituation process. With regard to the viewing time recoveries, we expected different results for the PC and CC conditions. In the PC condition, there might be higher variability in the stimulus sequence all the time, and there would not be a clear changing point of time period (i.e., from the 16th to 17th century). Contrary to the PC condition, in the CC condition, there might be lower variability in the stimulus sequence in each category of the paintings, but the change of the category (i.e., from landscape to portrait) would be salient. Accordingly, if the habituation is a likely cause of the attention decreased phenomenon from repeated art viewing, the viewing time would not show recoveries in the PC condition after the period of the artworks changed—because this ‘change’ is not salient to the viewers—but that the viewing time would show recoveries in the CC condition after the category of artworks changed—because the change in the stimulus sequences will be salient. Together, we hypothesized that response recovery occurs in the CC condition but not in the PC condition (Hypothesis 1).

Question 2: Does higher stimulus novelty in the stimulus sequence counteract/delay the decreased attention over repeated art viewing?

Second, we examined how we can mitigate the decreased attention over repeated art viewing. To this aim, in line with the past studies of affective habituation, we assessed if the higher stimulus novelty could mitigate/delay the decreased in the viewing time.

As shown in Leventhal et al. (2007), when the presented stimuli contain higher variability, this might delay/counteract the decrease in response. Accordingly, if the decreased attention could be mitigated by introducing novelty in the stimulus sequence, we expected that, in the PC condition the slope of the viewing time would be more gradual compared to the CC condition, since there were higher stimulus variability in the stimulus sequence in the PC condition (Hypothesis 2).

Pilot Study

Since the research question as well as the study idea itself are quite new to the field, we first conducted a pilot study to assess the feasibility of the study idea as well as to decide the proper number of participants to test the hypotheses for the main study. In the following section, we will first present the pilot study, and then present the main study.

3.2.1. Method

Participants

Thirty-one native Japanese psychology students (Tokyo, Japan, 20 male, 12 female; $M_{age} = 21.40$ years) participated in the pilot study for monetary compensation (1,000 Japanese yen, approximately \$ 10 US). All participants had no academic background in art history and aesthetics and normal or corrected-to-normal vision—measured after the testing session via verbal assessment. They were unaware of the purpose of the study, completed the experiment individually in a soundproof room at Keio university. This study was approved by the local ethical committee of Keio University. The participants were randomly assigned either to the PC and CC conditions,

leading 17 participants in the PC and 14 participants in the CC condition (please see the detailed description of the two conditions below).

Stimuli and Apparatus

The same stimuli used in Study 2-1 were used in the present study. They consisted of total 160 images of paintings selected from two exhibitions both held in a Japanese museum, The National Art Center, Tokyo. Half of the stimuli (80 images) was from an exhibition “400 Years of European Masterpieces from the State Hermitage Museum”, and another half of stimuli was from “The Birth of Impressionism – Freedom in Painting: Masterpieces from the Musée d’Orsay”.

As aforementioned in Study 2-1, the exhibition of the State Hermitage Museum (Saint Petersburg, Russia) chronologically presented the works of art in order to show the historical development of the Western paintings made in 16th, 17th, 18th, and 19th century. Consequently, this exhibition has the equivalent contents of the PC condition in this study. On the other hand, the exhibition of the Musée d’Orsay (Paris, France) focused on one specific art school/style, *Impressionism*, and presented the works of art corresponding to four different art categories (i.e., history paintings, landscape, portrait, and genre paintings). Consequently, this exhibition has the equivalent contents of the CC condition. For the stimulus selection, 20 paintings in each time epoch in the exhibition of the State Hermitage Museum and in each category in the exhibition of the Musée d’Orsay were selected, resulting in 80 total paintings from one exhibition.

After the stimulus selection, the final sequence of each experimental condition was created for the PC and CC condition separately. First, in both experimental conditions, all stimuli were included in the final stimulus sequence—resulting in 160 images of paintings. Second, the original presentation order of paintings was kept when it corresponds to the experimental condition. For example, in the PC condition, the original presentation order of the paintings from the State Hermitage Museum exhibition was kept the same, since they are already presented in chronological order (i.e., from the 16th to the 19th century paintings). Further, in the CC condition, the original

presentation order of the paintings from the Musée d'Orsay exhibition was kept the same, since they are already presented by category (i.e., history paintings, landscape, portrait, and genre paintings). However, in the PC condition, the paintings from the Musée d'Orsay exhibition were rearranged into a chronological order. To do so, the paintings were sorted by data/year when the paintings were made. In contrast, in the CC condition, the paintings from the State Hermitage Museum exhibition were rearranged into category-wise order. For instance, if the category of a painting was landscape, then the painting was added to the landscape block. The information of the paintings i.e., when it was made, and which art category it belongs was obtained via Wikiart (<http://www.wikiart.org>). In both conditions, the participants viewed the stimuli in the same order. However, the order of the exhibition (i.e., paintings from which exhibition the participants viewed first or second) was counter-balanced between participants. It is important to note that in the original presentation sequences (i.e., the State Hermitage Museum exhibition in the PC condition, and the Musée d'Orsay exhibition in the CC condition), the number of the paintings in each period/category was the same—20 paintings in each period/category. However, when the paintings were rearranged, there was not always the same number of paintings from one period/category in each stimulus block.

The experiment was controlled via PsychoPy 3.0 (Peirce, 2007) on a MacBook Pro (Apple, USA). Stimuli were presented on the cathode ray tube monitor (Trinitron, CPD-G420, SONY) with a screen resolution of 1280×960 pixels. All stimuli were presented on a gray background with their horizontal dimension fixed at 600 pixels. Each participant sat in front of the monitor at a distance of about 80 cm away.

Procedure

After the participants were welcomed in the laboratory, they were instructed that a series of images of paintings are going to be presented on the screen monitor, and only one painting will be presented at a time. They were told that they can view each artwork as much as they wished, and

there are no time restrictions for viewing time. As with Study 1 and Study 2-1, participants were not informed how many paintings were in the experimental task nor how long the study would last in order to avoid any anticipations for the experiment.

To start each trial, participants pressed a key on a keyboard. After the key press, one painting appeared on the center of the screen monitor. When participants wanted to finish viewing a painting, they press the same key again. Following the second key press, the screen changed into the rating screen, where participants were asked to rate how beautiful the painting was by pressing the appropriate key with a 9-point Likert visual scale (1: *not beautiful at all* to 9: *very beautiful*). This beauty assessment was especially included, since the results of Study 1 showed that the beauty rating scores could significantly predict the viewing time. Considering that the presentation order of paintings is going to be fixed in both PC and CC condition, the beauty appraisals of the paintings can be a critical confounding factor i.e., the slope of the viewing time in the PC condition would be more gradual than in the CC condition, since there are more beautiful paintings at the end of the sequences compared to the CC condition. Thus, the beauty rating scores were measured as a control factor.

3.2.2. Results

First, we calculated the mean viewing time for both conditions. The mean viewing time across all participants was 7.04s ($SD = 5.37$) for the PC condition, and 8.00s ($SD = 7.04$) for the CC condition. The mean viewing time per participants ranged from 1.64s to 20.40s in the PC condition, and from 4.01s to 16.8s in the CC condition.

Dishabituation in the viewing time in the PC and CC conditions

To assess the first hypothesis: response recovery occurs in the CC condition but not in the PC condition, we compared viewing time right before and after the period/category changes in the PC and the CC conditions. In the original study where the dishabituation process was observed

(Groves & Thompson, 1970), the dishabituation was present only a short while after the presentation of the novel stimulus. Thereby, we compared the viewing times right before and after the stimulus block changes in the following analysis. It is important to note that only the changing points of the original presentation order (i.e., stimulus block change points in the paintings from Hermitage museum exhibition in the PC condition, and those in the paintings from Musée d'Orsay exhibition in the CC condition) were used for the analysis below, since, again, the number of the paintings included in each stimulus block was not always identical to 20 when the presentation order was rearranged for the final stimulus sequences.

In the PC condition, the stimulus block was changed three times, from the 16th to 17th, from the 17th to 18th, and from the 18th to 19th century. In the CC condition, the category of the paintings was changed three times, from historical paintings to landscapes, from landscapes to portraits, and from portraits to genre paintings. Thus, we conducted pair-wise *t* tests for three times using the viewing time right before and after the block changes for each condition. The mean viewing time right before and after the stimulus block changes are presented in Figure 9.

As a result, in the PC condition, the viewing time before and after the first and third block changes significantly differed (first: $t(16) = 1.80, p = .09, d = 0.55$, third: $t(16) = 2.70, p = .02, d = 0.59$). As shown in Figure 9, the viewing times after the first and third stimulus block changes were significantly shorter than those before the block changes—the viewing time kept decreasing after the block changes. Further, there was no significant difference in the viewing times before and after the second block change in the PC condition ($t(16) = .24, p = .80, d = 0.05$). In contrast, in the CC condition, there were no statistically significant differences in the viewing times before and after the stimulus block changes (first: $t(13) = 0.91, p = .38, d = 0.24$, second: $t(13) = -0.22, p = .83, d = 0.06$, third: $t(13) = 1.63, p = .013, d = 0.43$).

Considering that the number of the participants included in the pilot study in each condition was quite small (PC: $N = 17$, CC: $N = 14$), we also describe the results of the effect sizes. Following the interpretation of the Cohen's *D* scores (Cohen, 1988), the *D* score .20 can be

interpreted as small, = .50 as moderate, = .80 as strong effect. Hence, the differences in the viewing time before and after the first and third block change in the PC condition had moderate effect sizes, $d = 0.55, 0.59$ respectively. Further, the differences in the viewing time before and after the first and third block change in the CC condition also had small/moderate effects, $d = 0.24, 0.43$ respectively. Importantly, after the second and third block changes in the CC condition, the average viewing time become longer than before, while in the other block changes including the results of the PC condition, the viewing time kept decreasing over the changes. Given these results, one might argue that while the viewing time before and after the block change constantly decreased in the PC condition, those in the CC condition first decreased, then stayed the same level with slightly longer viewing time after the change and showed a recovery in the last block change.

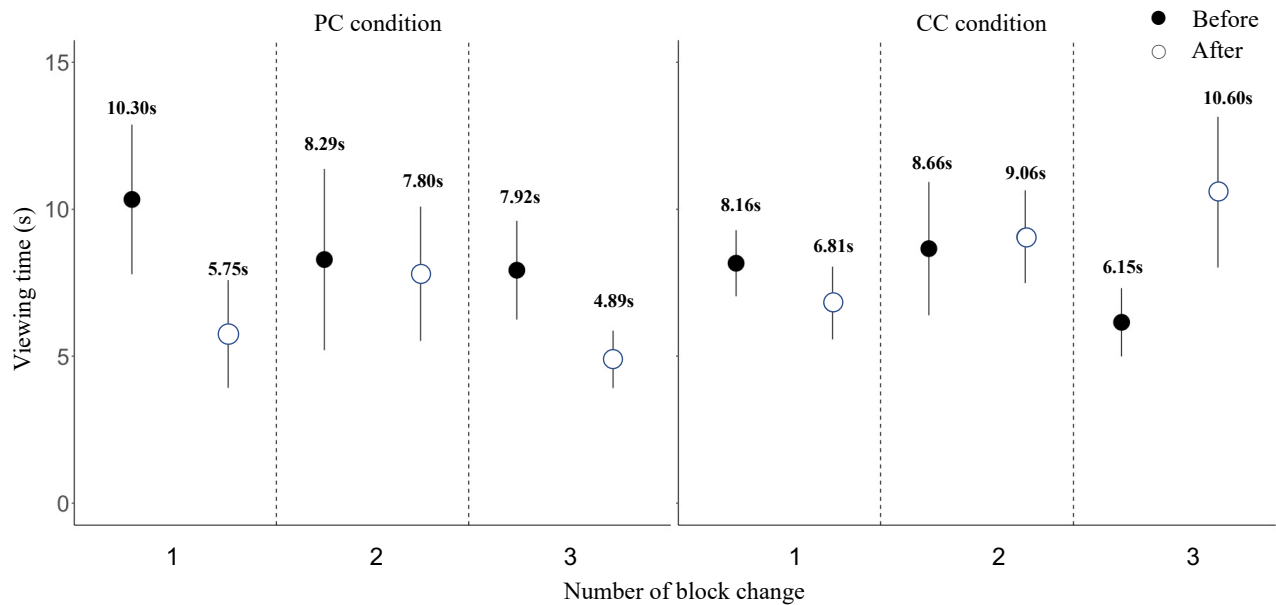


Figure 9. The mean viewing times before and after the block changes across all participants for each condition. The left side of the graph shows the result from the PC condition, and the right side from the CC condition. The error bars indicate the Standard Error (SE) for each point. Black points represent the mean viewing time before each stimulus block change, while white ones represent the mean viewing time after each stimulus block change. The numbers in the figure represent the exact mean viewing time in each point.

The effect of stimulus novelty in the presented stimulus sequences on viewing time

To assess the second hypothesis: in the PC condition the slope of the viewing time would be more gradual compared to the CC condition due to the *novelty effect*, we compared the slope of the viewing time between two experimental conditions using a LMM (Linear Mixed Model).

Before building up the LMM model, the distribution of the viewing time was visually assessed. The viewing time seems to have the shape of log-normal distribution, which is very typical for reaction time data. Since LMMs postulate the data normality, the viewing time was log-transformed in order to assure the data normality. The mean log-transformed viewing times for each trial per condition are shown in Figure 10.

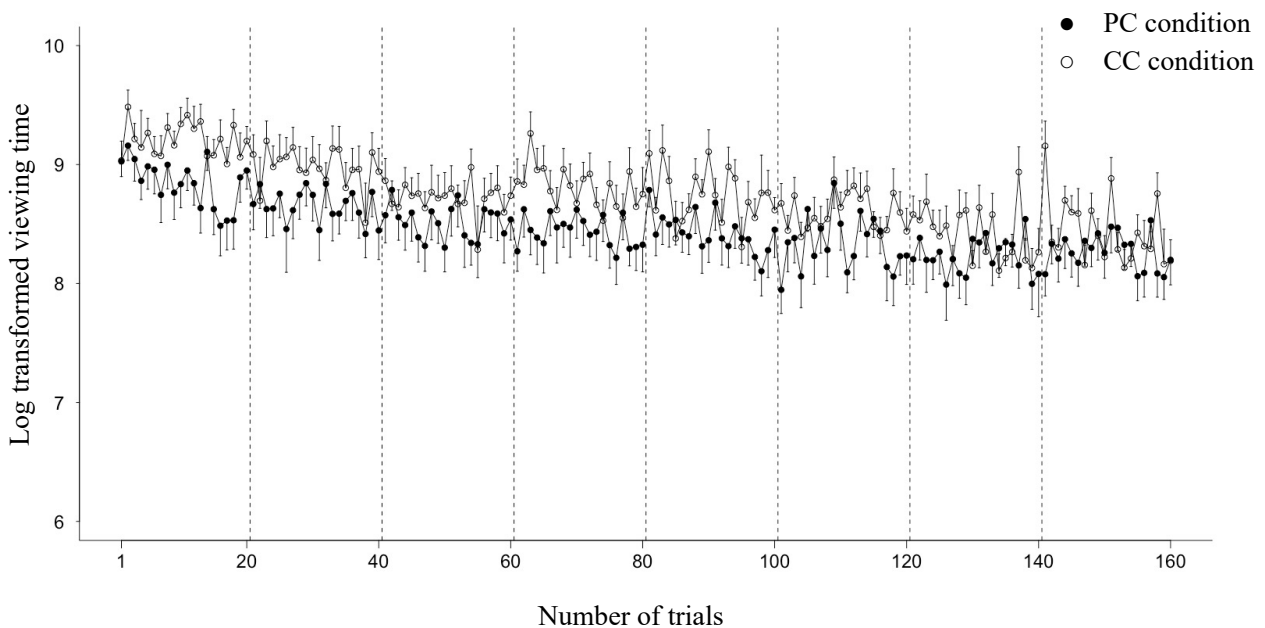


Figure 10. The mean log-transformed viewing time for each trial in the PC and CC conditions. Error bars indicate the Standard Error (SE) for each point. The black points represent the results of the PC condition, and the white points represent those of the CC condition. Vertical lines highlight the points where the stimulus period/category changed in the original presentation order.

In the final LMM, the log-transformed viewing time for each trial (coded continuous) was set as the dependent variable. The experimental condition with the CC (baseline) against the PC condition, the trial numbers (0 to 159, coded continuous), and the interaction between the two independent variables were set as the fixed effects. Note that when we first set the log-transformed viewing time and raw number of trials, the model did not converge and suggested rescaling variables. Accordingly, in order to ease the model complexity, log-transformed viewing time as well as the number of trials were *Z* standardized. For the random effects, random intercept, and slope for trials by participants as well as the random intercept for the beauty rating score (ranged from 1 to 9, coded continuous) were included. The random intercept and the slope of the viewing time for trials by participants were set, in order to assess the inter-individual differences in the viewing time i.e., some participants might view artworks generally longer than the others, some participants might show the decrease in viewing time quicker than other participants. Note that the beauty rating scores were treated as a control variable in the present experiment. Especially in the pilot study, since the presentation order was fixed, the beauty rating scores can skew the results of the viewing time. According to Barr et al. (2013), it is not essential to specify random slopes for control variables as long as the interactions with factors of interest are not present in the model. In the present model, the interactions between the interested factors (i.e., experimental conditions and the number of trials) were not included. Accordingly, the random slope for the beauty rating score was not included in the final model.

The LMM was carried out in the R environment (Version 1.1.423. R Development Core Team, Bates et al., 2017) with lme4 package (Version 1.1-13; Bates, Mächler, Bolker, & Walker, 2015) applying Satterthwaite approximation for the estimations of *p* values. Further, the effect sizes, Cohen's *D*, for each fixed effect were computed with EMAtools package (Version 0.1.3.; Kleiman, 2017). All fixed effects and the results are shown in Table 8. As can be seen in Table 8, there was a significant main effect of the number of trials, meaning the viewing time decreased as a function of trial in both PC and CC conditions. This result provides another replication of the

viewing time decreased phenomenon over repeated art viewing in a lab-based setting. Further, there was no significant main effect of the experimental condition as well as the interaction between the number of trials and experimental condition. These results suggest that the intercept of the viewing time as well as the slope of the viewing time were not significantly different between two experimental conditions.

Given the results of the effect sizes, D scores, the effect of the number of trials can be regarded as a strong effect, and main effect of the experimental condition as well as the interaction between the trial number and the experimental condition also had moderate effects. Regarding the main effect of the experimental condition, the estimated coefficient for the PC condition was -0.338, meaning that the intercept of the PC condition was shorter than that of the CC condition. Regarding the interaction between the trial number and the experimental condition, the estimated coefficient for the PC condition was 0.067. Considering the coefficient has a positive result, this result can be interpreted that the slope of the viewing time in the PC condition was more gradual compared to the CC condition.

Table 8. Fixed effects in the LMM predicting the viewing time in the pilot study

Fixed effect	Estimate	95 % CI		t value	Cohen's D	Pr ($> t $)
		Lower	Upper			
(Intercept)	0.123	-0.248	0.493	8.489	-	0.52
Number of trials	-0.291	-0.379	-0.203	25.374	-2.392	< .001
PC condition	-0.338	-0.826	0.149	26.483	-0.505	0.18
Number of trials * PC condition	0.067	-0.052	0.186	6.321	0.407	0.28

3.2.3 Discussion of the pilot study

The purpose of the pilot study was to test the feasibility of the study idea and research questions: (1) Does habituation process cause the decreased attention from repeated art viewing?,

(2) Does higher stimulus novelty in the stimulus sequence counteract/delay the decreased attention over repeated art viewing?

The first question was assessed by comparing viewing time before and after the changes of the stimulus blocks in the PC and CC conditions. We expected that while the PC condition would not show any recovery in the viewing time after the block changes, the CC condition would show the recovery in the viewing time if the habituation process was a likely cause of the decreased attention from repeated art viewing. As hypothesized, the participants in the PC condition did not show any recovery in viewing times, rather, the viewing times after the block changes were either the same level or shorter compared to the viewing time right before the block changes. In contrast, the results of the CC condition seem to have a different tendency in the following two points. First, seeing the mean viewing time before and after the first block change in the CC condition (see Figure 9), the viewing time after the first block change was shorter than that before the change as in the PC condition. However, the result of the *t*-test for the first block change in the CC condition showed that the block change had a small effect on the viewing time ($d = 0.24$), while we observed a moderate effect in the PC condition in the first block change ($d = 0.55$). Hence, the decrease in the viewing time after the first block change was smaller in the CC condition than in the PC condition. Second, notably, the viewing time did recover after the third block change, while in the PC condition, the viewing times kept decreasing over the third block change. Combining all results together, though the viewing time did not always show the recoveries in the CC condition, the salient block changes in the CC condition seem either to mitigate the decrease in the viewing time or to extend the viewing time after a block change. Hence, our first hypothesis seems to be, though not fully, supported.

A brief note should be made on why there were no viewing recoveries in the first/second block changes in the CC condition. In the pilot study, the order of the paintings in both experimental conditions was the same across participants. Thus, the paintings right before and after the block changes were always the same, meaning that it could be possible that only the third block

change in the CC condition was salient to the participants, while other changes were not well recognized. For example, the first block change in the CC condition was from history painting block to landscape block. In history paintings, there are some paintings which are mainly focus on the buildings/scenes rather than events with humans/animals. Then, it would be not so salient for the participants to see the block changes in the CC condition. Especially, considering that all participants in the pilot study did not have any background education history/deep knowledge on art, it could be plausible to assume that some of the changes might not have been salient enough to them. In order to cancel out the effect of specific paintings as well as the specific block changes (i.e., the change from history paintings to landscape was vague than other block changes), it might be helpful to randomize the presentation order of the paintings within one stimulus block.

The second question was assessed by comparing the slopes of the viewing time between the PC and CC condition. We expected that the slope in the PC condition would be more gradual than that in the CC condition due to the novelty effect. As reported in Table 8, the LMM did not show the significant interaction between the number of trials and experimental condition ($\beta = 0.067$, $d = 0.41$). However, as aforementioned, considering that the small sample size in the pilot study, it might be fair to argue the result by seeing the effect size. Since the coefficient has a positive value, this result means that, in the PC condition, the slope of the viewing time from one trial to the next trial is 0.067 times more gradual compared to the CC condition with a moderate effect size. Hence, the result is in line with our second hypothesis, and thereby, it seems like the higher novelty in the stimulus sequence can mitigate the decreased attention from repeated art viewing.

Overall, the results of the pilot study partially supported the two hypotheses, and it seems to be possible to assess the study idea with the methodology used in the pilot study. However, as aforementioned, there are several limitations which should be further addressed/improved in the pilot study i.e., the small sample size, the effect of painting before and after the stimulus block changes. In the next section, we will present the main study where we improved the study design given the results of the pilot study.

3.3. Main Study

In the pilot study, there were specifically four major methodological limitations which might skew the results: 1. Relatively small sample size in each experimental condition, 2. Observed floor effect at the end of the experiment and skewed categorization of the paintings, 3. The uneven number of the paintings when they are rearranged, and 4. The effect of paintings before and after the block changes. Given those limitations, in the main study, the experimental procedure was further improved. In the following section, we will describe each limitation in the pilot study, and how to improve the study design in the main study.

Regarding the first limitation, in the pilot study, there were only 31 participants in total. Especially for multilevel modeling analysis including (G)LMMs, it is valuable to consider how much variance in the data the random effects can explain when we decide the proper sample size. Thus, I retrieved this information from the results in the pilot study and used it for the power analysis. The detailed descriptions on how we performed the power analysis as well as on how we decided the sample size can be found in the *Participants* in the *Method* section.

Regarding the second limitation, there are two points which should be mentioned. First, as shown in Figure 10, in the pilot study, the viewing time did not seem to decrease after the first 120 trials. Presumably, the viewing time met the floor effect, and did not show further decrease. Thus, in order to avoid the floor effect, it seems plausible to reduce the number of stimuli in the main study. Second, in both PC and CC conditions, there was one stimulus block which did not fit the concepts of the two conditions, namely, the 16th century period block in the PC condition and the genre painting block in the CC condition. In the PC condition, the stimulus sequence is expected to have high novelty within one stimulus block with variety of art contents/styles. Though these premises are kept in the 17/18/19th century period blocks, most of the paintings included in the 16th century period block were religious paintings, leading to less variety in the contents of the stimulus sequence. Similarly, in the CC condition, the stimulus sequence is expected to have low variability within one stimulus block with similar art contents/styles. However, in the genre

painting block in the CC condition, there are variety of art contents i.e., some paintings are mainly depicting humans, some are focusing on the buildings and scenes of everyday life or still-lives such as pianos/flowers. Thus, those two stimulus block conditions did not fit the ideas of the PC and CC condition. Given these points, in the main study, the 16th century period stimulus block as well as the genre painting block were excluded from the stimulus sequences. Thereby, the total number of stimuli was changed from 160 to 120, with 6 stimulus blocks each containing 20 paintings.

Regarding the third limitation, when the paintings were rearranged for making the final sequences (i.e., the paintings from State Hermitage Museum exhibition in the CC condition, and the Musée d'Orsay exhibition in the PC condition), the stimulus block did not always contain 20 paintings. Consequently, in the pilot study, I could only use the three block changes in each experimental condition. In order to solve this issue in the main study, the stimuli were partially replaced, so that the stimulus blocks in both experimental conditions can always have 20 paintings even when they are rearranged. The detailed stimulus selection criteria is going to be shown in the *Stimuli and Apparatus* in the *Method* section.

Regarding the last limitation, since the presentation orders of the paintings in both experimental conditions were fixed in the pilot study, the results could be strongly influenced by the specific paintings. Specifically, when I compared the viewing time before and after the stimulus block changes, the viewing time for specific paintings were used. Consequently, in the pilot study, the changes themselves might not be salient if paintings with similar structure/contents were presented over the block changes in the CC condition. Further, if the painting after the block change was evaluated generally as more beautiful by the participants than that before the block change, the recovery in the viewing time was observed not due to the stimulus novelty, but rather due to the perceived beauty of that specific painting—as aforementioned, more beautiful paintings tend to be viewed longer. Accordingly, in the main study, the presentation order of the paintings within one stimulus block as well as the presentation order of stimulus block themselves were randomized in order to cancel out the effect of specific paintings.

3.3.1. Method

Participants

As in the pilot study, for assessing the first hypothesis, we intended to run a series of paired *t*-tests using the viewing times before and after stimulus block changes. We calculated the sample size for the *t*-test with G*Power 3 (Faul et al., 2007). For this calculation, we set a medium effect size, Cohen's *d* of 0.43, considering the third block change in the CC condition had $d = 0.43$. The power analysis indicated a total sample size of 70 participants (35 participants in each experimental condition) to achieve a power of .80 with an α of .05. Further, for assessing the second hypothesis, we intend to run a Linear Mixed Model. The model structure of the main study will be identical to the pilot study—in case the model did not converge, the structure of the random effects will be changed. The sample size for the LMM was calculated by *study_parameters* function in the *powerlmm* package (Magnusson, 2021) using R environment. In this package, we have to set the total number of trials which one participant will go through, the number of the participants, the Intra Class Correlation (ICC) for the factor 'participants', the correlation between the random intercept and slope for the level-1 factor, and the effect size. The first value was set to 120, since in the main study, all participants will go through 120 trials. The second value was set to 35, given the fact that the power analysis for the *t*-tests suggested 35 participants in each experimental condition. The values for the ICC, the correlation between the random intercept and slope for the level-1 factor, and the effect size was obtained from the results of the pilot study. For the effect size, considering the purpose of this analysis to compare the slopes of the viewing times in two experimental conditions, we set the value as 0.41, since this was the estimated *d* score for the interaction between the number of trials and experimental condition in the pilot study. The result indicated that, the experimental structure with 70 participants (35 in each experimental condition) going through 120 trials will achieve a power of 1.00 with an α of .05 to detect a moderate effect size i.e., $d = 0.41$.

Since the data for the main study was collected during the COVID-19 pandemic, the main study was conducted online, while the pilot study was conducted in a laboratory. In the main study, participants were recruited via emails and the link to the online study was attached in the email. Participants could join the study whenever/wherever they wanted via their own laptops. We recruited participants until we had complete data from 35 participants in each experimental condition. Note that when we sent emails to the first 70 participants, there were several people who did not join/complete the experiment. In case they did not join the experiment, we sent a reminder two weeks after the first email. When the participants still did not join the study after three weeks from the first email, we recruited seven more participants (10% of the whole population) until we reached 70 total participants with complete data.

For the final analysis, 71 participants with full-completed data were included in the main study (22 male, 49 female, $M_{age} = 25.73$, $SD = 9.46$). They were participated in the present study with monetary compensation (1,500 Japanese yen, approximately \$ 15 US). Only native Japanese participants who were raised in Japan participated in the present study. They were all unaware of the purpose of the study. Importantly, they were asked to join the experiment alone with a quiet/composed condition where they could concentrate on the task e.g., do not chat with their families/friends about the presented artworks on the screen. This study was approved by the local ethical committee of Keio University, and was carried out in accordance with the Declaration of Helsinki. The participants were randomly assigned either to the PC and CC conditions, leading 35 participants in the PC ($M_{age} = 26.60$, $SD = 10.60$) and 36 participants in the CC condition ($M_{age} = 24.80$, $SD = 8.19$). In order to see if there is a significant difference in the ages of the participants between the two conditions, especially because in the main study there are not only students but people with ranged age groups were participated, we conducted a two-sample t -test using the ages of the participants in the two groups. The result showed that there was no significant different in the mean age of the participants between the two conditions ($t(66) = -0.81$, $p = 0.40$, 95% CI [-6.32, 2.64], $d = 0.19$).

Stimuli and Apparatus

Most of the paintings used in the main study were identical to those in the pilot study. However, importantly, some of the stimuli in the main study were replaced in order to address the third limitation in the pilot study—unequal number of the images of the paintings in each stimulus block when they are rearranged. In the original 17th/18th/19th stimulus blocks, only mythological/religious paintings, landscapes, portraits were kept. If one painting belonged to other categories i.e., genre painting, still-life, the painting was replaced by another painting from the same artist which is categorized as either mythological/religious paintings, landscapes, or portraits. Further, in each century stimulus block, the equal number of mythological/religious paintings, landscapes, portraits are included, so that there are the same number of paintings in each category, when the paintings are rearranged by category for the CC condition. Note also that in order to eliminate the impact of one specific artist, no more than three paintings from one artist were included. When the paintings were rearranged for the CC condition, paintings included in the 17th/18th/19th stimulus blocks was sorted by the category of the paintings, leading three stimulus categories i.e., methodological/religious paintings/landscape/portrait with 20 paintings in each block.

The paintings in the original historical painting/landscape/portrait stimulus blocks were almost kept identical, since the dates when the paintings were made were already scattered. However, as with another condition, no more than three paintings from one artist were included. When the paintings were rearranged for the PC condition, paintings included in the historical paintings/landscape/portrait stimulus block was first sorted by the date in each stimulus block. Then the one third of the paintings which were made in the early period in each stimulus block were combined together, constituting ‘early period stimulus block’. Similarly, another third of the paintings which were made in the middle period in each stimulus block were combined, constituting ‘middle period stimulus block’. The last stimulus block ‘late period stimulus block’ were made in the same manner. This method was introduced, since when al paintings were sorted

together by period as in the pilot study, the paintings in each stimulus block tended to have similar art categories i.e., the early period paintings are religious/historical paintings, leading the high similarity in each stimulus block. Thus, in order to make sure that there are variety of art contents in each stimulus block for the PC condition, one third of the paintings from each stimulus category was combined.

Finally, in order to address the last limitation in the pilot study—the same paintings presented before and the after the stimulus block changes, the presentation order of the paintings within each stimulus block was randomized. Further, the presentation of the stimulus blocks were randomized in each exhibition. Also, the order of the exhibition (i.e., paintings from which exhibition the participants viewed first or second) was counter-balanced between participants as done in the pilot study. For example, when a participant is assigned to the PC condition, s/he might experience the stimulus sequence starting from 18th century, 17th century, 19th century, to landscape, portrait, historical paintings, while another participate in the PC condition might experience the stimulus sequence starting from historical paintings, landscape, portrait, to 19th century, 18th century, 17th century. By doing so, we aimed to eliminate the effect of specific paintings on the recoveries of the viewing time after the block changes, and to assess only the effect of stimulus block changes.

As aforementioned, due to the restrictions for the COVID-19 pandemic, the main study was an online study. Each participant participated the study with their own laptops, and thereby, the sizes of the screen monitors differed across participants. In order to make sure that the paintings were presented in the same size for all of the participants, the vertical size of the paintings were always set to 600pixel as with the pilot study. The biggest horizontal size of the painting in the stimulus set was 1025pixel, and thus, the participants whose laptop could not present this size of the image could not join the online study. The experiment was controlled via Labvanced (Finger et al., 2017).

Procedure

Before the Japanese participants started the online study, a written introduction, explanations of the experiment structure, approximate time to complete the study, and the information that they will get monetary compensation by participating this study was provided to them. Only when they understood/agreed on the handed information, they were asked to jump to the experimental link, guiding them to the online study.

At the beginning of the experiment, the gender, age, first language, nationality of the participants as well as the country where they were raised were asked to the participants. After the participants' demographic information was collected, there was a practice trial session. In the past experiments in the present dissertation, the participants tended to view artworks quite long at the beginning of the experiment. If the participants spend long time to view each artwork at the beginning of the experiment to get used to the experimental task itself, the steep decrease in the viewing times at the very beginning of the experimental task could be an artificial effect and might have an impact on the slope of the viewing time as a whole. Thus, in the present study, there were five practice trials for the viewing task. Note that the stimuli included in the practice trials were not included in the main session. Before the participants started the practice session, they were instructed that they will see a series of artworks in the experiment, and that they can view each artwork as long as they wish. However, they were not informed how many paintings are in the task to avoid any time anticipations for the experiment. To start each trial, they pressed either the space or enter key. Followed by the key press, one painting appeared on the center of the screen. When participants finished viewing a painting, they pressed the same key again to complete the viewing. The duration from the first key press to the second the key press was recoded as the viewing time. After the second key press, the screen was changed into the rating screen, where participants was asked to rate how beautiful the painting was by pressing the appropriate key with a 7-point Likert visual scale (1: *not beautiful at all* to 7: *very beautiful*). After they typed the rating score, the screen

was again back to the instruction screen, saying '*please press the space/enter key to start the next trial*'.

After participants completed the practice trials, they were instructed to press the space key to start the main viewing task. Importantly, before the main viewing task started, they were asked not to leave the room and stop the experiment as much as possible during the experiment. However, if they had to stop the experiment, they were asked to do so when the instruction screen for the next trial is presented on the screen monitor. Further, if they left the room or stopped the experiment for a while, they were asked to report this to the experimenter. In the end, no participants reported that they did so during the experiment.

In the main viewing task, participants repeated the same structured trial until they finished viewing all 120 paintings. After the main viewing task, the experiment proceeded to the post-hoc assessment to examine the art interest and expertise of the participants. Since the main study was an online study, these questions were important in order to understand the participants' demographic information related to art.

First, the participants' art interest was assessed using the questions included in the VAIK (Specker et al., 2018). It is important to note that only the questions in art interest part of the VAIK was included. As note by Specker et al. (2018), in the VAIK, art interest and knowledge are two different concepts e.g., there are some people who do not have rich knowledge of artworks, but still like to go to museums/art galleries and enjoy viewing artworks. When the VAIK questionnaire was developed, only the data from Western participants was included. Further, all questions included for assessing art knowledge scores were related to Western artworks. Thereby, they are not valid to assess the art knowledge level of the people with different cultural background (see also Specker et al., 2018). Hence, in the present study, the questions assessing art knowledge were not included. As with the original VAIK questionnaire, for each question, participants could select one score with a 7-point scale.

Second, art expertise of the participants were assessed via one question ‘*Have you ever joined any practical/academic/research activities related to art/aesthetics? e.g., I am an artist, I am/was studying art/aesthetics/art history in the university*’ with two options either yes or no.

Last, what kind of visual artworks the participants prefer/know the most was assessed via two questions ‘*Which of the following artworks do you prefer the most to appreciate?*’ and ‘*Which of the following artworks do you think you know the most?*’ *The definition of ‘you know the most’ here is ‘the extent to which you have knowledge of the historical and cultural background of a work of art, as well as its composition and arrangement, regardless of your tastes or interests’. Since all of the artworks in the presented sequences were Western artworks, these questions would help to understand the familiarity of the participants towards the presented artworks. For the answers, there were five options available: 1. Western artworks, 2. East Asian artworks (including Japanese artworks), 3. Islamic artworks, 4. African artworks, 5. Other.

3.3.2. Results

Art interest and expertise of the participants

Before diving into the main part of the analysis, the results of the assessment of the art interest and expertise level of the participants were analyzed.

First, we calculated the art interest scores for each participant. Originally, there were 11 questions in the VAIK art interest questionnaire (Specker et al., 2018) e.g., I like to talk about art with others, I’m interested in art. The art interest score can be calculated by summing up the scores for each question ranged from 1 to 7. Note that, in the present experiment, most of the answers for two questions i.e., ‘*How often do you visit art museums and/or galleries?*’ and ‘*How often do you read books, magazines or catalogs about art?*’ were lost due to the recording issue. Hence, the art interest scores for each participant were calculated only using the scores for the rest of nine questions. Note that the questionnaire was developed and valid with the use of the full 11 questions. Hence, what we present here is not a completely valid measure for art interest. Since there were

nine questions in total, the possible lowest score was 7, and the possible highest score was 63. Overall, the mean art interest score across all participants was 33.21 ($SD = 8.66$) ranged from 10 to 51. In order to see if the overall mean art interest scores was higher than the theoretical average score i.e., 31.5 points, we conducted one sample t -test comparing the mean art interest scores across all participants to 31.5. The result did not show a significant difference ($t(70) = 1.70, p = 0.10, 95\%$ CI [31.16, 35.26], $d = 0.19$). For the PC condition, the mean art interest score was 34.03 ($SD = 9.45$) ranged from 14 to 51. For the CC condition, the mean art interest score was 32.42 ($SD = 7.86$) ranged from 10 to 46. In order to assess if there is a significant difference in the mean art interest score between the PC and CC condition, we conducted two sample t -test using the art interest score from all participants in each experimental condition. The result showed that there is no significant difference in the mean art interest scores between the two conditions ($t(66) = 0.78, p = 0.40, 95\%$ CI [-2.51, 5.73], $d = 0.18$).

Second, regarding the art expertise of the participants assessed via the question '*Have you ever joined any academic/research activities related to art/aesthetics? e.g., I am an artist, I am/was studying art/aesthetics/art history in the university*', none of the participants selected 'yes'. Thus, all participants were art laypersons.

Last, the participants' favorite and the most knowledgeable artworks were assessed to see what kind of artworks they like and are familiar with. In Table 9, the answers from the participants are shown. Over 70% of the participants answered that they prefer Western artworks, and they know Western artworks the most. The percentage of the participants who prefer/know the East Asian paintings the most was 16.90%. The percentage of the participants who prefer the Islamic/African/Other artworks was 5.63%, and that of the participants who know the Islamic/African/Other artworks the most was 11.27%. Though the participants are originally from Japan, most of them prefer/know the most about Western artworks. Considering that the famous museums/art galleries often hold exhibitions of Western artworks in Japan, this result seems reliable. In sum, the Japanese participants in the present study could be assumed to be art

laypersons with a moderate level of art interest, and the majority preferred/knew about Western artworks the most.

Table 9. The number of the participants who preferred/know the most about five distinct categories of visual artworks

	Western artworks	East Asian artworks	Islamic artworks	African artworks	Other
Favorite art					
All	N = 55	N = 12	N = 1	N = 0	N = 3
PC condition	N = 26	N = 5	N = 1	N = 0	N = 3
CC condition	N = 29	N = 7	N = 0	N = 0	N = 0
Knowledgeable art					
All	N = 51	N = 12	N = 0	N = 1	N = 7
PC condition	N = 27	N = 3	N = 0	N = 1	N = 4
CC condition	N = 24	N = 9	N = 0	N = 0	N = 3

Pre-processing for the viewing time data

Before analysing the viewing time data, the trials were checked for each participant. Trials with a viewing time that was under 300ms as well as above 3 *SD* from the within-participant mean were excluded. Since the main study was an online study and the experimenter could not check if the participants skipped a painting by mistakenly pressing the space/enter key or if they were sleeping/doing something else during the main task, this data pre-processing was applied in order to keep the data quality as high as possible. This amounted to 1.76% of data of viewing time in total—150 trials in total were recognized as outliers. The maximum number of the trials in one participant, which were recognized as outlier, was four, 0.047% out of the whole data. Thus, all participants' data without the excluded trials were retained for the analyses below.

Dishabituation in the viewing time in the PC and CC conditions

In the main study, there were five stimulus block changes in both experimental conditions. As with the pilot study, I conducted five pair-wise *t* tests using the viewing time right before and

after the block changes separately for each condition to examine if the viewing times recovered after the block changes only in the CC condition (hypothesis 1). The mean viewing time right before and after the stimulus block changes in the PC condition are presented in Figure 11. The same results for the CC condition are presented in Figure 12.

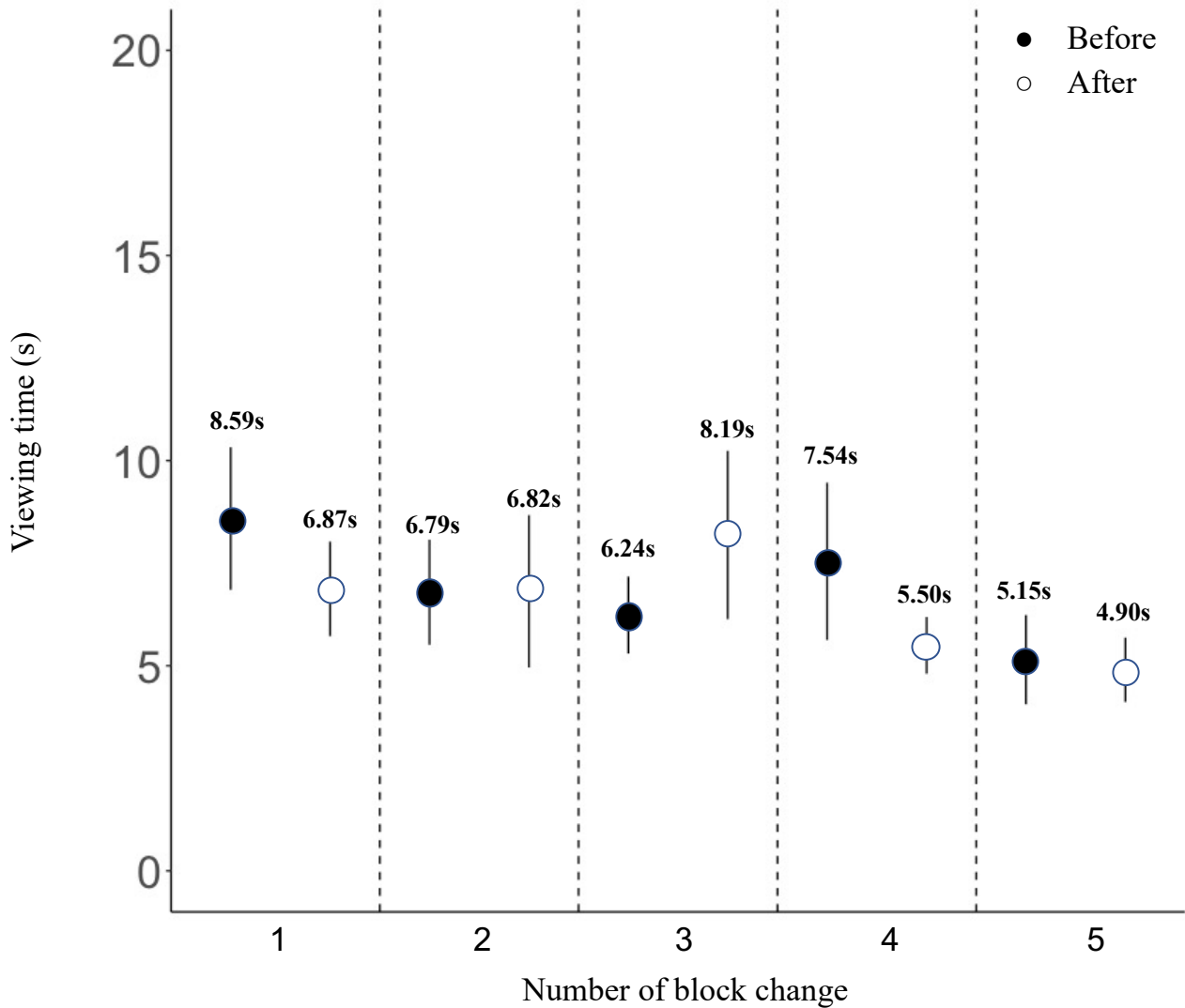


Figure 11. The mean viewing times for right before and after the stimulus block changes in the PC condition. Error bars indicate the Standard Error (SE) for each point. The black points represent the mean viewing times before the stimulus block changes, and the white points represent those after

the stimulus block changes. Vertical lines highlight the number of the block changes from one to five. The numbers in the figure represent the exact mean viewing time in each point.

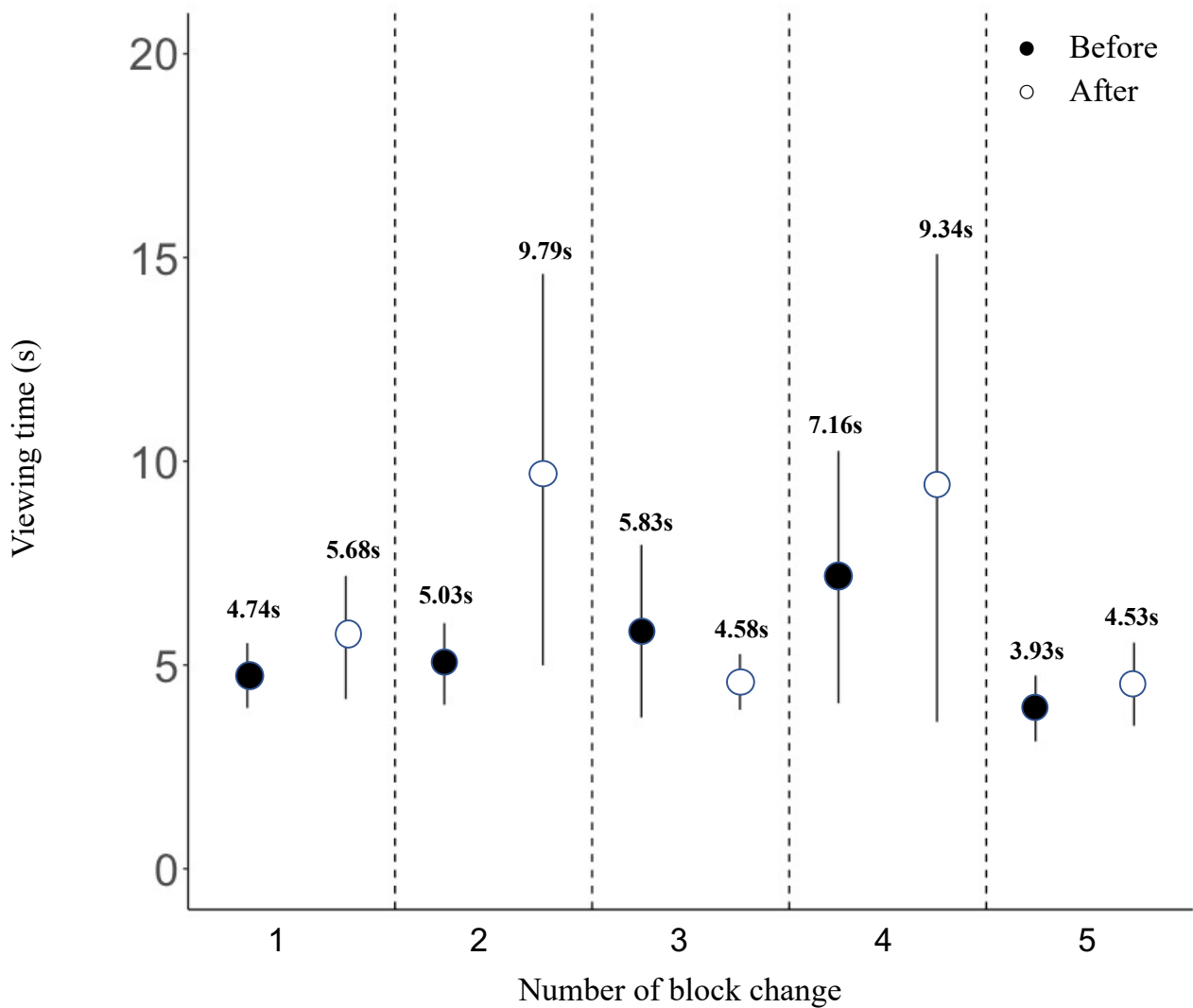


Figure 12. The mean viewing times for right before and after the stimulus block changes in the CC condition. Error bars indicate the Standard Error (SE) for each point. The black points represent the mean viewing times before the stimulus block changes, and the white points represent those after the stimulus block changes. Vertical lines highlight the number of the block changes from one to five. The numbers in the figure represent the exact mean viewing time in each point.

The results of all t -tests for both conditions are shown in Table 10. As shown in the table, the viewing time before and after the stimulus block changes were not significantly different in both experimental conditions.

Table 10. The results of t -tests comparing the mean viewing time right before and after the block changes for both experimental conditions.

	<i>t</i> value	df	95% CI		Cohen's D	<i>p</i> value
			Lower	Upper		
PC condition						
First	1.20	34	-1.16	4.59	0.19	0.20
Second	-0.02	34	-2.33	2.29	0.00	1.00
Third	-1.10	33	-5.82	1.71	0.20	0.30
Fourth	1.30	33	-1.21	5.71	0.24	0.20
Fifth	-0.78	31	-1.97	0.89	0.04	0.40
CC condition						
First	-0.80	32	-4.11	1.79	0.14	0.40
Second	-1.00	35	-14.33	4.80	0.17	0.30
Third	0.55	32	-3.30	5.76	0.10	0.60
Fourth	-0.76	35	-8.03	3.66	0.13	0.50
Fifth	-0.42	34	-2.95	1.93	0.07	0.70

The effect of stimulus novelty in the presented stimulus sequences on the viewing times

In order to assess if the slope of the viewing time in the PC condition would be more gradual compared to the CC condition due to the novelty effect (hypothesis 2), another LMM was performed. As with the pilot study, the distribution of the viewing time was visually assessed to see the data normality before building up a model. Again, the viewing time was not normally distributed, but rather had the shape of log-normal distribution. In order to assure the data normality, the viewing time data was log-transformed. The mean log-transformed viewing times for each trial per condition are shown in Figure 13.

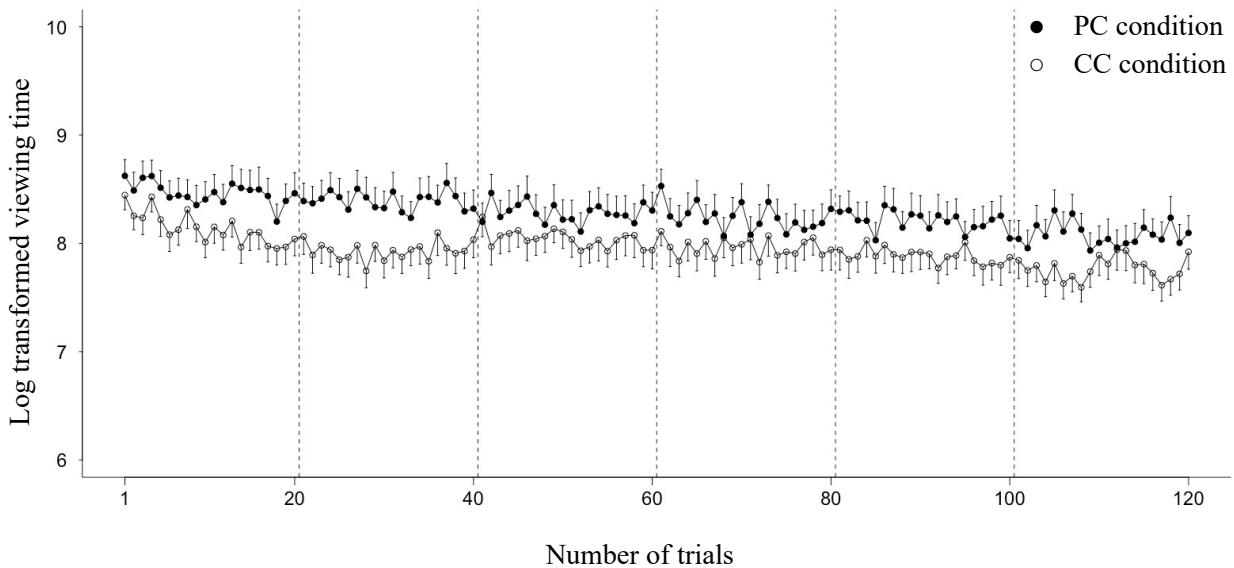


Figure 13. The mean viewing times for each trial for both experimental conditions. The black points represent the results from the PC condition, and the white ones represent those from the CC condition. The dashed lines express the points where the stimulus block changes occurred. The error bars represent SE for each data point.

The structure of the final model was identical to the one in the pilot study. The LMM was carried out in the R environment (Version 1.1.423. R Development Core Team, Bates et al., 2017) with lme4 package (Version 1.1-13; Bates, Mächler, Bolker, & Walker, 2015) applying Satterthwaite approximation for the estimations of p values. As in the pilot study, effect sizes for the fixed effects were estimated by EMAtools package (Version 0.1.3.; Kleiman, 2017). All fixed effects and the results are shown in the upper part of Table 11.

The result shows a significant main effect of the number of trials. Hence, again, the viewing time in both PC and CC conditions significantly decreased as a function of the number of trials. However, the main effect of the experimental condition as well as the interaction between the experimental condition and the number of trials were not significant. Seeing the effect sizes, the effect size of the number of trials was -0.91. Thus, indeed, the effect of repeated art viewing has a

strong effect on the viewing time. The effect size of the main effect of the experimental condition was 0.38, suggesting it has a moderate effect. Considering that the estimated coefficient of the PC condition was 0.30, the result shows that the intercept of the viewing time was longer in the PC condition than that in the CC condition. Last, the effect size of the interaction between the number of trials and the experimental condition was -0.17, meaning that the effect of the interaction was small. Hence, in the main experiment, there was no impact of the experimental condition on the slope of the viewing time.

As could be seen in Figure 13, the data suggests that the viewing times in the PC condition did not show a large decrease until around the 30th trial, afterwards showing a gradual decrease. However, the viewing times in the CC condition seemed to start decreasing from the beginning of the experiment. This might suggest that the slopes of the viewing times in the PC and CC conditions did differ at the beginning of the experiment. Consequently, another LMM using only the first 30 trials was performed. The structure of the fixed effects as well as the random effects were identical to the original model, which included the whole 120 trials. The lower part of Table 11 shows the results of the LMM. The result showed again a significant main effect of trial number, meaning that the viewing time decreased as a function of trial number in both experimental conditions. Further, there were a significant main effect of the experimental condition as well as the interaction between the trial and the experimental condition, suggesting that the intercept of the viewing time was longer in the PC condition than in the CC condition and the slope of the viewing times in the PC condition was more gradual compared to the CC condition. Considering the effect size for the main effect of the experimental condition and the interaction between the trial and the experimental condition were 0.62 and 0.52 separately, those factors had moderate effects on the viewing times.

Table 11. Fixed effects in the LMM predicting the viewing time in the main study

Fixed effect	Estimate	95 % CI		t value	Cohen's <i>D</i>	Pr (> t)
		Lower	Upper			
General impact of repeated exposure and experimental condition						
(Intercept)	-0.20	-0.48	0.09	-1.34	-	<i>p</i> = 0.19
Number of trials	-0.10	-0.15	-0.05	-3.77	-0.91	<i>p</i> < .001
PC condition	0.30	-0.08	0.68	1.56	0.38	<i>p</i> = 0.12
Number of trials * PC condition	-0.03	-0.10	0.05	-0.72	-0.17	<i>p</i> = 0.48
General impact of repeated exposures and culture in the first 30 trials						
(Intercept)	-0.81	-1.25	-0.37	-3.59	-	<i>p</i> < .001
Number of trials	-0.56	-0.77	-0.34	-5.09	-1.22	<i>p</i> < .001
PC condition	0.80	0.19	1.42	2.56	0.62	<i>p</i> = 0.02
Number of trials * PC condition	0.34	0.03	0.64	2.16	0.52	<i>p</i> = 0.03

3.3.3. Discussion of Study 2-2

The pilot and main study in Study 2-2 aimed to assess two main questions: 1. Does habituation process cause the decreased attention from repeated art viewing? 2. Does higher stimulus variability in the stimulus sequence counteract/delay the decreased attention over repeated art viewing? Here, we mainly discuss the results of the main study in order to avoid the redundancy, since the results of the pilot study is already discussed in the section, *Discussion of the pilot study*.

The first question was assessed by seeing if the viewing time recovered after the stimulus block changes in the PC and CC condition. We hypothesized that, if the habituation process is one of the possible causes of the decreased attention phenomenon, in the PC condition, the viewing time would show no recoveries after the stimulus block changes, since the changes themselves were not salient compared to the CC condition. In contrast, in the CC condition, the viewing time might show recovery, since the stimulus block changes were set more salient than in the PC

condition, which might lead the response recovery, dishabituation process, due to the novelty effect. The results of a series of *t*-tests did not show any significant results, meaning that our first hypothesis was not supported by the statistical tests.

Nevertheless, here, we further discuss the results considering the data shown in the figures as well as the effect sizes. First, as shown in Figure 11, in the PC condition, the mean viewing time after the stimulus block changes had mainly two patterns: 1. stayed at the same level as before the changes (in the second and fifth block changes), or 2. decreased compared to before the changes (in the first and forth block changes), with one exception in the third block change, showing longer viewing time after the block change. In contrast, as shown in Figure 12, the results of the CC condition had different pattern: increased compared to before the changes (in the first, second, forth, and fifth changes) with one exception in the third block change, showing shorter viewing time after the block change. The different trends in the mean viewing times before and after the block changes between the two conditions can also be seen from the effect sizes. For example, in the PC condition, the first and forth block changes had small effect sizes ($d = .19$ and $.24$ respectively), which led to the shorter viewing times after the block changes, while those in the CC condition also had small effect sizes ($d = 0.14$ and 0.13 respectively) with different directions, leading to longer viewing time after the changes. Likewise, in the PC condition, the second block change had no effect on the viewing time, while that in the CC condition had small effect size ($d = .17$), leading to longer viewing time after the change. Second, regarding the Figure 11 and 12, another note should also be made on the size of the variances in the mean viewing times especially in the CC condition. It is quite visually obvious that especially the variances in the viewing time after the second and forth block changes in the CC condition are quite large compared to the mean viewing times in other block changes.

Combining these results, the generally longer viewing times after the block changes as well as the bigger variances in some blocks in the CC condition might suggest that we could not detect *significant* viewing time recoveries in the CC condition, since the changes or the stimulus novelty

after the block changes might not have been equally salient to each participant. Specifically, there might have been some participants who showed the recoveries after the block changes in the CC condition, while there were also some who did not show any recoveries, leading to a larger variance in the viewing times in some blocks. Especially considering that there were no art experts included in our samples, it might be possible that some of the participants did notice the change/novelty itself, but some did not. We could not tell if each participant was really aware of the changes of the stimulus blocks, since we did not include an assessment, such as asking if there were any changes in the presented sequences and how many times it happened. Also, it is hard to tell if they had enough knowledge to see the difference in the blocks, since we did not directly measure the level of art knowledge of our samples. Nevertheless, for future studies, it might be interesting to manipulate the novelty not only in the contents of the paintings, but also in the physical features of the artworks such as hue/saturation/contrast between the stimulus blocks, so that the block changes or novelty in the stimulus sequences could even more salient regardless of the knowledge on the artworks.

The second question was assessed by comparing the slope of the viewing time in the PC and CC condition. We hypothesized that, in the PC condition, the slope of the viewing time would be more gradual due to the novelty effect, which is known as a way to counteract habituation process. The results of the LMM showed that there was not a significant interaction between the number of trials and experimental condition. Thus, contrary to our second hypothesis, there was no significant difference in the slope of the viewing time in the two conditions. The LMM using only the first 30 trials showed that the viewing time was longer in the PC condition compared to the CC condition, and also the slope of the viewing time was more gradual in the PC condition. Combining all results together, it seems that the novelty in the stimulus sequences could counteract the decrease in the viewing time at the beginning of the art viewing task, but this effect did not last for the entire study. Considering that Leventhal et al. (2007) reported the novelty effect with 20 presentations of the

images, it seems plausible that the novelty effect could counteract the habituation process for only short while.

Overall, though our hypotheses were not fully supported, we could find some hints in the results of the main study. Indeed, the results should be further replicated. It is hard to tell the effect of any factor of interest with one single study. Nevertheless, it is expected that the present study, where we showed empirical studies built on the past knowledge from museum and psychology studies, will inspire future studies, following our footstep.

3.4. General discussion of Study 2

In Study 2, though the target phenomenon of Study 2-1 and 2-2 were different, they had a common motivation to study: to investigate the way to counteract the ever-presented attention decrease phenomenon as well as its possible causes in a lab-based setting. These attempts in a lab-based setting were especially motivated, since, as manipulations of the factor of interest and controlling confounding variables are quite difficult in actual museum/gallery settings.

Study 2-1 investigated if the ever-presented way to counteract this phenomenon, presenting the information about the artworks, could stop the decrease in viewing time with an updated procedure. The results showed that the manipulation did not stop the viewing time, but rather maintained the attention towards artworks in general in one condition and at the beginning of the task in another condition.

Study 2-2 investigated if one of the ever-proposed causes, habituation, attributes the decrease attention by observing dishabituation process in the viewing time, and if the novelty in the stimulus sequences could counteract the decrease in the viewing time. Though the results did not fully support our hypotheses, we could partially observe the dishabituation, and could confirm that the stimulus novelty in the presented sequences, though not fully, could counteract the decrease at the beginning of the presentation.

As could be seen in the introductions for Study 2, there has been only a few studies which empirically tried to assess the possible causes and ways to counteract this phenomenon. In this sense, Study 2 in the present study was an example that postulate a possible factor attributing this phenomenon and assess a certain manipulation could mitigate the decreased in attention with concrete theoretical background. However, as a general limitation in Study 2, we could not find a way to fully stop the decrease in viewing time. Indeed, as in Bitgood (2009a), the decreased attention phenomenon is a result of a combinations of multiple factors, and hence, it might be difficult to fully stop the decrease by focusing on one single factor. Since the studies in the present dissertation only focused on habituation (satiation) as a factor of interest, in the future, the effect of other causes as listed in Bitgood (2009a) such as information overload, objective competition, stress, and decision-making process should be further investigated. As a second general limitations, in Study 2, we only focused on the viewing times towards the artworks and did not examine how to mitigate the decrease in art appraisals. Considering that Study 1 suggested that not only the viewing times but also art appraisals might decrease over trials, it will be worth studying in the future, for example, the novelty effect would also be effective to the other metrics as well.

As pointed out in Davey (2005a), the further investigations on the causes of this phenomenon as well as how to counteract it should be elaborated with a collaboration between museum visitor researchers who are specializing in observing the museum visitors in an ecologically valid setting, and cognitive scientists/psychologists who are specializing in investigating complex human behavior with properly controlled and solid methods. As such, we hope that Study 2 will inspire future research by presenting a way to connect the problems/issues in museums to the lab-based paradigm, where the reality and basic research complete each other hand in hand.

4. Study 3—Cross-cultural similarity/difference in aesthetic evaluation and art viewing

In Study 3, the effect of culture on different aspects of art evaluation/behavior is further assessed. Specifically, three empirical studies are shown, examining (1) the universality of the preference/values for symmetry, (2) the difference in the artwork features attributing the art appraisal *beauty*, and (3) cross-cultural diversity in eye movement patterns while viewing art and non-art. As shown, each experiment included in Study 3 deals with different aspect of the art experience, and hence the detailed introduction, purpose, methods, and results of each study will be respectively presented in the following sections.

Nevertheless, the attempt to assess the effect of culture on art experience might be a challenge to the implicit premise which presumably most of the past studies in the field holds: the universality in art experiences. Thereby, by explicitly assessing the cultural effect on art experiences, it is expected to provide further knowledge in the field of empirical aesthetics and inspire future studies.

4.1. Study 3-1: the universality of the preference for symmetry

Manuscript in preparation

Leder, H., **Mikuni, J.**, Kawabata, H., Pelowski, M., & Rosenberg, R. (*in prep*). Symmetry as an Inter-cultural Feature Constituting Beauty: Implicit and Explicit Beauty Evaluation of Visual Symmetry in Japan.

My Contribution:

I shared the first authorship with HL. I collected the data and analyzed and interpreted the data together with HL. HL and I are the main writer of the current draft, elaborating the draft with HK, MP, and RR.

As a starting point for a challenge against the universality assumption in empirical aesthetic, we focused on one visual feature, *symmetry*. Not only in psychology, but also in other academic fields such as biology and physics, symmetry has been recognized as one of the most fascinating and important visual features in human perception and evaluation (Leder et al., 2019). For example, studies on facial attractiveness have suggested that people show their preference for symmetric faces/bodies (Grammer & Thornhill, 1994; Rhodes, 2006; Tinio et al., 2013), and this preference is associated with biological advantages related to perceived health as well as actual physical health (Fink et al., 2006) and good genes (Thornhill & Møller, 1997). Further, people show their preferences/values for symmetric features even when it does not signalize biologically important cues in the environment (Gartus & Leder, 2013; Höfel & Jacobsen, 2007; Jacobsen & Höfel, 2002; Tinio et al., 2013; Tinio & Leder, 2009). For instance, Jacobsen and Höfel (2002) showed 252 abstract patterns to their participants, half of which were symmetric, and the other half were asymmetric patterns. As a result, they found that most of the participants judged symmetric patterns as more beautiful than asymmetric ones.

Virtually all aforementioned studies had been conducted in Western cultures, and from this respect they do not talk much about symmetry as a universally preferred/valued feature. However, there have been a few studies including cross-cultural comparisons, providing evidence that there is an inter-cultural agreement on the preference/evaluations for symmetry (e.g., Bode et al., 2017; Makin et al., 2018; Uduehi, 1995; Weyl, 1952). Uduehi (1995) conducted a cross-cultural study recruiting participants from the U.S. and Nigeria, examining if there is a cross-cultural difference in the preference for symmetry. In this study, multiple visual patterns were shown to the participants. As a result, symmetry was a strong determinant of preference in both cultures. Second, Bode et al. (2017) compared the preference for symmetry (and complexity) between participants from the U.K. and Egypt. They assumed that the traditional styles of artworks in each culture might influence on the preference for symmetric patterns: “Indeed historically, representational art has not been practiced in the Islamic world, especially in relation to the depiction of humans and animals. This is

one reason that makes the cross-cultural comparison interesting” (Bode et al., 2017, p.397). As a result, they found “an overall pattern of preference for symmetry, and a dislike for asymmetry that was similar across the two cultures” (p.397). Makin et al. (2018) conducted a similar study, reporting again that the British and Egyptian participants preferred symmetric patterns over non-symmetrical ones.

Hence, past studies consistently shown the strong preference for symmetry with American and European participants and also suggested an inter-cultural agreement on the preference and higher values for symmetry. Thereby, it seems plausible to assume the universality, especially considering that symmetric features are tied to the importance in biological cues. However, interestingly, more recent studies demonstrated that symmetry is not always the determinant of preference and aesthetic judgements. For example, Gartus et al. (2020) studied how abstract asymmetric, symmetric, and “broken” patterns—“slightly deviating from symmetry as well as more representational patterns resembling faces” (Gartus et al., p.1)—evaluated from lay persons and art experts. In this study, participants were asked to rate the liking and interest for all patterns. As a result, they found profound differences in the rating scores between these groups of participants. While laypersons liked symmetric patterns more and found them more interesting than asymmetric patterns, art experts liked asymmetric patterns more and found them more interesting than laypersons did. As another example, Leder et al. (2019) reported that art experts and non-art experts evaluated the beauty of multiple patterns differently. In this study, they showed multiple patterns which were systematically varied according to symmetry (either symmetric or asymmetric) and complexity (either simple or complex) to their participants and asked them to judge the liking of the patterns. As a result, art experts did not prefer symmetric patterns over the non-symmetric ones. These findings are evidence against the assumption: symmetry is a universal law for preference and aesthetic values. Weichselbaum et al. (2018) further tested if lay persons (psychology students) and art experts (art history students) evaluate beauty for symmetric and asymmetric visual patterns differently in explicit and implicit evaluations—by directly asking participants to evaluate the

patterns' beauty as well as by an Implicit Association Test (IAT). In order to assess the individual art expertise level specifically, they asked the participants to fill in an art knowledge questionnaire. The IAT task revealed a strong implicit preference for symmetric patterns regardless of the individual art knowledge level. The findings of the IAT employed here are also in accordance with Makin et al. (2012) where they found an implicit positive response to symmetric stimuli. The results of explicit rating task showed that both laypersons and art experts preferred symmetric patterns over asymmetric ones. However, importantly, art experts evaluated the asymmetric patterns as more beautiful than laypersons. Thus, they found an interaction between the patterns' symmetry and art expertise level. Note that, this interaction was not found in the results of the IAT task. They concluded that there is, indeed, general preference for symmetry independent from art expertise. However, experts tended to emphasize the beauty of asymmetrical depictions, eventually resulting in different criteria of aesthetic judgements in explicit evaluation.

Combining all results together, it seems that, indeed, humans generally share preference/values for symmetric visual features regardless of the cultural background. Weichselbaum et al., (2018) further discussed that "symmetry preference is based on evolutionary adaption, similarly for almost everybody" (p. 16). However, considering that there was a variety of results depending on inter-individual factors i.e., the level of art expertise, it might be possible that the importance of symmetry for preference and/or aesthetic judgements could also differ to a certain amount between individuals. Weichselbaum et al., (2018) argued that familiarity might explain the findings from art experts, who might see deviations from symmetry in artworks more often than laypersons. Hence, art experts might be more familiar to the asymmetry features and be able to process asymmetry more fluently, leading to their higher evaluation for asymmetry. In other words, difference in the bottom-up process—more exposure towards asymmetric features—might eventually have had an impact on the evaluation for visual symmetry.

This argument might even coin some questions for the past findings regarding the inter-cultural agreement in the preference/evaluation for symmetry. Specifically, though past studies

(e.g., Bode et al., 2017; Makin et al., 2018; Uduehi, 1995) found inter-cultural agreement for preference for symmetry across U.S., U.K., Nigeria, and Egypt, they did not have specific hypotheses on why the comparisons between the cultures might challenge the universality claim. Indeed, Bode et al., (2017) explicitly stated that the comparison between U.K. and Egypt is interesting, since the forms of art between the two cultures are quite different. Nevertheless, in Islamic art, the symmetrical features are also quite popular, as we can see in the shapes in their famous pyramids as well as in the geometric mosaic patterns. Hence, in this sense, the participants from the two cultures are assumed to be equally exposed to symmetry. In other words, they share a similar visual culture regarding symmetry. Thus, in order to claim that humans truly share universal aesthetic attitude for symmetry, the above evidence seems to be a bit weak.

Regarding the challenge against the universality claim for symmetry, a replication in Japan could bring a new insight to this topic. Japan is especially an interesting candidate for comparing the results to those from the past studies in Western, Islamic, and African cultures, since Japanese aesthetic culture has a special relation to asymmetry. For instance, in traditional Japanese aesthetics, '*wabi-sabi*' is an attitude that demands the acceptance of imperfection. According to this view, beauty is possible through imperfection, impermanence, or incompleteness. More concrete features of the '*wabi-sabi*' is that aesthetic comprise (among other principles) asymmetry, and roughness. Grounded in Zen aesthetics, the principle of '*Fukinsei*', which might be translated to asymmetry or irregularity, is proposed to produce a composition by explicitly using elements of irregularity and asymmetry. In Japanese pottery and brush paintings, imperfection of shapes often is considered as a beautiful, valuable, and positive aesthetic feature. Further, in Japanese design the term "*Hacho*" means intentional unevenness. Similarly, "*kintsugi*" is a special way of fixing cracks in plates/cups/pots often with gold inlays, which makes the distortion salient, yet aesthetically valuable. All these principles could be found in everyday lives in Japan e.g., when people see artworks, when they use tableware, in the flower arrangements in their houses. Consequently, this appreciation for asymmetry is, perhaps implicitly, implemented in their visual culture, and thereby,

it can be assumed that Japanese people might be more familiar to asymmetric features and be able to process asymmetry more fluently than people from other cultures where they appreciate symmetry more over asymmetry.

In the present study, we assessed the aesthetic judgements from Japanese laypersons for symmetric/asymmetric visual features via explicit and implicit assessments, largely following Weichselbaum et al. (2018) where students in Vienna (Austria) were participating. By conducting a follow-up test with Japanese participants, we aim to challenge the universality claim regarding symmetry. If human share the preference/values for symmetry over the process of evolution and adaptation, we expect that Japanese participants would also show both implicit/explicit preference/values for symmetry. In contrast, if the preference for symmetry is not shared by humans but modulated by an external factor as suggested for the Western art experts i.e., exposures to visual features in everyday lives, we expect that the Japanese participants would show the implicit/explicit preference/values for asymmetry. Finally, as a hybrid of the two patterns, if humans share the preference/values for symmetry over the process of evolution but this can be also modulated by the exposure to the visual features, we expect that the Japanese participants would show their implicit preference/values towards symmetry but show their explicit preference/values for asymmetry.

4.1.1. Method

Participants

Thirty-one students at Keio University (Japan, Tokyo) participated in the present study (17 females, $M_{age} = 21.25$, $SD = 1.12$, age range 19-25 years). All participants were native Japanese speakers and had normal or corrected-to-normal vision. Participants received ¥1,000 (about 10\$) for their participations. Importantly, none of the participants had academic background of art related field. All filled out the consent before they started the experiment and were allowed to

withdraw the experiment at any point for any reason. However, no participants quit the experiment, and thus, the data from all participants were used in the analysis below.

Stimuli and Apparatus

160 visual patterns were used as our stimuli. Half of the stimuli were symmetric, and the other half were asymmetric patterns acquired from Jacobsen and Höfel (2001, 2002). Note that, in Weichselbaum et al. (2018), there were only 80 visual patterns included in the stimuli. However, while they have a large number of participants in total ($N = 79$), we did not have enough resources to recruit that many participants, ended up with the final 31 participants. Consequently, in order to achieve a higher statistical power to perform analysis using LMM, we used the double number of stimuli.

For the IAT task, as with Weichselbaum et al., (2018), 10 positive- and 10 negative-valenced words were used as word stimuli from the BAWL-R (Vo et al., 2009; Võ et al., 2006). All words were translated from German to Japanese. The words stimuli are presented in Table 12.

To assess the individual art-knowledge level as in Weichselbaum et al., (2018), we used the art interest and knowledge questionnaires included in VAIK (Specker et al., 2018). All questions previously used in Weichselbaum et al., (2018) are included in the VAIK. Different from Weichselbaum et al., (2018) where they only asked their participants to fill the art knowledge questionnaire, we used the full version of the VAIK, which also includes the questionnaire measuring art interest level.

All tasks were conducted on a 24.5-inch monitor (AW2518H, Dell) with a resolution of $1,024 \times 768$ pixels, individually in a test room at Keio University (Tokyo, Japan). Both experimental tasks of the explicit rating and IAT were controlled by PsychoPy 3.0 (Peirce, 2007), and Google Form was used for collecting responses of the VAIK from each participant.

Table 12. The list of word stimuli used in the IAT task in English and Japanese

Words with positive valence		Words with negative valence	
English	Japanese	English	Japanese
savior	救世主	tyrant	暴君
cheer	元気	penalty	罰則
talent	才能	atrocious	残虐
celebration	祝福	grief	悲嘆
party	パーティー	victim	犠牲
profit	利益	exhaust gas	排気ガス
humor	ユーモア	fraud	詐欺
hobby	趣味	plague	疫病
journey	旅行	arrest	逮捕
joy	喜び	command	命令

Note: Only the Japanese translation of the words were used in the present study.

Procedure

The experimental procedure largely followed Weichselbaum et al., (2018). The experiment consisted of three parts: the VAIK to classify participants' art interest and knowledge, explicit beauty rating task, and an IAT task for the measurement of implicit association between symmetry/asymmetry and positive/negative. The VAIK questionnaire was asked always after the participants finished the IAT task and the explicit rating task. In the following sections, the detailed procedure in each task is described.

The VAIK questionnaire

The participants answered the questions in VAIK (Specker et al., 2018) after they finished the IAT/explicit rating task. The questionnaire can be divided into three parts. One part is assessing the general art interest, followed by additional two parts assessing the knowledge on art. The questions included in the VAIK were developed based on the ones which are used in

Weichselbaum et al., (2018), and thereby, in the present study, we used the updated version of the questionnaire after the proper validations.

As aforementioned, in the present study, we assumed that all native Japanese participants are more familiar to the asymmetric patterns, since this familiarity is rooted in the visual culture in Japan which is presumably shared by all the members of the society. Hence, in the present study, the art interest/knowledge scores were only analyzed in order to see if there are Japanese participants who are highly familiar with Western artworks and art history, since this deep knowledge on the art contents/styles/history of the Western cultures could skew the familiarity towards asymmetric features, which might eventually conflate the results. To this aim, the VAIK (Specker et al., 2018) is the best questionnaire, since the questions aim to assess the expertise in Western artworks.

Implicit Association Test (IAT)

As in Weichselbaum et al., (2018), in the IAT task, 10 positive and 10 negative words as well as 10 symmetric and asymmetric patterns were used as the stimuli. 10 symmetric/asymmetric patterns were randomly chosen from the 160 visual stimuli. The presented patterns in the IAT task were identical for all participants.

Before they started the IAT task, the participants were instructed to press a key (the left or right arrow key) in response to the property of each stimulus, either the valence of a word or the symmetry of a pattern. The mapping of the key presses was counterbalanced across participants. Further before the main trials in the IAT tasks, they completed a training either for patterns or words, followed by one main experimental block. After the first main trials, the participants again completed another training block also either patterns or words depending on which training/main trials were presented at the beginning, followed by the second main experimental block. In each trial, one word (positive/negative) or pattern (symmetrical/asymmetrical) was presented on the screen center. Participants had to press the right or left arrow key in response to the stimulus. Before the next trial

started, a gray screen was presented for 250ms. If the participants made a wrong answer, the Japanese word for incorrect, “間違い”, was presented on the screen monitor for 500ms, then another 250ms gray screen was presented before the next trial started.

The main experimental block was either a compatible or incompatible block, and the key mapping was the same in the training trials right before the experimental block. According to Weichselbaum et al., (2018), this version of IAT is “a variant of the original method by Greenwald et al. (1998) and the depiction is inspired by the one used in Makin et al. (2012)” (p.8), and the present study followed the same procedure in order to make the results more comparable to the original study.

Explicit rating task

In the explicit rating task, participants were asked to judge ‘how beautiful the presented pattern was?’ with a 7-points Likert-type scale (1: *not beautiful at all* to 7: *very beautiful*). Each trial was started with a fixation cross presented on the screen center for 200ms, followed by a presentation of one pattern. A pattern was presented at a time for 3 s, and the participants could only see one pattern for this fixed duration. After the presentation of a pattern, rating scale asking the beauty of the pattern was presented, and the scale was shown on the screen until the participants made a response. Finally, before the next trial started, a gray blank screen was presented for 2.5s, and the participants repeated this until all stimuli were presented. The presentation order of the patterns was randomized.

4.1.2. Results

The VAIK questionnaire

In the VAIK (Specker et al., 2018), there are two parts aiming to assess the art interest and knowledge respectively. The art interest score can range from 7 to 77. With our participants, the art interest score ranged from 15 to 59, with a mean of all participants 35.42 ($SD = 12.86$). The art knowledge score can range from 1 to 36. With our participants, the art knowledge score ranged

from two to 25, with a mean of all over participants 7.32 ($SD = 4.67$). Thus, none of the Japanese participants' art knowledge scores were over the 70% of the possible highest score (25.20). Further in Weichselbaum et al., (2018) using the same art knowledge questionnaire, the mean art knowledge score in lay group was 12.67. Given the results above as well as none of our participants had academic education background in art-related field, it is plausible to postulate that there were no participants who are extremely familiar/knowledgeable with Western artworks.

Note that the art interest scores ranged more widely than art knowledge scores. The questions which assessed the art interest included e.g., *I'm interested in art, how often do you visit art museums and/or galleries?* Accordingly, it is possible that people with higher art interest scores were exposed more to Western artworks and/or were reading art educational texts in museum exhibitions, which can be seen as a partial art training. Consequently, the art interest scores will be used as one of the control factors in Linear Mixed Models in the following analysis.

Implicit Association Test

In the analysis for IAT task, all 31 participants' data was included, since none of them made more than 20% error. The error rate per participant ranged between 1.25% and 11.87%. If the participants made a wrong answer or if the response time was exceeding 2.5 SD s from the mean of the correct response times per participant as well as per condition, the trials were excluded from the analysis below. With these criteria, in total, 8.20% of the trials were excluded. We followed the same data exclusion criteria as in Weichselbaum et al. (2018).

To assess the impact of the task compatibility on the response times, we ran a LMM. In the model, response time to make a key press was set as the dependent variable. A treatment coded contrast comparing the task compatibility, with compatible (baseline) against incompatible, the task switching, with switch (baseline) against repeated trials, and the interaction between the two variables were included as the fixed effects in the model. Following the model structure in Weichselbaum et al. (2018), for the random effects, we included random by-participants intercepts

and slopes for compatibility and task switching, random by-stimuli intercept and slopes for compatibility, and random intercept for centered sum of the art interest score were included. Note that random by-stimuli intercept and slopes for task switching was not included in the final model, since otherwise the model did not converge. Note also that all response times were log-transformed to ensure the data normality. All fixed effects and the results are shown in Table 13. The results showed that a significant effect of task compatibility, indicating that the response times in the compatible block were faster than those in the incompatible block. Also, the effect of task switching was significant, suggesting that the response times for the repetition trials were faster than those in the switching trials. Further, the interaction between the task compatibility and the task switching was also significant, which was the same direction as in Weichselbaum et al. (2018). This interaction suggests that the slower response brought by the task switching was larger in the incompatible condition than in the compatible condition. In sum, our Japanese participants made more faster correct response for the compatible blocks, where they pressed the same arrow key for symmetrical patterns and positive words, and asymmetrical patterns and negative words, showing an implicit association between symmetry and positive words, and asymmetry and negative words. These results were in line with the ones from participants in Vienna in Weichselbaum et al. (2018).

Table 13. Fixed Effects and Results of the LMM for the IAT test.

Fixed effect	Estimate	95 % CI		t value	Cohen's <i>D</i>	Pr (> t)
		Lower	Upper			
(Intercept)	-0.21	-0.29	-0.13	-5.24	-	$p < .001$
Task compatibility - incompatible	0.34	0.28	0.40	10.66	3.71	$p < .001$
Task switching - repeated	-0.21	-0.24	-0.17	-12.15	-4.45	$p < .001$
Task compatibility * Task switching	-0.24	-0.28	-0.20	-11.71	-0.35	$p < .001$

Complementary analysis

Following Weichselbaum et al. (2018), we also computed D scores, which were used as a standard measure for the IAT effect (Makin et al., 2012) for each participant: “They represent the difference between incompatible and compatible RTs divided by the pooled SD per participant” (Weichselbaum et al., 2018, p.14). The mean D scores across all participants was 0.70 ($SD = 0.34$). Further, one-sample t test using individual participants’ D scores showed that the mean D score was significantly differed compared to 0 ($t(30) = 11.41$, 95% CI[0.57, 0.83], $p < .01$). This result indicates that the response times included in the compatible block were shorter compared to those in the incompatible block. Only two participants showed negative D scores, indicating faster RTs in the incompatible block. Given the result of LMM together, as with the participants in Weichselbaum et al. (2018), our Japanese participants also showed implicit association between symmetric and positivity.

Explicit Rating task

To assess the impact of the patterns’ symmetry on the explicit beauty rating, we ran another LMM. The mean beauty rating scores for symmetrical/assymetrical patterns are shown in Figure 14. In the model, explicit beauty rating scores for patterns was set as the dependent variable. For the fixed effects, a treatment coded contrast comparing the patterns’ symmetry (baseline) against asymmetry was set. For the random effects, we set random intercepts for the interaction between participants and patterns’ symmetry, for the interaction between art interest scores and images, and for patterns’ complexity. Though we largely followed the model structure of Weichselbaum et al. (2018), the models including by-participants intercepts and slopes for patterns’ symmetry and by-stimulus image intercepts and slopes for art interest scores did not converge. Hence, we simplified the random effect structure in our final model. All fixed effects and the results are shown in Table 14. The LMM showed a significant main effect of pattern’s symmetry, suggesting that participants evaluated asymmetric patterns less beautiful by 1.51 scale points compared to the symmetric ones.

Hence, the Japanese participants also explicitly preferred symmetric patterns over the asymmetric ones as the participants from the Western, Islamic, and African cultures.

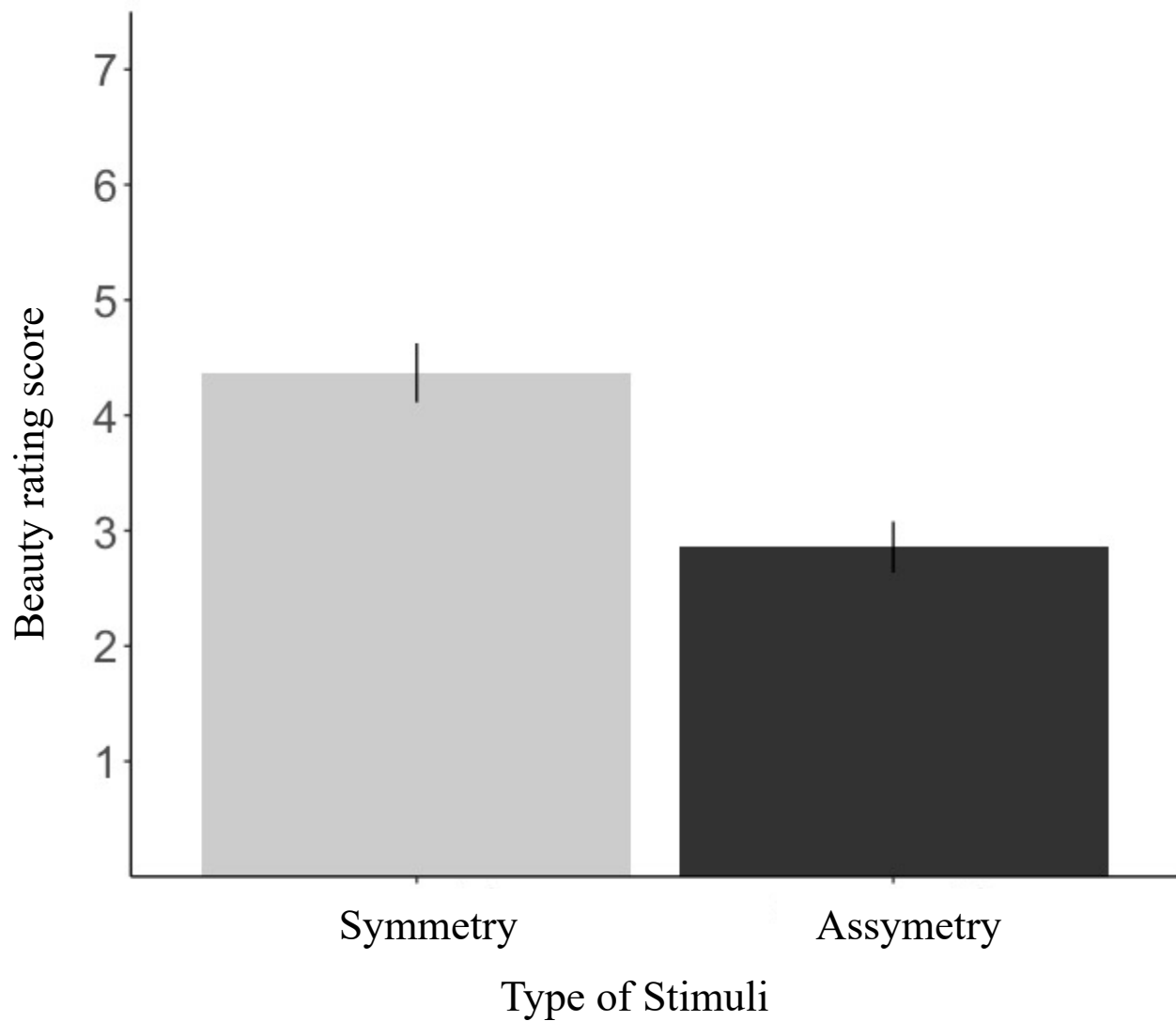


Figure 14. The mean beauty rating scores for each patterns' symmetry. The error bars represent the *SE* for each bar.

Table 14. Fixed Effects and Results of the LMM for the Explicit ratings.

Fixed effect	Estimate	95 % CI		t value	Cohen's <i>D</i>	Pr (> t)
		Lower	Upper			
(Intercept)	4.37	3.98	4.76	24.66	-	$p < .001$
Pattern's symmetry - assymetry	-1.51	-1.85	-1.17	-8.84	-2.27	$p < .001$

4.1.3. Discussion

In the present study, we aimed to challenge the universality claim regarding symmetry by testing implicit/explicit evaluations for symmetry/asymmetry features with participants from Japan, where a deviation from symmetry is one of the important principles of beauty. The results of IAT task as well as explicit rating task are going to be discussed by comparing them to the results in Weichselbaum et al. (2018).

Regarding the results of IAT task, our results showed that the Japanese participants made significantly faster response in the compatible block, where the same key responses were required for symmetric patterns and positive words, asymmetric patterns, and negative words, than in the incompatible block, where the same key responses were required for symmetric patterns and negative words, asymmetric patterns, and positive words. This significant main effect of block compatibility was also found in Weichselbaum et al. (2018). Thus, there was no inter-cultural differences. Importantly, in the analysis for the IAT task using LMM, the main effect of task switching (either repeat or switch) as well as the interaction between the task switching and block compatibility were especially included in the model, so that “we made sure that our IAT task measure symmetry preference in an implicit way rather than only by means of a deliberate recording of the picture- versus word- discrimination responses.” (Weichselbaum et al., 2018, p.16). As with Weichselbaum et al. (2018), we did find the interaction between the two factors, suggesting that the task switching effect e.g., when the task was switched from a word to a pattern, the delay in the response became bigger compared to when the same task was repeated (a word trial

to another word trial). Nevertheless, the main effect of the task switching was also significant in the compatible block in the same direction, and thereby, we also postulate that the IAT effect as an implicit measure is valid. Regarding the results of explicit rating task, our results showed that the Japanese participants evaluated symmetric patterns as more beautiful compared to the asymmetric ones. Thereby, again, as with Weichselbaum et al. (2018), there was no inter-cultural differences on explicit rating for symmetry. To summarize the results, though Japanese participants were assumed to be more familiar to asymmetric visual features, they showed implicit preference and higher explicit beauty evaluation for symmetry. Thus, as argued by the past researchers, our results support the assumptions that, indeed, humans from different culture share the aesthetic values for symmetry. This result might even foster the argument that the response to symmetry could be biologically hard-wired.

Perhaps, our results could also put further insight on a variety of results in the past studies, while laypersons showed implicit/explicit preference for symmetry, art experts tended to emphasize the beauty of asymmetry in explicit measures. There might be two possible explanations on why the art experts emphasized the beauty of asymmetric objects. Firstly, as argued in Weichselbaum et al. (2018), it could be possible that art experts are more exposed to the deviations from symmetry in artworks. This bottom-up process (i.e., visually exposed to more asymmetric features) might lead to a higher familiarity for asymmetric objects and to a more fluent processing of asymmetry, eventually resulting in emphasizing the beauty in explicit measures. Secondly, contrary to this bottom-up explanation, Lindell and Mueller (2011) suggested more top-down based account. According to them, “the importance of symmetry on judgements of aesthetic beauty decreases as artistic training increases” (p. 459). Hence, in their explanation, more top-down information/knowledge might produce this difference. Nevertheless, these two accounts have not been investigated in detail in the past studies, since the Western art experts are typically exposed to the asymmetry as well as are taking art training. Thus, it was hard to examine the effect of one factor per se. However, our Japanese participants were assumed to be highly familiar to asymmetry,

while not taking any art trainings as could be seen in the scores in art knowledge. Further, they did not show any explicit emphasis on beauty in asymmetry. Hence, all together, it seems that the later account: not the familiarity towards asymmetric features, but rather artistic trainings would impact the explicit evaluation for symmetry holds, since, if the exposure and the higher familiarity are the important factor, our Japanese participants should also have shown the higher beauty rating score for asymmetry. Thereby, it seems plausible to conclude that, humans indeed share preference/values for symmetry, yet education/knowledge on art could modulate the values for symmetry, only explicitly.

Importantly, a brief note should be made on the impact of object classes. For example, Little (2014) suggested that people would show different levels of preference for symmetry depending on the classes of object. More specifically, he showed that the preference for symmetry was much smaller for artworks compared to faces. As such, considering that the stimuli used in the present study was a series of geometric patterns, it is still an open question if people would show a similar result when evaluating more complex stimuli e.g., artworks, scenery. For the future study, it would be an interesting question if the symmetry in more complex stimuli would show similar effect as in the present study or not.

To conclude, indeed, the present study showed that there is a potential universality in aesthetics. As pointed out, the present study used rather simple geometric patterns as our stimuli, and thereby, potential effect of culture on especially experience in art was not assessed. In the following section, Study 3-2, we go one more step further, and present an explorative study where we assess if cultures have an impact on art evaluation.

4.2. Study 3-2: The differences in the concept of beauty across culture

The state of the research

This is a cross-cultural study where the Ph.D. candidate was acted as the first author, collaborating with researchers in the University of Vienna (Spee, B., Leder, H., Steyrl, D., Scharnowski, F., Nakamura, K.), in Waseda University (Watanabe, K.) and in Keio University (Kawabata, H.). At the moment, we finished recruiting participants in Japan (Tokyo) and Austria (Vienna), and now are in the phase of analyzing data. Though the manuscript of this study is not ready yet, since the topic of this study is related to the dissertation and this study also represents one of the first author work of the Ph.D. candidate, I present the results, though partially, in this dissertation. Note that it might be possible to present full results in the public oral defense of the thesis.

In Study 3-1, we examined the evaluation for symmetry/asymmetry from Japanese participants, hypothesizing that they would evaluate asymmetry as more beautiful, since asymmetry is one of the principles of beauty in Japanese aesthetics, which might lead to the higher exposures and easier process for processing asymmetry. In other words, Study 3-1 tested a specific hypothesis (i.e., Japanese people might prefer and evaluate asymmetry as more beautiful due to the higher exposures to asymmetry) which might bring inter-cultural differences in aesthetics. Since past studies have extensively discussed the possible universality in the preference/values in symmetry, we could build a concrete hypothesis based on the past findings and conduct a study with so-called *hypothesis-driven approach* (Sutherland et al., 2017) i.e., focusing on one specific feature (dependent variable) and target factor (independent variable), building a concrete hypothesis for the relationship between the variables, and testing the effect of the target factor on the given feature.

As reviewed in the general introduction, regarding other properties, features, or aspects which bring aesthetic values, it is still hard to see if they are appreciated universally or to make concrete hypotheses on inter-cultural variations, due to the limited number of the research in the past (Che et al., 2018). What we see as a result is, for example, while there are famous quotes like *beauty is in the eye of the beholder*, *standards of beauty are cultural artifacts*, we do not know much about what factors in culture/what difference in culture might bring the different standards of beauty. In fact, while questions regarding the universality in physical attractiveness such as faces, bodies have been extensively assessed in the past (e.g., Broer et al., 2014; Cunningham et al., 1995; Furnham & Baguma, 1994; Rhee, 2018) similar attempts for artworks have little been studied. Of course, hypotheses could also be built not only by the results of past studies but also by personal experiences and/or discussions in other field of research. Nevertheless, further effort to systematically describe if culture has significant impacts on aesthetics as well as to consider what differences in culture might bring the differences in aesthetics might be a key for future studies to build interesting hypotheses and expand possible discussions.

In Study 3-2, we tried to capture and observe how culture has impact especially on beauty appraisal towards artworks. More specifically, we listed art attributes which are the most commonly used in the past studies in empirical aesthetics and investigated (1) if there is significant effect of culture on beauty appraisals for artworks, and (2) if there are inter-cultural differences in significant art attributes for predicting beauty. Along with the other studies in the present dissertation, the beauty appraisal was set as the main dependent variable in Study 3-2, since, indeed, beauty has been a core concept in empirical research of aesthetics (e.g., Jacobsen et al., 2004; Leder et al., 2004). Also, as mentioned in Che et al. (2018), considering that past studies with cross-cultural components in empirical aesthetics mainly focused on preference/liking towards art, systematic observations for the cross-cultural differences in beauty appraisal are expected to add another layer of knowledge in the field.

This study is rather explorative and inductive, and thereby, we did not have clear hypotheses on in which factors we would find inter-cultural differences and why. It is important to note that we do acknowledge that the research based on hypothesis-driven approach as a powerful tool to reveal the relationship between two variables in past research. Studies regarding the effect of culture on aesthetics should also be elaborated in the future with this approach in order to see the impact of specific factor(s). Nevertheless, a systematic and quantitative approach in Study 3-2 to investigate inter-cultural differences might shed the light on inter-cultural differences in aesthetics.

4.2.1. Method

Participants

This study involved a final sample of 102 native Japanese students (56 female, $M_{age} = 21.28$, $SD = 2.10$) from Waseda University (Tokyo, Japan) and 85 native German students from the University of Vienna in Austria (58 female, $M_{age} = 23.79$, $SD = 3.43$). All participants had normal or corrected-to-normal vision. It is important to note that the study setting was different in the two culture groups. Due to the COVID-19 pandemic, it was not possible to conduct a study in person in

Waseda University. Consequently, while the experience was conducted in person in a testing room in the University of Vienna for native German speakers, the native Japanese speakers from Waseda University joined an online study. Participants in Waseda University were recruited via the website of the university, while those in the University of Vienna were recruited via participants' pool in the university. Participants from Waseda University were paid 3,000 Japanese yen (approximately \$30 US) for their participations, and those from the University of Vienna got course credits for their participations. The study was approved by the local ethical committee of Waseda University (Japan) as well as of the university of Vienna (Austria), following the declaration of Helsinki. Note also, the final sample in both culture groups was reduced from an initial collection of 143 and 92 participants from Waseda University and the University of Vienna respectively due to issues with data collection (i.e., some participants did not complete the online study, participant inputs were not recoded correctly) or to having neither Japanese nor German as their native languages. In total, 41 participants from Waseda University and 7 participants from the University of Vienna were excluded from the following analysis.

Stimuli and Apparatus

We used 54 images of paintings, which were chosen from the Vienna Art Picture System (VAPS: Fekete et al., *under review*). This picture system contains 999 fine art paintings from Western cultures with several ratings for each picture (i.e., liking, valence, arousal, visual complexity, and familiarity). In order to make sure that we have artworks, representing a wide range of styles, contents, and period, the final set of the stimuli contained equal number of art motives (i.e., portrait, landscape, and still-life), of art categories (i.e., representative; Baroque and Rococo paintings, impressionist; Impressionism and Post Impressionism paintings, and abstract art; Cubism and Surrealism paintings). Further, since the ratings for each picture are available in the VAPS, we applied Shapiro-Wilk tests for each rating scale to see if the ratings are normally distributed e.g., it might impact the results if we only have paintings with quite high liking scores.

Overall, we found that liking ($W = 0.98, p = .52$), valence ($W = 0.98, p = .54$), arousal ($W = 0.95, p = .06$), visual complexity ($W = 0.95, p = .054$) were normally distributed. However, only familiarity rating was not normally distributed ($W = 0.74, p < .01$). Considering that the mean familiarity rating across all raters in the VAPS were 1.37 ($SD = 0.81$) and that all of them were psychology students, the ratings were skewed to the lower scores due to the limited knowledge for the artworks.

Our main dependent variable, *beauty*, was assessed via a question; *How beautiful do you find the artwork?* For the independent variables, we used the identical list of art attributes which was used in Spee et al. (*submitted*), representing 17 most influential ones for rating artworks in the past studies (e.g., Arnheim, 1957; Berlyne, 1975; Chatterjee et al., 2010; Leder et al., 2004; Pelowski et al., 2016b). The full list of art attributes is presented in Table 15. Note that all attributes were translated to Japanese and German versions, and were presented in accordance with the native languages of the participants. Note also that though we also asked the participants to evaluate impressions/preference of the artworks other than beauty (i.e., liking, understanding, interesting), the relationship between beauty and other impressions is under analysis at the moment. Thereby, the dissertation will only focus on the analysis for the relationship between beauty appraisal and the art attributes.

Since the native Japanese speakers participated in an online study, they joined the experiment from their own laptops. Hence, the size of PC screen was different across the participants. In contrast, native German speakers were coming to the testing room in the University of Vienna, and thereby, the testing setting was identical for all participants. In the testing room, the pictures were presented on a screen monitor (19" Iiyama ProLite B1906S, 1280×1024 pixel), and each participant sat in front of the screen monitor at a distance of about 80cm. For both Japanese and German native speakers, the experiment was controlled by an online questionnaire tool Sosci Survey (Leiner, 2019, <https://www.soscisurvey.de>). Further, for both Japanese and German native speakers, the horizontal dimension of paintings were fixed at 500pixels.

Table 15. 17 art attributes used in present study to predict beauty.

Art attributes	Presented scale to the participants	
	Negative pole (minimum value)	Positive pole (maximum value)
<i>Formal perceptual attributes</i>		
1. Visual harmony	visual harmony, proportional	peculiar, strange shapes
2. Depth	two-dimensional	three-dimensional
3. Complexity	simple	complex
4. Color saturation	soft, pastel	intense, strong
5. Color variety	less colors	many colors
6. Color temperature	warm colors	cold colors
7. Color world	dark color world	light color world
8. Brushwork	fine brushwork	rough brushwork
9. Utilization of drawing area	little utilization of the painting area	all utilization of the painting area
<i>Content representational attributes</i>		
10. Abstraction	representative	abstract
11. Realism	realistic content/topic	imaginary, unreal, fantastic
12. Symbolism (ambiguity)	distinct/clear (unambiguous interpretation)	symbolic (more room for interpretation, ambiguous)
13. Accuracy of depicting objects	photorealistic	painterly
14. Liveliness/animation	dynamic	still
15. Emotional expressiveness	emotionless	emotionally loaded
16. Valence	negative valence	positive valence
17. Focus	much context/environment in the image	focused content

Procedure

Regarding the native Japanese speakers, they joined the study via URL link, which was provided by the experimenter. Before they started the online study, they consented that they agreed with their participations, the data use and they understood the task in the online study. Only if they agreed with all options, they started the experiment. Regarding the native German speakers, they came to the testing room at the University of Vienna and made a written consent to participate the study. Following procedure was identical for both native Japanese and German speakers.

Before they started the main part of the experiment, they filled the demographic information such as age, gender, their mother tongue in order to see if the cross-cultural comparisons were correctly done. When the participants reached the main part of the experiment, the instruction was presented, saying they are going to see a series of artworks, and are going to be asked to rate all of the rating scales/dimensions for each picture. In the rating task, one painting was presented at a time with all of the scales/dimensions presented in Table 15. The image was fixed on the screen center, and thereby, participants could always see the presented painting. When participants scrolled the mouse, they could see more scales/attributes to be rated. The participants could make a rating for each picture for sliding the scale (i.e., 100point Likert-type scale ranged from 1 to 100) for each art scales and attributes. For the art scales (e.g., beauty, liking etc.), the edge of the left side of the scale implied 1 (not at all), and the edge of the right side implied 100 (very much so). For the art attributes (e.g., few colors – many colors, warm colors – cold colors etc.), the point in the middle of the scale (50) represented the values which are equally far from both poles. Taking one art attribute, few colors – many colors, as an example, the edge of the left side represented that the painting indeed used few colors, while the edge of the right side represented that the painting indeed used many colors. The presentation order of the artworks as well as of the scales/attributes were randomized. When the participants did not answer some questions, an instruction, “please answer all of the questions” was presented on the screen monitor. The participants repeated this procedure until all the 54 paintings were presented. Following the rating task, the participants answered the art interest and knowledge questionnaire (VAIAK, Specker et al., 2020).

4.2.2. Results

To assess if culture has an impact to predict beauty appraisals and if the art attributes which significantly predict beauty are difference across the two cultures, we conceived a three-steps analysis. First, we assessed if machine learning can predict beauty rating scores from art attributes at all and if the native language of the art viewers is important for predicting beauty ratings (Step

1), and then, we assessed which art attributes are significantly contributing for the prediction of the beauty ratings using data from native Japanese (Step 2.1) or German speakers (Step 2.2) separately. Finally, we compared the results from Step 2.1 and 2.2 statistically in order to find differences in the art attributes contributing to predict beauty rating between native Japanese and German speakers (Step 3). In the following section, the result for each stage will be presented separately.

Step 1. Assessing if machine learning can predict beauty rating scores from art dimensions and if the native language of the art viewers actually have significant impacts on predicting beauty ratings

First, we examined if the art attributes as well as the mother tongue of the art viewers are the significant predictors for the beauty judgement. If our main independent variables do not explain significant amount of the variance in the data, any speculations on the attributes predicting beauty and the difference in the two groups seems not to be warranted.

To this aim, we trained a predictive regression model for the beauty rating scores using beauty rating scores from all participants as the dependent variable, and all 17 art dimensions as well as the mother tongue of the participants (either Japanese or German) as the independent variables. Further, the contributions of independent variables to the final scores of the beauty rating e.g., how much the mother tongue of the art viewers contributed to predict the beauty ratings, were examined with SHAP (SHapley Additive exPlanations, (Lundberg & Lee, 2017)).

The regression task was carried out with extreme randomized trees (ET) ensembles. This machine learning method is computationally efficient and generally highly accurate, since it can model non-linear relationships between independent and dependent variables as well as variable interactions (Python v3.8.5, scikit-learn library v0.23.2) (Geurts et al., 2006; Hastie & Tibshirani, 2009; Pedregosa et al., 2011). Regression performance—the prediction accuracy—was assessed using a nested cross-validation (CV) procedure (Cawley, 2010). CV implements repeated train-test splits of the data to assess the generalizability of a model to new unseen data. A shuffle-split

scheme with 128 repetitions (20% of the participants in the testing-set, 80% in the training-set) was applied in the main (outer) CV loop. In each repetition, the training-set was used for data scaling (min-max scaling) and model complexity optimization. Model complexity optimization was carried out in a nested (inner) CV procedure using a sequential Bayesian optimization procedure in combination with a shuffle-split scheme (20% testing, 80% training, 128 repetition, 64 initial points) to find the best performing complexity parameters (BayesSearchCV, scikit-optimize, v0.8.1, min_samples_split 2 to n_samples, min_samples_leaf 1 to n_samples/2, max_features fraction 0 to 1). The complexity parameters that led to the highest prediction accuracy in the training-set were subsequently used along with the constant parameters n_estimators = 333, criterion = friedman_mse and all other left to default, to train ET regressor models in the main CV loop. The models were subsequently tested on the respective testing-set of the main CV loop. The testing-set was explicitly not used in the inner CV loop. Regression performance was measured with (1) the prediction coefficient of determination (prediction R^2) (Hastie 2009). R^2 values are scaled that 0 equals a performance as good as using the average target value as predictor (the trivial predictor), and that 1 means no error at all. However, R^2 values can have values between minus infinity (a model that performs worse than using the mean target value as the prediction) and 1 (perfect model, no error at all). Notably, the prediction R^2 will be smaller than R^2 values of conventional statistical models because the prediction R^2 measures actual predictions for unknown data. Therefore, the prediction R^2 measures existing structures in the data and not post hoc model fit.

Model analysis was carried out with SHAP. SHAP is a method from interpretable machine learning and is based on Shapley values – a method from cooperative game theory – and measures the contributions of each independent variable (here 15 art dimensions and mother tongue of the participants) for a models' prediction. Pooling those contributions over many predictions allows a comprehensive analysis of the importance of single independent variables for the regression task (Lundberg & Lee, 2017).

Statistical significance of the R^2 values as well as of feature importance was determined using a modified t-test that takes the sample dependence due to the CV into account (Nadeau 2003). T-test results were Bonferroni corrected for multiple comparisons. The result of the t-test for Step 1 is shown in Table 16. The average R^2 value of the predictive model was 0.22, meaning that the model could explain 22% of the variance in the beauty rating scores. The average R^2 value of 0.22 is statistically significantly higher than the average R^2 value obtained with shuffle data models ($p < .0001$).

Figure 15 presents the contributions of the 18 independent variables to the beauty ratings. The total impact of each independent variable was measured via the mean of the absolute SHAP values. Importantly, the independent variable: Japanese0_German1 representing the mother tongue of the art viewers is a significantly important independent variable for the prediction of beauty rating scores when compared with SHAP results from shuffle data models. Accordingly, in the following analysis, we separate the data of native Japanese and German speakers (Step 2) and aim to see how the important variables for predicting beauty ratings scores differ between the two groups (Step 3).

Table 16. The results of model predictions in distinct phases.

Models	R^2 original mean (SD)	R^2 shuffle mean (SD)	p value for the R^2 shuffle mean
Step 1	0.22 (0.03)	-0.02 (0.02)	$p < .0001$
Step 2.1 Japanese	0.22 (0.04)	-0.03 (0.03)	$p < .0001$
Step 2.2 German	0.21 (0.04)	-0.02 (0.02)	$p < .0001$

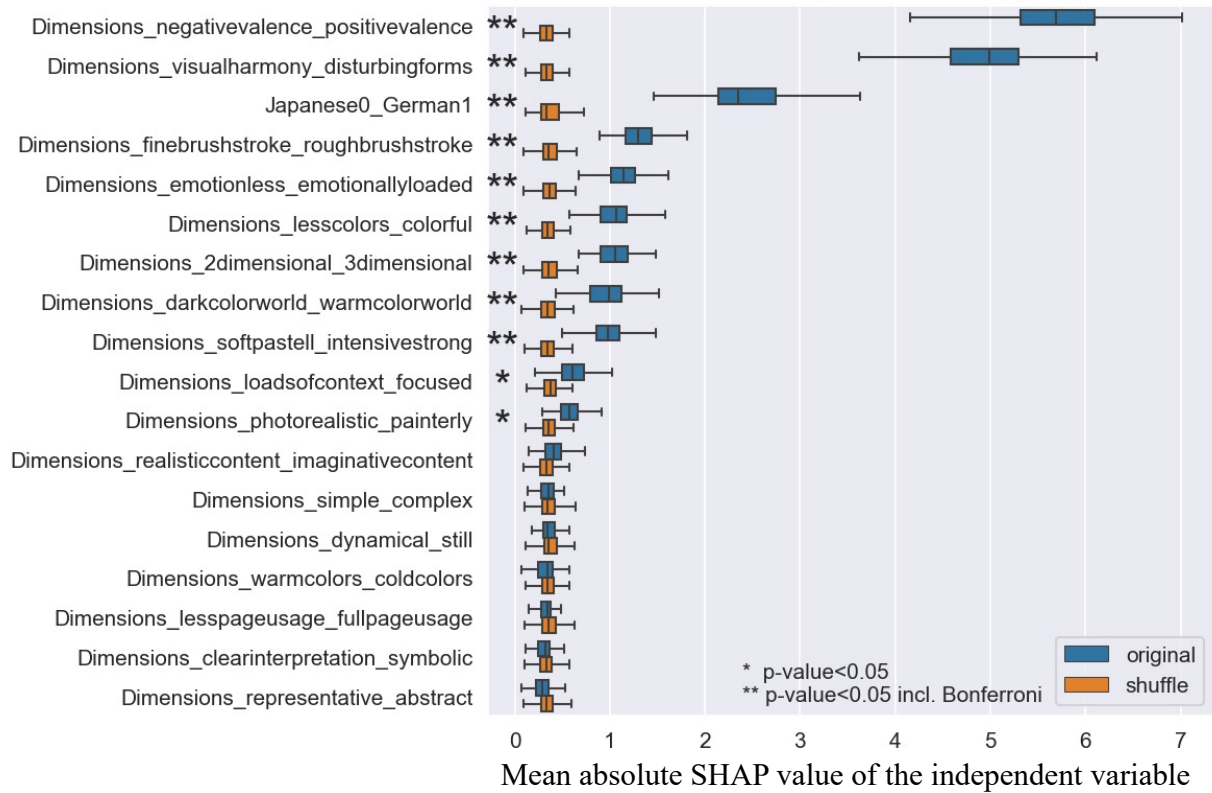


Figure 15. The mean absolute SHAP values for all 18 independent variables in the model in Step 1, including 17 art attributes and the native language of the participants.

Step 2.1 Assessing what art dimensions are significantly contributing for the prediction of the beauty ratings made by native Japanese speakers

Second, we examined what are the important variables for predicting beauty rating scores for native Japanese speakers. We trained a predictive model for the beauty rating scores only using the data from native Japanese speakers. However, note that the independent variable representing the mother tongue of the art viewers was excluded in the trained model, since only the data from native Japanese speakers were included in the model in this step. Thus, only the 17 art attributes were included as the independent variables in step 2.1. Note that all procedures for building a

predictive regression model for the beauty rating scores as well as evaluating the model was identical to the Step 1.

The result of the predictive model in Step 2.1 is shown in Table 16. The average R^2 value of the predictive model was 0.22, meaning that the model could explain 22% of the variance in the beauty rating scores. Further, Figure 16 presents the contributions of the 17 independent variables to the beauty judgements made by native Japanese participants. As in Step 1, the impact of each independent variable was measure via the mean absolute value of SHAP values. In Figure 16, the independent variables are listed by their importance for predicting beauty ratings. For example, *Dimension_visualharmony_disturbingforms* was the most important feature to predict beauty (it was most influential), and the second most important feature was *Dimension_negativevalence_positivevalence*.

The results showed that nine independent variables were significantly important to predict beauty rating scores from native Japanese speakers, namely, *Dimension_visualharmony_disturbingforms*, *Dimension_negativevalence_positivevalence*, *Dimension_finebrushstroke_roughbrushstroke*, *Dimension_softpastell_intensivestrong*, *Dimension_darkcolorworld_warmcolorworld*, *Dimension_simple_complex*, *Dimension_2dimensional_3dimensional*, *Dimension_lesscolors_colorful*, *Dimension_realisticcontent_imaginativecontent*. Note that we only present the significant results with Bonferroni correction here to discuss the results. The directions of the results are shown in Table 17.

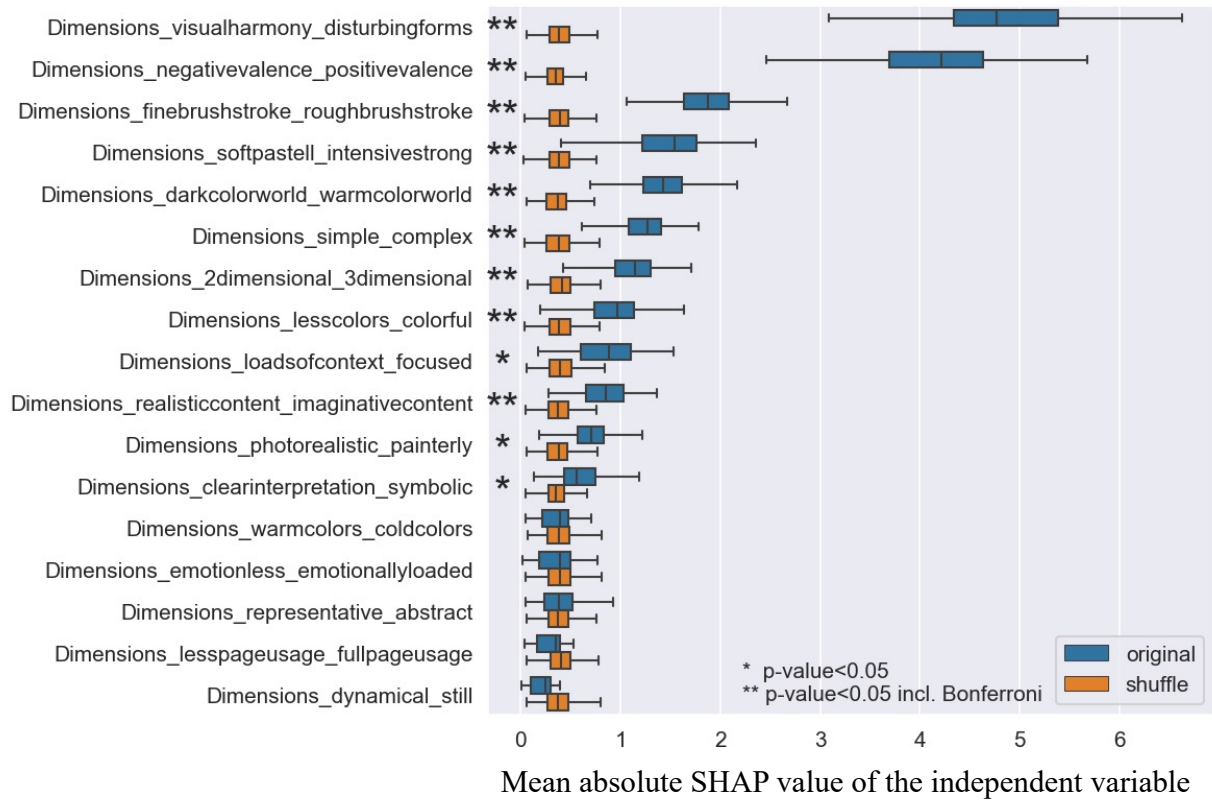


Figure 16. The mean absolute SHAP values for all 17 art attributes in the model in Step 2.1 using only the data from native Japanese speakers.

Table 17. The directions of the results in the model in Step 2.1

Significant Independent Variable	The direction of the results
Dimensions_visualharmony_disturbingforms	Native-Japanese speakers found paintings with high visual harmony as more beautiful, and with disturbing form as less beautiful.
Dimensions_negativevalence_positivevalence	Native-Japanese speakers found paintings with positive valence as more beautiful, and with negative valence as less beautiful.
Dimensions_finebrushstroke_roughbrushstroke	Native-Japanese speakers found paintings with fine brush stroke as more beautiful, and with rough brush stroke as less beautiful.
Dimensions_softpastell_intensivestrong	Native-Japanese speakers found paintings with soft pastel color as more beautiful, and with intensive strong color as less beautiful.
Dimensions_darkcolorworld_warmcolorworld	Native-Japanese speakers found paintings with warm color world as more beautiful, and with dark color world as less beautiful.
Dimensions_simple_complex	Native-Japanese speakers found simple paintings as more beautiful, and complex ones as less beautiful.
Dimensions_2dimensional_3dimensional	Native-Japanese speakers found 3 dimensional paintings as more beautiful, and 2 dimensional ones as less beautiful.
Dimensions_lesscolor_colorful	Native-Japanese speakers found colorful paintings as more beautiful, and paintings with less colors as less beautiful.
Dimensions_realisticcontent_imaginativecontent	Native-Japanese speakers found realistic paintings as more beautiful, and imaginary ones as less beautiful.

Step 2.2 Assessing what art dimensions are significantly contributing for the prediction of the beauty ratings made by native German speakers

Third, we examined the important variables for predicting beauty rating scores for native German speakers. We trained a predictive model for the beauty rating scores only using the data from native-German speakers. As in Step 2.1, the independent variable i.e., Japanese0_German1 was excluded in the trained model, and only the 17 art dimensions were included as the independent variables in Step 2.2. Again, all procedures for building a predictive regression model for the beauty rating scores as well as evaluating the model was identical to the Step 1.

The result of the predictive model in step 2.2 is shown in Table 16. The average R^2 value of the predictive model was 0.21, meaning that the model could explain 21% of the variance in the beauty rating scores. Further, Figure 17 presents the contributions of the 17 independent variables to the beauty judgements made by native-German participants. As in step 1 and 2.1, the impact of each independent variable was measure via the mean absolute value of SHAP.

The results showed that five independent variables were statistically significantly important to predict beauty rating scores, namely, *Dimension_negativevalence_positivevalence*, *Dimension_visualharmony_disturbingforms*, *Dimension_emotionless_emotionallyloaded*, *Dimension_simple_complex*, *Dimension_lesscolors_colorful*. Again, we only present the significant results with Bonferroni correction here to discuss the results. The directions of the results are shown in Table 18.

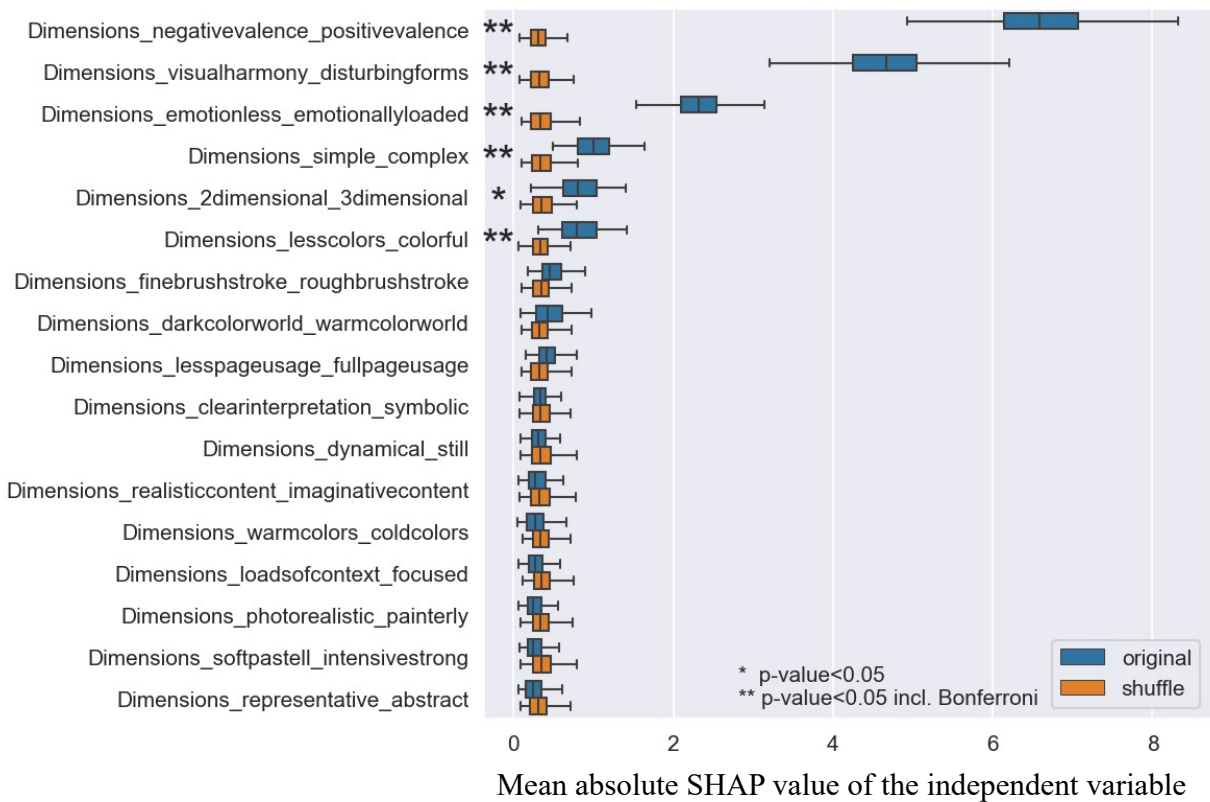


Figure 17. The mean absolute SHAP values for all 17 art attributes in the model in Step 2.2 using only the data from native German speakers.

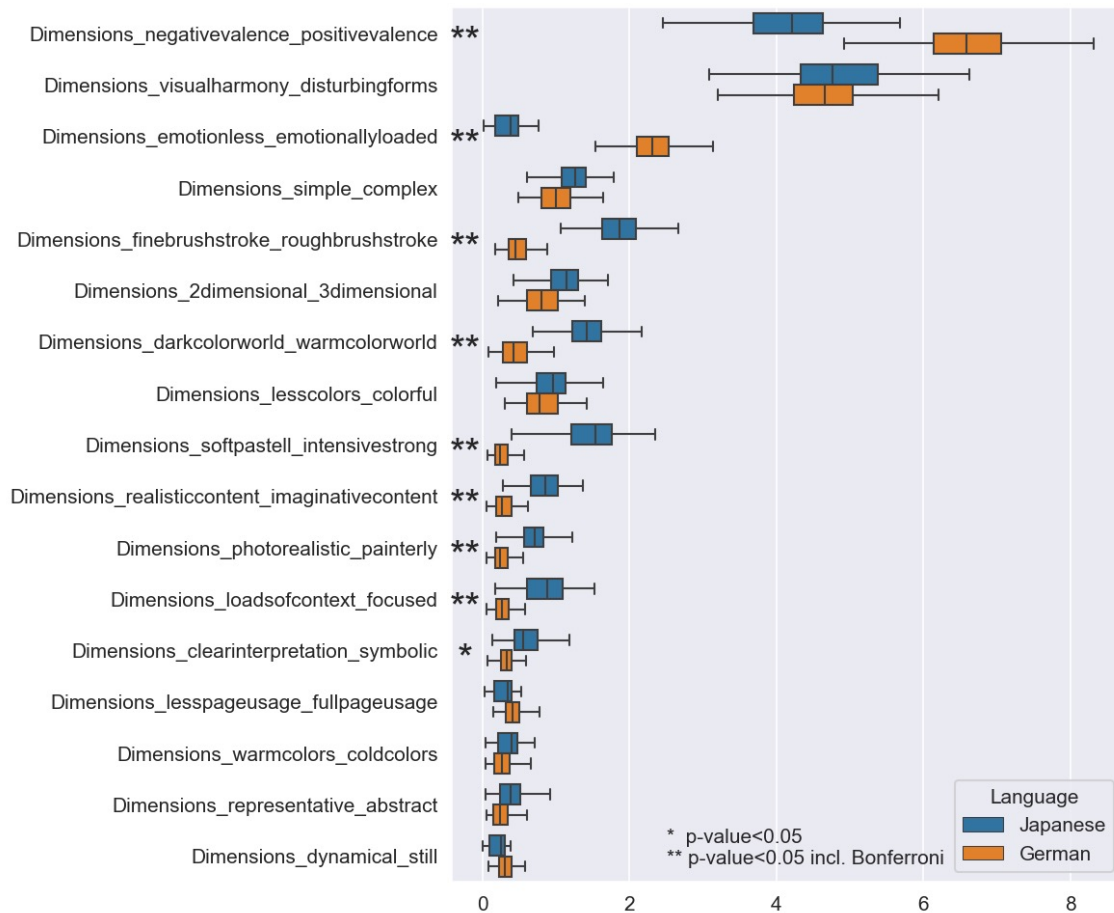
Table 18. The directions of the results in the model in Step 2.2

Significant Independent Variable	The direction of the results
Dimensions_negativevalence_positivevalence	Native-German speakers found paintings with positive valence as more beautiful, and with negative valence as less beautiful.
Dimensions_visualharmony_disturbingforms	Native-German speakers found paintings with visual harmony as more beautiful, and with disturbing forms as less beautiful.
Dimensions_emotionless_emotionloaded	Native-German speakers found emotion loaded paintings as more beautiful, and emotion less ones as less beautiful.
Dimensions_simple_complex	Native-German speakers found complex paintings as more beautiful, and simple ones as less beautiful.
Dimensions_lesscolor_colorful	Native-German speakers found colorful paintings as more beautiful, and paintings with less colors as less beautiful.

Step 3. Assessing if there are differences in the art attributes which contribute to predict beauty rating between native Japanese and German speakers

Last, we compared the two models build in Step 2.1 and 2.2, and tested if the contribution to predict beauty of each independent variable significantly differed between the two groups. The statistical comparison of the feature importance was carried out using a modified *t*-test that takes the sample dependence due to the CV into account (Nadeau 2003). T-test results where Bonferroni corrected for multiple comparisons. Figure 18 presents differences in the contributions of the 17 independent variables to the beauty judgements between native Japanese (Step 2.1) and German (Step 2.2) speakers.

The results showed that there were significant differences between two groups in the importance of eight independent variables, namely, *Dimension_negativevalence_positivevalence*, *Dimension_emotionless_emotionallyloaded*, *Dimension_finebrushstroke_roughbrushstroke*, *Dimension_darkcolorworld_warmcolorworld*, *Dimension_softpastell_intensivestrong*, *Dimension_realisticcontent_imaginativecontent*, *Dimensions_photorealisticPainterly*, *Dimensions_loadofcontext_focus*. Again, we only present the significant results with Bonferroni correction here to discuss the results. How the results were differed for the eight variables are described in Table 19.



Mean absolute SHAP value of the independent variable

Figure 18. The results of the differences in the contributions of the 17 independent variables to the beauty judgements between native Japanese (Step 2.1) and German (Step 2.2) speakers.

Table 19. The directions of the cross-cultural differences

Variables	Results
Dimensions_negativevalence_positivevalence	This feature was important for both cultures, but more important for German speaking people: same direction, always the positive was appraised as more beautiful.
Dimension_emotionless_emotionallyloaded	While this feature was important for German, it was not for Japanese people: German evaluated more emotionally loaded paintings as more beautiful.
Dimension_finebrushstroke_roughbrushstroke	While this feature was important for Japanese people, it was not for German people: Japanese evaluate more fine strokes as more beautiful.
Dimension_darkcolorworld_warmcolorworld	While this feature was important for Japanese people, it was not for German people: Japanese evaluate warm color as more beautiful.
Dimension_softpastell_intensivestrong	While this feature was important for Japanese people, it was not for German people: Japanese evaluate pastel color as more beautiful.
Dimension_realisticcontent_imaginativecontent	While this feature was important for Japanese people, it was not for German people: Japanese evaluate realistic painting as more beautiful.
Dimensions_photorealisticPainterly	While this feature was important for Japanese people, it was not for German people: Japanese evaluate photorealistic painting as more beautiful.
Dimensions_loadofcontext_focus	While this feature was important for Japanese people, it was not for German people: Japanese evaluate paintings with loads of contents as more beautiful.

4.2.3. Discussion

In Study 3-2, we aimed to quantitatively assess (1) if there is significant effect of culture on beauty appraisals for artworks, and (2) if there are inter-cultural differences in significant art attributes for predicting beauty. In the present study, 17 most influential art attributes for rating artworks were used as the predictors for beauty appraisal. To achieve to build statistical models with large sets of art attributes, we used machine learning method.

The results from the Step 1 showed that the mother tongue of the art viewers was a significantly important variable to predict beauty appraisal, and thereby, the answer for our first question is “yes, culture does have a significant impact on predicting beauty appraisal”. Thus, our result adds another insight that culture has significant impact not only on the preference/liking (Che et al., 2018), but also on beauty appraisal towards artworks. Regarding the inter-cultural differences in art attributes predicting beauty, there were different patterns e.g., there are some art attributes

which are important to predict beauty appraisal regardless of culture, which are important to either one of the cultures to predict beauty but not to the other culture etc. In the following section, we discuss the results dividing into the different patterns.

Art attributes which are important to predict beauty regardless of the cultures

In total, there were four art attributes which were important to predict beauty regardless of the cultures of participants, specifically, *visual harmony*, *color variety*, *valence*, and *complexity*. Regarding the first three attributes, participants from both cultures evaluated paintings with higher visual harmony/with more color/with positive valence as more beautiful and one with disturbing forms/with less color/with negative valence as less beautiful. However, note that the results of Step 3 revealed that though the attribute *valence* was an important factor to predict beauty for both groups, the importance of this art attribute was significantly larger for the German speaking group. Thus, though this attribute was important to predict beauty in both cultures, the importance was difference across the cultures.

The fact that visual harmony and the positive valence were the important factors to predict beauty in both cultures seem to be compatible with the discussion regarding beauty from the ancient Greek to more recent theories of aesthetics. Plato and Aristotle, in their arguments about aesthetics, both agreed that visual harmony is one of the important components in aesthetics (see Palmer et al., 2013), and beauty has been always connected with something pleasurable, positive, and desired (e.g., Leder et al., 2004). Thus, our results supported these arguments and its universality.

The attribute *complexity* was equally importance for predicting beauty across the two cultures. However, the Japanese and German speakers evaluated the opposite pole as more beautiful. Specifically, while Japanese speakers evaluated simpler paintings as more beautiful, German speakers evaluated more complex paintings as more beautiful. Hence, here, we see an inter-cultural difference. This result is especially interesting, since the past studies in empirical aesthetics repeatedly showed that people tend to prefer/value complexity in artworks and music (e.g., Berlyne, 1973, Chmiel & Schubert, 2017 for a review; Marin et al., 2016). Berlyne's theory (Berlyne, 1973) which emphasizes the importance of the mediating role of arousal, explains the preference/values for complexity by arguing that the increase in the complexity might evoke

arousal, resulting in the preference for complexity up to an optimal point. Our results from German native speakers are thus compatible with the past findings, where *Western* participants were the majority of the participants. However, native Japanese participants showed the opposite result. Perhaps, this result might provide the evidence that aesthetic experience/judgement is not completely shared universally by nature, but nurture can also play an important role. Specifically, if the Berlyne's arousal theory holds, the Japanese participants would have also evaluated more complex paintings as more beautiful, since more complex paintings might require more detailed information processing (e.g., a number of colors, interpretation of the meanings in the painting) which might provoke higher arousal regardless of cultures. Nevertheless, the native Japanese speakers evaluated simpler paintings as more beautiful. Here, though speculative, the reason why the Japanese participants valued more simple paintings as more beautiful could be found in Japanese aesthetics. Especially in one of the Zen principles, simplicity/quietness are important and admired concepts. As can be seen in terms in Zen principles such as *Kisha* (喜捨), meaning discard with pleasure, being/living simple mentally/physically is something to be appreciated. Further, the higher simplicity could be also found in the forms of artworks. In Japanese as well as Chinese landscapes, it is quite famous that they utilize empty spaces quite often, which might make the space seem as simpler. Further, in Japanese traditional gardens, simplicity as well as seen/experienced quietness are intentionally implemented in the design. Though it is unclear whether the higher visual familiarity of simpler object or the concept in the Japanese culture resulted in the opposite results in the present study e.g., simple is something to be appreciated, indeed, this result implies that nurtured experience could have impacts on aesthetics in paintings.

Inter-cultural differences in art attributes which are important to predict beauty

The results of Step 3 showed that seven attributes were only important either one of the cultures but not for the other one. For German speakers, *emotional expressiveness* was an important attribute to predict beauty, but not for Japanese speakers. In contrast, for Japanese speakers, *brushwork*, *color world*, *saturation*, *realism* were important attributes to predict beauty, but not for German speakers. Additionally, though the results of the Step 2.1 did not show the significant

impacts to predict beauty of *accuracy of depicting objects* and *focus*, the importance of the two attributes were significantly larger for the Japanese speakers than German speakers.

These results might even foster the argument that nurtured experience might have significant impact of aesthetics. For example, the attribute of brushstroke can be seen as a representative one, which could reflect the differences in everyday practice in the two cultures. In Japan, especially when people write Kanjis, brush stroke is a core feature to see if one person can write beautifully/well or not. Similarly, Shodo (書道) or Shuuiji (習字), which is a form of calligraphy, is one of the popular forms of art in Japan, and they are also in the curriculum in Junior/High school. In Shodo and Shuuiji, we use ink and tick brush, and write Japanese words/Kanjis on thin Japanese traditional paper. Obviously, brush stroke is a critical component to make Shodo/Shuuiji as beautiful, and people learn how to write properly and beautifully with tick brushes. On the New Year's Day, it is a very traditional practice to make this calligraphy with family members, stating aspirations for the coming year. Hence, considering that the importance/evaluation for brush strokes are quite commonly implemented in everyday lives for the Japanese people, it seems plausible that this attribute was one of the important attributes to predict beauty for only Japanese speakers, but not for German speakers.

Going further, the attribute *emotional expressiveness* had a significant impact only to predict beauty for native German speakers but not for Japanese speakers. Also, as presented in the last paragraph, the importance of the attribute *arousal* was significantly larger for the native German speakers. Combining those two results together, it seems like all emotional-related attributes were more important to predict beauty for German speakers than Japanese speakers. This result is somehow along with the cliché in the two cultures. Of course, there are inter-individual differences within culture, people in Europe are generally described as more expressive, while Japanese people are as not expressive but rather quiet. From my personal experiences, for example, the form of university class represents these differences very well. Specifically, when I was taking a class in the University of Vienna, it was common that students stopped the teacher and started discussing in the class. Hence, not only to express one's emotion, but also one's thought/idea seemed to be something to be shared and appreciated. However, in Keio University and in most of the universities in Japan to the best of my knowledge, it is not so common for students to stop the

teachers in the classes, and they were expected to be quiet during the class, and to ask questions after the classes, and thus, discussions were not spontaneously happening. Hence, the differences in the emotion-related attributes across the two cultures are consistent with the practices/norms shared by the society in each culture. The impact of the practices/social norms could also be found in the differences in the attributes, *realism* and *accuracy of depicting objects* and *focus*. Our results suggested that the Japanese native speakers evaluated realistic contents/photorealistic paintings as more beautiful. For example, Kim and Sherman (2008) reported that, while North American people show their preference to something unique e.g., an uncommon pen, East Asians prefer something representing conformity e.g., a common pen. Ishii et al. (2013) found similar tendency in colorings of patterns. They reported that while colorings made by North American/European adults and children were evaluated as more unique, those made by Japanese adults and children were evaluated as more harmonious. Further, the participants preferred their own-culture colorings more than different-culture ones. Further, Ishii et al. (2013) discussed that “child rearers in Japan and Canada gave feedback about the children’s colorings that were consistent with their culture’s value. These findings suggest that simple geometric patterns can embody cultural values that are socialized and reinforced from an early age” (p. 175). The results in the present study, the native Japanese speakers tended to evaluate realistic paintings as more beautiful, are in line with the research above, suggesting that reinforcement cycle in cultures might shape aesthetic values. Finally, the differences in the art attribute *focus* is along with the theories and findings from Nisbett and his colleagues (Chua et al., 2005; Nisbett et al., 2001; Masuda & Nisbett, 2001): while *East Asian* people tend to focus on more contextual information in the visual field, *Western* people tend to focus more on focal objects. In our study, the results showed that Japanese speakers evaluated paintings with load of context/environment in the image as more beautiful, showing that the differences in the strategy of visual exploration are also reflected in the beauty appraisals towards artworks.

To conclude, our results showed that art attributes constituting beauty appraisal were not completely shared universally, showing that culture might impact on art experience, evaluation, and/or art processing itself. Further, it seems like the practices/values share in a culture could

contribute to the differences. In the next section, Study 3-3, we go one more step further, and test the impact of reading/writing system in culture on art experience.

4.3. Study 3-3: the cross-cultural diversity in eye movement patterns while viewing art between two cultures

Manuscript submitted to Journal of Cross-Cultural Psychology

Brinkmann, H., **Mikuni, J.**, Dare, Z., Kawabata, H., Leder, H., & Rosenberg, R. (*submitted*).

Cultural Diversity for Austrian and Japanese Viewers in Saccade Direction and Amplitude when Viewing Art and Non-Art. Submitted date: 30.06.2021.

My Contribution:

I shared the first authorship with HB. I collected the data and analyzed and interpreted the data together with HB and RR. HB, HL, RR and I were the main writers of the current draft, elaborating the draft with ZD and HK.

In past studies in the field of empirical aesthetics as well as in the field of art history, scholars have assumed that the way people interact with artworks e.g., how people look around artworks or the aspects in the artworks which people would focus, might diverse across cultures. For example, Thausing (1873) argued that the history of (stylistic) changes in art derives from a history of changes in perception. Later, that idea figured in several prominent theories of art history, as for example, Baxandall's concept of the *period eye* (Baxandall, 1988). He assumed that people develop different cognitive styles due to education and a lifelong exposure to different visual cultures. Thus, culture would affect the way people perceive and process art due to an unconscious collective visual memory based on shared experiences in everyday life. Consequently, people from different cultures would perceive art differently.

Those assumptions were further supported by the work from Nisbett and his colleagues. They postulated “that there is a causal chain running from social structure to social practice to attention and perception to cognition” (Nisbett et al., 2001). They suggested that there are basic differences of social structures between individualism in *Western* and collectivism in *Eastern* culture, resulting in different systems of thought—analytic in the Western vs. holistic in the Eastern cultures (Nisbett, 2003). Later on, researchers tested this hypothesis by measuring eye movements on photographic scenes (Boland et al., 2008; Chua et al., 2005) and found that Western participants spent more time focusing on focal objects, while East Asians distributed their fixations more evenly between foreground and background. The authors interpreted the latter as support of holistic perception and concluded that culture does influence the way we see. While not all studies could reproduce their findings (Rayner et al., 2009), a number of other studies also reported cultural influence on perception. Buswell noted differences in placement of fixations between American, Chinese, and Japanese viewers as early as 1935 (Buswell, 1935).

More recently, Liu et al. (2013) found Chinese viewers focus more on the empty spaces in Chinese ink paintings than Western viewers. Kliegl et al. (2015) found that Western viewers found Mount Fuji in Hokusai's famous print *Great Wave off Kanagawa* quicker when it was shown to them

horizontally flipped, attributing the difference to reading direction. Further, Brinkmann (2021) suggested that saccades differ between cultural groups and that the direction of script might have an influence on their direction. Following these findings and arguments, in the present study, we also postulate that oculometric patterns such as the direction and amplitude of saccades, the duration of fixations, and dynamic changes across the viewing time, which are more often considered in studies of reading (Rayner et al., 2009), might be relevant in studies of art perception.

Oculometric parameters of persons viewing images could also differ depending on whether they perceive, and or expect to perceive, art vs. non-art. This is suggested by two studies that tested aesthetic vs. non-aesthetic tasks: Molnar (1981) asked students from an art school to look at paintings, either in order to answer questions about their aesthetic qualities or their content. Participants in the aesthetic group made significantly longer fixations than those in the semantic group. Cronin et al. (2020) compared participants looking at non-artistic photographs either in order to assess their aesthetic qualities or to memorize them. They found effects of the task on fixation duration and saccade amplitudes and their changes over time.

In the present study, in order to explore cultural effects on oculometric parameters depending on the script systems, again, the eye movement patterns between the participants from Japan and Austria were compared, while they saw a series of artworks. Japanese and Austrian visual culture are historically different in several respects. Two fundamental differences lay in their ways of representing space in images (perspective) and their writing systems (Brinkmann, 2021). Since the Renaissance, Europeans are used to pictures constructed with central perspective, whereas central perspective was unusual in Japan before the Meiji period (1868 -1912) and is still not the norm today. Austrians employ the German variant of the Latin script, which, just like English, is largely phonemic and uses the Latin alphabet with 26 characters and four special letters (i.e., ä, ö, ü, ß). They are grouped in distinct words, written in lines from left to right, with lines ordered from top to bottom. In contrast, Japanese does not use word boundaries and have a lot more characters. The Japanese writing system is far more complex than the Austrian one, with three

different alphabets (*Hiragana, Katakana, Kanji*). A significant difference lies also in the direction of script. Traditionally, Japanese is written in columns from top to bottom, with columns ordered from right to left (*tategaki*), which is still commonly used, for example in newspapers/text books/novels. At the same time, Japanese books and internet contents are also commonly written in lines from left to right (*yokogaki*), as in English and German.

Consequently, Study 3-3 compared the eye movement parameter of Japanese and Austrian participants, varying the type of image, and for the first time also analyzing saccades—their direction and amplitude—as well as the dynamic changes of oculometric parameters over time. In accordance with other studies that assume that reading direction influences perception (Kazandjian & Chokron, 2008; Treiman & Allaith, 2013) we hypothesized that while looking at images 1. the Japanese participants will show a higher amount of vertical and a smaller amount of horizontal saccades compared to the Austrian ones and 2. vertical saccades of the Japanese participants will have a larger amplitude than in the Austrian ones, whereas horizontal saccades of the Austrians participants will have a larger amplitude than in the Japanese ones. Beyond these hypotheses, as in Cronin et al. (2020), Study 3-3 aimed to explore any differences that might occur during the time of viewing—cultural differences between participant groups, and or those resulting from the types of images.

4.3.1. Method

Participants

We recruited 30 Austrian and 30 Japanese participants between 18 and 30 years ($M_{age} = 22.00$ for the Austrian and 20.00 for the Japanese participants) among the students from Keio University (Tokyo, Japan) and the University of Vienna (Vienna, Austria). Students with an art historical background were explicitly excluded. Of the 30 Austrian participants, 20 were women and of the 10 Japanese participants, 13 were women. The experiment was carried out in accordance

with the Declaration of Helsinki and approved by the local Ethic committee of the University of Vienna. All participants had normal or corrected to normal sight. Participants were rewarded with 5 Euro / 1000 Japanese Yen.

Stimuli and Apparatus

Forty images were used as the stimuli. The stimulus set can be divided into four blocks: the everyday photographs block included 5 photos taken in Japan and 5 in Europe, the Japanese artworks block included 10 Japanese pictures from different periods and styles, the European artworks block included 10 European paintings from different periods and styles, and the “art” photographs block included another 5 and 5 photographs taken in Japan and Europe similar to the non-artistic photographs.

The experiment was controlled via Experiment Builder (Version 1.10.1630), and the eye movement patterns were measured via EyeLink 1000 remote eye tracker (SR research, Canada), recording a single (dominant) eye at 1000Hz. Images were presented on a monitor (24-inch, 16:10 format) with a 1920x1200 resolution. Events were detected using Eye Link’s proprietary velocity and acceleration based event detection algorithm, following a comparison of four alternatives. EyeTrace (Kübler et al., 2015) was used for classifying saccade directions.

Procedure

After being welcomed in the lab, participants signed a consent form and took a seat in front of an EyeLink 1000 remote eye tracker (SR research, Canada). For each participant, the height of the chin rest as well as the chair for each participant were adjusted in order to keep the distance from the screen monitor from the participants constantly at about 80cm, followed by the calibrations of eye fixations. The calibrations were conducted prior to the main part of the experiment as well as to each stimulus block. For the calibration, the calibration protocol, EyeLink

API, was used which is implemented in EyeLink 1000. After the calibrations, the fixations of the participants were validated with the EyeLink software, and repeated until they reached the validation criteria in the software. When a calibration trial started, four fixation crosses appeared on the four quadrants of the screen, and the participants were instructed to fixate each cross to perform a drift correction. If the drift correction was more than 0.5° , a new calibration was conducted to assure an optimal tracking accuracy. The participants repeated the calibration phase, until the drift correction was under 0.5° .

After the calibration phase was successfully completed, participants started the main part of the experiment. Before the beginning of each stimulus block started, an introduction was presented explaining what types of images (i.e., everyday photos, Japanese artworks, European artworks, ‘artistic’ photos) they are going to view in the next block.

When one block started, a fixation cross was presented on the screen center, until the participants kept making their fixations for more than 500 ms, following an image presentation screen. One image was presented for the fixed duration (i.e., 12 s), and the participants could explore the screen freely while the image was presented. After all of the images included in one block was presented (i.e., 10 images in each stimulus block), the screen went back to the calibration screen again. Before participants started a new stimulus block, the same calibration procedure was again completed, and they proceeded to the next stimulus block. Participants repeated this flow four times in total, and viewed all images used in the present study. The order of the stimulus block was counter-balanced across participants, and the presentation order of the images within one stimulus block was randomized.

4.3.2. Results

Proportion of saccades according to direction

In order to assess the cross-cultural differences in the proportion of saccade directions, first, all the saccade data was divided into four directions each covering 90° (up, down, right, left). Second, in order to estimate the statistical relevance of the proportion of different saccade directions, we ran GLMMs with a binominal distribution. In the model, all vertical saccades were coded as 0, and all horizontal saccades were coded as 1. The proportion of horizontal/vertical saccades (dependent variable) was predicted by a treatment coded contrast comparing the participants' culture with Austrian (baseline) against Japanese. The model also contained random intercepts for participants and images in order to evaluate the effect of fixed effects while modeling the natural individual differences for saccade directions and for specific art styles/artworks within the analysis. We ran the same structured model separately for the four distinct stimulus blocks. Fixed effects and the results of the GLMMs are presented in the upper part of Table 20. The results of all GLMMs showed significant intercept for all stimulus types, confirming the hypothesis that the Austrian participants made more horizontal than vertical saccades. The estimated coefficients and the range of the confidence intervals for the coefficients were similar across the four stimuli blocks, suggesting that the Austrian participants made similar proportions of horizontal and vertical saccades in all stimuli blocks. However, we found significant main effects of culture in regard to artworks (Japanese and European): while looking at artworks, the Japanese participants still made more horizontal than vertical saccades—the coefficients are bigger than 0—but significantly more vertical saccades than the Austrian group. In contrast, we did not find a main effect of culture in

either of the photograph blocks, indicating that there was no significant difference in the amounts of horizontal/vertical saccades between the Japanese and Austrian participants.

Table 20. Fixed effects in the GLMMs predicting proportion of saccades

	Estimate	95% CI		z value	Pr(> t)
		Lower	Upper		
<i>Proportion of vertical and horizontal saccades</i>					
non-artistic photographs					
(Intercept)	0.69	0.55	0.84	9.430	< .001
culture - Japan	-0.073	-0.27	0.12	-0.760	0.45
Japanese artworks					
(Intercept)	0.73	0.59	0.87	10.26	< .001
culture - Japan	-0.34	-0.52	-0.17	-3.83	< .001
European artworks					
(Intercept)	0.69	0.56	0.83	10.08	< .001
culture - Japan	-0.43	-0.59	-0.27	-5.25	< .001
‘artistic’ photographs					
(Intercept)	0.71	0.55	0.86	8.96	< .001
culture - Japan	-0.18	-0.38	0.02	-1.78	0.08
<i>Proportion of left and right saccades</i>					
non-artistic photographs					
(Intercept)	-0.05	-0.11	0.01	-1.57	0.12
culture - Japan	0.00	-0.09	0.08	-0.03	0.98
Japanese artworks					
(Intercept)	-0.04	-0.09	0.01	-1.44	0.15
culture - Japan	-0.01	-0.09	0.07	-0.20	0.84
European artworks					
(Intercept)	-0.01	-0.06	0.05	-0.21	0.83
culture - Japan	-0.02	-0.10	0.06	-0.47	0.64
‘artistic’ photographs					
(Intercept)	0.00	0.00	0.00	0.00	1.00
culture - Japan	-0.02	-0.09	0.06	-0.40	0.69
<i>Proportion of down and up saccades</i>					
non-artistic photographs					
(Intercept)	0.24	0.15	0.33	5.21	< .001
culture - Japan	-0.27	-0.40	-0.14	-4.07	< .001

Japanese artworks					
(Intercept)	0.22	0.13	0.31	4.71	< .001
culture - Japan	-0.19	-0.31	-0.06	-2.94	< .001
European artworks					
(Intercept)	0.24	0.16	0.31	6.54	< .001
culture - Japan	-0.21	-0.30	-0.11	-4.30	< .001
‘artistic’ photographs					
(Intercept)	0.26	0.17	0.34	6.25	< .001
culture - Japan	-0.23	-0.34	-0.12	-4.24	< .001

Note: The top part of the table shows the fixed effects and the results of the GLMMs predicting the proportion of horizontal (including right and left saccades, coded as 1) and vertical (including up and down saccades, coded as 0) saccades. In the middle, the results of the GLMMs predicting the proportion of right (coded as 1) and left (coded as 0) saccades are shown. The bottom shows the results of the GLMMs predicting the proportion of up (coded as 1) and down (coded as 0) saccades. In all of the models, binominal distribution with logit link function was used. If the coefficients are 0, the participants made equal portion of horizontal/vertical, right/left, and up/down saccades respectively. As one direction was coded as 1 (or 0), if the coefficients have the plus (minus) values, the results suggest that the participants made bigger portion of horizontal (vertical), right (left), and up (down) saccades respectively.

To predict the proportions within horizontal and vertical saccades, and assess cross-cultural differences, we ran the same structured model for horizontal and vertical saccades separately. In the models, all leftward saccades were coded as 0, all rightward saccades were coded as 1 for analyzing the horizontal saccades, and all downward saccades were coded as 0, all upward saccades were coded as 1 for analyzing the vertical saccades. The model structures were identical to the one used for horizontal and vertical saccades. The fixed effect and the results of GLMMs for predicting the horizontal saccades are presented in the middle part of Table 20, and those for predicting the vertical saccades are presented in the lower part of Table 20.

The results for horizontal saccades, showed no significant intercepts and main effects of culture for all of the stimulus blocks indicating that both participant groups made equal proportions of left and right saccades for all stimuli. In contrast, the results for vertical saccades showed significant intercepts and main effects of culture for all of the stimulus blocks, indicating that within vertical saccades Austrian participants made more up compared to down saccades, independently of the stimuli, while Japanese participants made equal proportions of up and down saccades.

Saccade amplitude according to direction

In order to analyze whether the saccade amplitudes differ between the two participant groups, we conducted a series of LMMs. The saccades were again split into four directions, and we ran four LMMs separately—one for each direction as well as for each stimulus block. We set the saccade amplitude for each saccade as the dependent variable. Independent variables and random effects were set as the same manner as for the analysis of saccade directions. Full descriptions of the statistical results are presented in Table 21. The results show significant main effects of culture in both horizontal saccade directions, regardless of the stimulus type: the amplitudes were significantly smaller in the Japanese group compared to the Austrian group. In contrast, we did not find main effect of culture in either of the vertical saccade directions.

Table 21. Fixed effects in the LMMs predicting the saccade amplitudes (°) for each saccade direction and for each stimulus type

	Estimate	95% CI		t value	Pr(> t)
		Lower	Upper		
<i>Saccade Direction - Right</i>					
non-artistic photographs					
(Intercept)	6.28	5.89	6.67	31.69	< .001
culture - Japan	-0.68	-1.22	-0.15	-2.49	0.02
Japanese artworks					
(Intercept)	6.48	6.02	6.94	27.63	< .001
culture - Japan	-1.08	-1.67	-0.49	-3.61	< .001
European artworks					
(Intercept)	6.33	5.89	6.77	28.36	< .001
culture - Japan	-1.24	-1.81	-0.66	-4.22	< .001
‘artistic’ photographs					
(Intercept)	6.48	5.99	6.97	25.99	< .001
culture - Japan	-0.96	-1.62	-0.30	-2.86	0.006
<i>Saccade Direction - Left</i>					
non-artistic photographs					
(Intercept)	6.03	5.66	6.40	31.94	< .001
culture - Japan	-0.54	-1.07	-0.02	-2.03	0.047
Japanese artworks					
(Intercept)	6.17	5.76	6.57	29.86	< .001
culture - Japan	-0.86	-1.39	-0.32	-3.16	< .001
European artworks					
(Intercept)	6.19	5.77	6.61	28.93	< .001
culture - Japan	-1.29	-1.87	-0.71	-4.38	< .001
‘artistic’ photographs					
(Intercept)	6.31	5.88	6.74	28.86	< .001
culture - Japan	-0.87	-1.42	-0.32	-3.09	0.003

Saccade Direction - Up

non-artistic photographs					
(Intercept)	4.30	3.97	4.63	25.50	< .001
culture - Japan	-0.02	-0.46	0.42	-0.10	0.92
Japanese artworks					
(Intercept)	4.40	4.05	4.74	25.08	< .001
culture - Japan	-0.43	-0.89	0.04	-1.78	0.08
European artworks					
(Intercept)	4.27	3.94	4.60	25.40	< .001
culture - Japan	-0.15	-0.56	0.26	-0.72	0.47
‘artistic’ photographs					
(Intercept)	4.31	3.89	4.72	20.17	< .001
culture - Japan	-0.08	-0.63	0.47	-0.28	0.78
<i>Saccade Direction - Down</i>					
non-artistic photographs					
(Intercept)	4.43	4.07	4.80	23.83	< .001
culture - Japan	0.06	-0.38	0.51	0.27	0.79
Japanese artworks					
(Intercept)	4.78	4.39	5.18	23.63	< .001
culture - Japan	-0.38	-0.85	0.09	-1.57	0.12
European artworks					
(Intercept)	4.65	4.28	5.03	24.40	< .001
culture - Japan	-0.31	-0.72	0.10	-1.50	0.14
‘artistic’ photographs					
(Intercept)	4.70	4.25	5.15	20.49	< .001
culture - Japan	-0.28	-0.80	0.23	-1.09	0.28

Note: From the top to the bottom, LMMs predicting the saccade amplitudes for each stimulus type are reported separately for the saccade directions, Right, Left, Top, Down.

Saccade amplitude over time

In order to assess changes in saccade amplitude over time within one trial, independent of their direction, we ran another set of LMMs separately for every stimulus block. We set the saccade amplitude as dependent variable, and the time when the saccade started, treatment coded contrast comparing the participants' culture with Austrian (baseline) against Japanese, and the interaction between two variables as independent variables. In order to ease the complexity of the model, we used centered scaled viewing time and the starting time. For random effects, we set the random slope by participants and the random intercept for images. All fixed effects and the results are presented in Table 22. The result shows that the saccade amplitude was consistently higher in the Austrian group compared to the Japanese group, except for the non-artistic photograph block. In all stimulus blocks, there are significant main effects of start time, suggesting that the saccade amplitude became smaller over time in the Austrian group. This trend was similar for the Japanese participants with one exception: when Japanese people looked at the Japanese artworks, the saccades amplitude did not decrease over time.

Table 22. Fixed effects in the LMMs predicting the saccade amplitudes over time for each stimulus condition

	Estimate	95% CI		t value	Pr(> t)
		Lower	Upper		
non-artistic photographs					
(Intercept)	0.04	-0.02	0.11	1.30	0.20
start time	-0.07	-0.10	-0.04	-4.62	< .001
culture – Japan	-0.09	-0.18	0.00	-1.87	0.07
Interaction between the start time and culture (Japan)	0.03	-0.01	0.07	1.32	0.19
Japanese artworks					
(Intercept)	0.09	0.02	0.16	2.45	0.02
start time	-0.05	-0.07	-0.03	-4.44	< .001
culture – Japan	-0.18	-0.27	-0.08	-3.67	< .001
Interaction between the start time and culture (Japan)	0.04	0.01	0.07	2.60	0.011
Japanese artworks – Japanese participants only					
(Intercept)	0.00	-0.08	0.09	0.09	0.09
start time	0.00	-0.03	0.01	-0.84	0.40
European artworks					
(Intercept)	0.11	0.04	0.19	2.89	0.005
start time	-0.06	-0.08	-0.04	-5.50	< .001
culture – Japan	-0.22	-0.32	-0.12	-4.25	< .001
Interaction between the start time and culture (Japan)	-0.02	-0.01	0.05	1.39	0.17
‘artistic’ photographs					
(Intercept)	0.07	-0.01	0.15	1.76	0.08
start time	-0.06	-0.09	-0.04	-4.46	< .001
culture – Japan	-0.15	-0.25	-0.04	-2.75	0.007
Interaction between the start time and culture (Japan)	0.04	0.00	0.08	1.80	0.08

Note: Considering the individual differences in the slope of saccade amplitude, we set the random slopes for participants as the random effect. Although we tried to set the row data of saccade amplitude and of start time as the independent variables, the model could not converge. Thus, I centered those two variables in the LMMs.

Fixation duration over time

Saccade amplitude over time is a parameter generally discussed in relation to fixation duration. Therefore, we also analyzed the duration of fixations over time for both groups for four stimulus blocks. We run a series of LMMs, setting fixation duration as the dependent variable. The structure of the independent variables and random effects was the same as the analysis of saccade amplitude over time. However, I did not include the random intercept for images since the models did not converge. All fixed effects and the results are presented in Table 23. In all stimulus blocks, there were significant main effects of start time, meaning that the fixation duration became longer over time in the Austrian group. Though a significant interaction between start time and culture was only found for European artworks, it can be seen that fixation durations for the Japanese group show less changes over time compared to the Austrian group. Consequently, we ran the same structure models using only the Japanese participants' data. The results show the fixation duration for the Japanese group only became bigger when they viewed 'artistic' photographs, but did not for the rest of three stimulus blocks.

Table 23. Fixed effects in the LMMs predicting the fixation duration over time for each stimulus condition

	Estimate	95% CI		t value	Pr(> t)
		Lower	Upper		
Impact of time and culture on fixation duration					
non artistic photographs					
(Intercept)	-0.02	-0.09	0.06	-0.44	0.66
start time	0.03	0.01	0.05	2.73	0.009
culture - Japan	0.08	-0.02	0.18	1.49	0.14
Interaction between the start time and culture (Japan)	-0.03	-0.06	0.00	-1.69	0.10
Japanese artworks					
(Intercept)	0.00	-0.06	0.07	-0.05	0.96
start time	0.06	0.02	0.09	3.50	< .001
culture - Japan	0.04	-0.05	0.13	0.85	0.40
Interaction between the start time and culture (Japan)	-0.04	-0.08	0.00	-1.78	0.08
European artworks					
(Intercept)	-0.01	-0.09	0.07	-0.23	0.82
start time	0.04	0.01	0.06	2.93	0.005
culture - Japan	0.07	-0.04	0.18	1.26	0.21
Interaction between the start time and culture (Japan)	-0.04	-0.08	0.00	-2.18	0.03
artistic' photographs					
(Intercept)	-0.01	-0.10	0.07	-0.30	0.77
start time	0.04	0.01	0.06	2.55	0.014
culture - Japan	0.09	-0.03	0.21	1.43	0.16
Interaction between the start time and culture (Japan)	0.00	-0.04	0.04	0.21	0.83
Impact of time and culture on fixation duration: Japan					
non artistic photographs					
(Intercept)	0.03	-0.05	0.11	0.72	0.48
start time	0.00	-0.02	0.03	0.27	0.79
Japanese artworks					
(Intercept)	0.02	-0.05	0.10	0.04	0.50
start time	0.01	-0.02	0.05	0.02	0.42
European artworks					

(Intercept)	0.03	-0.06	0.12	0.71	0.48
start time	0.00	-0.03	0.02	-0.26	0.80
artistic' photographs					
(Intercept)	0.04	-0.06	0.15	0.83	0.42
start time	0.04	0.0	0.08	2.10	0.047

Note: Considering the individual differences in the slope of saccade amplitude, I set the random slopes for participants as the random effect. Although I tried to set the row data of saccade amplitude and of start time as the independent variables, the model could not converge. Thus, I centered those two variables in the LMMs.

4.3.3. Discussion

The results confirm our hypotheses and hence do support the assumption that culture had the predicted impact on image viewing.

Firstly, in comparison to the Austrian participants the Japanese did perform a lower amount of horizontal and a higher amount of vertical saccades, but only when viewing artworks. Moreover, the Austrian participants performed more upward than downward saccades, whereas Japanese had a similar proportion in both directions, resulting in a higher percentage of downward saccades compared to the Austrian participants. The higher proportion of horizontal saccades is in accordance with a higher habituation to horizontal reading in the Austrian group. The higher proportion of upwards saccades in Austrian participants might be related to being used to perspectival pictures where the bottom is the foreground, the place next to the viewer where the gaze may start to inspect a scene. Conversely, the higher proportion of downward saccades in the Japanese group might be related to being used to write and read from top to bottom.

Secondly, in comparison to the Japanese participants the amplitude of horizontal saccades was larger in the Austrian group. This might also be related to habituation to horizontal reading. Yet, we did not find significant differences of amplitude for vertical saccades between both groups of participants.

Beyond the confirmation of our hypotheses a remarkable and unforeseen pattern emerged in the data. Whereas Austrian participants had similar oculometric parameters in all stimuli blocks, the parameter of Japanese participants varied in relation to the type of images they were looking at:

1. When looking at pictures (Japanese/European artworks) versus photographs, especially photographs declared as everyday scenes, Japanese participants performed, firstly, a higher amount of vertical and a lower amount of horizontal saccades, and secondly, saccades with smaller amplitudes, especially in horizontal direction.
2. When looking at Japanese art the amplitude of the saccades of Japanese participants did not decrease over time.
3. Whereas the fixation duration of Austrian participants increased over the viewing time in all conditions; fixation duration of Japanese participants remained unchanged over the viewing time except when viewing photographs declared as art.

In order to understand those unforeseen differences, it is helpful to recall research on the two successive phases of image viewing. Since Buswel (1935), several studies have shown that while viewing images, the amplitude of saccades diminishes while the duration of fixations augments. This has been interpreted as a succession of two different phases of perception: A first phase of general survey with longer saccades and shorter fixations followed by a phase of more detailed study with shorter saccades and longer fixations (e.g., Fischer et al., 2013; Pannasch et al., 2008). However, to the best of our knowledge, none of the studies enquiring into the changes of oculometric parameters during free picture viewing ever included Japanese (or any East-Asian participants).

Remarkably, our study does confirm the two phases model, but only for the Austrian participants. For Japanese participants, the dynamic of saccade amplitude and fixation duration depended on the type of stimuli:

1. When viewing Japanese art, both the amplitude of the saccades and the duration of fixations of Japanese participants remained largely unchanged. This result contradicts the previous studies and is rather a behavior connected to reading.
2. When viewing European art and everyday photographs, the amplitude of the saccades of Japanese participants did

decrease. However, the duration of their fixations remained largely unchanged. We interpret this as half way between image viewing on the basis of earlier studies and reading. 3. When viewing ‘artistic’ photographs, the amplitude of the saccades of Japanese participants decreased and the duration of their fixations augmented. This is the only stimuli block where Japanese participants had a similar behavior as the Austrians.

If differences between the Austrians and Japanese were only due to the habituation to different writing systems, they would impact participants independently of the stimuli. Habituation to writing systems seems significant, but it does not completely explain our results. We therefore suggest not only to consider diverse *cognitive styles*, as Baxandall assumes, but also diverse *cognitive expectations*: Austrian and Japanese participants in our study seem to have had different expectations depending on the type of stimuli. Those expectations could be based on their visual cultures: for Austrians (and more generally Europeans), looking at pictures is to a certain degree similar as looking through a window into the world. This is historically rooted in the Renaissance and its revolution of painting on the base of perspective. In 1435, Alberti delivered the first description of perspective and defined the painting as a perspectival window (Alberti & Sinisgalli, 2011). In contrast, for Japanese (and possibly Chinese as well) looking at a painting is to a certain degree related to reading calligraphy, rather than looking at the world through a window. Writing with ideograms (*Kanji*), is per se a pictorial act. In contrast to European paintings, traditional Japanese paintings have no sharp demarcation between calligraphy and painted image. Japanese scrolls used to decorate the walls of particular rooms (*Kakemono/Kakejikus*) can consist of images—with or without calligraphy—or of pure calligraphy. There is a long tradition for describing the proximity of art and written text as a peculiarity of Japanese (and Chinese) art and as a fundamental difference from the Western tradition. Shelton and Okayama’s (2006) for instance stated that “pictures in Japan are made to be read.” Our results show that this is more than a metaphor.

4.4. General discussion for Study 3

In Study 3, we investigated the cross-cultural similarity/difference in art evaluation and behavior through three empirical studies. These studies were motivated to tackle the second limitation in the field of empirical aesthetics: the skewed participant demographics in past research and to challenge the implicit assumptions of the universality in aesthetics in the past research.

In Study 3-1, we tested if one of the ever-suggested universal visual feature, symmetry, would be appreciated universally. As a result, our Japanese participant did associate symmetry and positivity, and evaluate symmetry as more beautiful than asymmetry as with the Austrian participants in Weichselbaum et al., (2018).

In Study 3-2, though rather explorative, we tested if the art viewers native language is a significant impact on predicting beauty appraisal, and if there are any differences in art attributes predicting beauty across cultures (Japan vs. Austria). Our results showed that while there were several attributes which were equally predicting beauty in both cultures i.e., *visual harmony*, *variety of colors* etc., there were also some attributes which were important to predict beauty only for native German speakers e.g., *emotional expressiveness*, and only for native Japanese speakers e.g., *brushstroke*, *realism*, *focus* etc.

In Study 3-3, we further tested if the habitual behavior in culture i.e., reading/writing system would have impact on the way we explore artworks and non-artworks. As a result, the participants from Japan and Austria did make different eye movement patterns in accordance with their own reading/writing system. However, our results further showed that the expectations for the seen objects might also play a role, since our results differed depending on the stimulus types.

Overall, the results of Study 3-1 and 3-2 did suggested that humans share some amount of criteria for aesthetics, especially when the visual features are tied to biological cues or to our visual information processing systems i.e., symmetric feature are connected to actual physical health (Fink et al., 2006) and good genes (Thornhill & Møller, 1997). However, the results from Study 3-2 and 3-3 clearly suggested that culture might have significant impacts on art appraisal as well as art

viewing behavior. Furthermore, our results further suggested that the differences could be stem from the visual cultures such as reading/writing systems, the cultural practice in each culture.

As Che et al. (2018) states, “Given the scarcity of cross-cultural studies, we can only assume that the findings of empirical aesthetics apply to all humans; to those living in large-scale industrialized and small-scale nonindustrialized societies alike” (p. 20). Considering that our comparisons were made between Japan and Austria, where both countries could be recognizes as industrialized ones, the differences between the industrialized and nonindustrialized culture would be more salient. Che et al. (2018) also argue that “the evidence supporting the assumption of cross-cultural extensibility is uneven, scarce, and weak” (p. 20). As aforementioned, as a general limitation especially for Study 3-2, one might argue that the study is rather explorative and do not have a clear focus to test. Nevertheless, as Che et al. (2018) argued, we believe that our studies, where we systematically tested if culture might have a significant impact on art experience and tried to describe where the difference across culture could be found, are going to inspire other researchers to build other solid hypotheses in the future studies.

5. General Discussion

This dissertation (“*Empirical Investigations on Art Viewing Behavior and Art Evaluation: An Interdisciplinary and Cross-Cultural Study*”) consists of three main sections: (1) Translating well-documented attention decrease phenomenon towards artworks from museum studies into lab-based settings, (2) Empirical investigation on how to counteract the attention decrease phenomenon in lab-based settings, and (3) Cross-cultural similarities/differences in art viewing behavior and evaluation of art between native Japanese and German speakers. Though the topics/methodologies in the present are quite diverse, they all have one common motivation that is to address general limitations in the field of empirical aesthetics: (1) a gap between empirical aesthetic research, typically conducted in a lab-based setting, and actual experience of art in real-life settings and (2) skewed participant samples—past studies in the field have been exclusively conducted either in North American or European countries, and other ones have been neglected. Here, I present a recap for each study.

Study 1 investigated if the ever-presented attention decrease phenomenon from museum studies could also be observed in a laboratory. Study 1-1 presented an updated procedure to measure the participants’ attention towards artwork where the participants could not truncate the total duration of the experiment by setting a fixed time for one trial. Study 1-2 provided one more replication, and further tested if the resting period during the successive art viewing could stop/mitigate the attention decrease phenomenon. Further, in order to see how robust this phenomenon could be found, we recruited participants from Japan and Austria. As a result, both studies showed that viewing time as a proxy of participants’ attention decreased over trials. Furthermore, we also found that not only the viewing times, but also art appraisal decreased over trials. Also, the results of Study 1-2 showed that the resting period during the art viewing did not

have significant impacts on the decrease in the viewing times, suggesting that the common practices in museums/galleries such as putting a bench to sit in exhibitions might not have power to counteract this phenomenon.

Study 2 further investigated how to counteract the phenomenon and why this happens. Specifically, Study 2-1 investigate if one of the presented ways to counteract it, providing information about artworks, could mitigate the decrease in the viewing times. However, our results showed that, contrary to the results reported from the past studies (e.g., Robinson, 1928), the manipulation only has power to maintain the attention of the participants at the very beginning of the experiments. Study 2-2 further investigated if habituation is a likely cause of the phenomenon, and if the stimulus novelty in the sequences could stop/mitigate the decrease in viewing time. As a result, though our hypotheses were not fully supported, we found similar tendencies as with dishabituation process, and also found that the stimulus novelty could mitigate the decrease in viewing times at the beginning of the experimental task.

Study 3 had a different focus compared to Study 1 and 2, specifically, we aim to study cross-cultural similarity/difference in art experience. Study 3-1 tested if symmetry is a universal feature determining preference/beauty appraisal. Indeed, our results showed that the Japanese participants preferred symmetry to asymmetry, and evaluated symmetry as more beautiful, suggesting an inter-cultural agreement in aesthetics. In Study 3-2, we systematically investigated if culture is a significant factor for beauty appraisal, and if there are any cultural differences in art attributes constituting beauty appraisal. Our results showed that art experience might not be fully shared across cultures, and visual culture and/or differences in cultural practice might influence art experience. Given the fact that our daily practice in culture might have an impact on art experience, in Study 3-3, we further tested if the different writing/reading system in culture would make different in how we visually explore the art and non-art stimuli. In general, we found that the Japanese participants made more vertical saccades, which is correspond to the writing/reading system. In Figure 19, an overview of the structure of the present dissertation is presented. The relationships between the studies included in each section were

already discussed in the discussion in each section, and thereby further remarks are not going to be made on this matter. Here, in the general discussion, I will rather discuss the connections between each study, caveats in this dissertation, and what to be expected in the future studies.

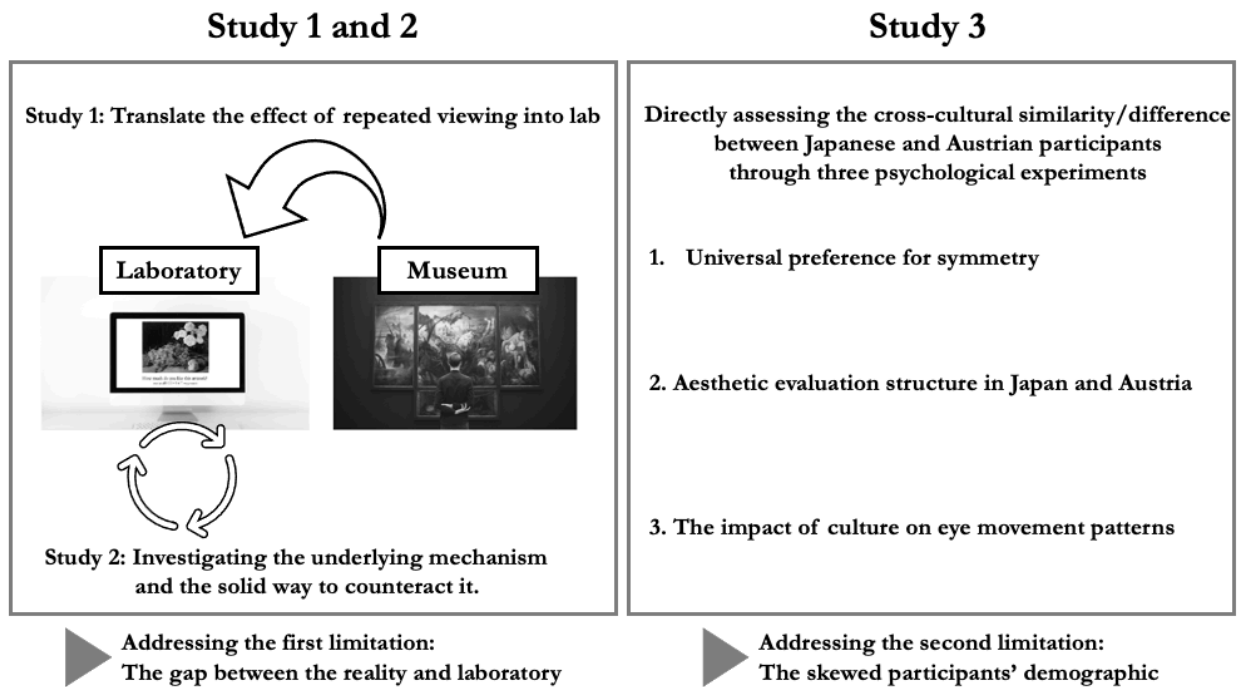


Figure 19. An overview of the structure of the present dissertation

The connection between Study 1 and Study 2

Study 1 and 2 constituted one marching study addressing the first limitation in empirical aesthetics. Study 1 provided the core basis for Study 2 by translating the ever-present attention decreased phenomenon into a laboratory with participants from different cultures, since if the effect is not present, further investigations for the causes in a laboratory does not seem to be valid. Part of the findings in Study 1 further constituted the basis of Study 2. Namely, in Study 1, we showed that the decrease in attention and art appraisal were quite robust across both cultures. The focus on habituation process in Study 2-2 was based on this robustness of this phenomenon. Since habituation

is one of the fundamental behavioral processes even across species, it seemed plausible to assume that such fundamental process could be an important factor.

As aforementioned in the general introduction, though the growing awareness has been made on the differences in art experiences between the real-life and lab-based settings, little has been studies how we can actually bridge/complete the findings from the two viewing contexts. Thereby, the two studies stand as a pioneering example of how we can conduct studies where the studies in ecologically valid setting and in a laboratory could complement each other's shortcomings. Of course, it is very time consuming and requires a lot of resources to perform multiple replications/pilot and main studies. However, considering that those works might bring key knowledge, I hope that these attempts are going to be applied in the future study as well.

Further, though it is not directly relevant to the main aim of Study 1 and 2, the two studies present a variety of analyses dealing with longitudinal data i.e., GLMMs, Tau-U, and LMMs. All these analysis methods could deal with time-series like data, which will be problematic with the usage simple *t*-tests and/or ANOVA. These analysis methods could be potentially applied to the topic such as the stream of aesthetic experiences (e.g., Muth et al., 2015). Thus, not only for assessing the data with multiple trials, but also for the behavior/physiological metrics changes while one person is appreciating one artwork for some time, the presented analyses would insightful.

The connection between Study 1/2 and 3

Study 3 seems to be isolated from Study 1 and 2, since it mainly addressed the second limitation by comparing the art viewing behavior and evaluation across two cultures (Japan and Austria). Importantly, the motivation of Study 3 could also be, though indirectly, connected to the one of Study 1 and 2. As mentioned, the studies included in Study 1 and 2 mainly focused on tackling the first limitation of empirical aesthetics, while Study 3 focused on tackling the second limitation. Study 1 and 2 were motivated to perform lab-based studies in order to examine one

phenomenon which is well-demonstrated in ecologically valid settings. In other words, they tried to capture/examine/understand a real-world phenomenon in lab-based basic psychological experiments. The studies in Study 3 actually shared this aim. Specifically, people who are visiting museums/art galleries, who are appreciating artworks at home/cafes are not all WEIRD people, but they are a mixed population from all over the world. In this sense, the samples in the past studies in empirical aesthetics are quite small populations compared to who are actually happening in the real-life setting, and thereby, one might argue that they do not reflect a real-world setting, which is in line with the arguments for Study 1 and 2. Hence, cross-cultural comparisons could contribute to the field by not only challenging the implicit universal claims in the past studies, but also by considering the diverse/international populations who reflect the art viewers in the real-life settings. Together, though of course this dissertation deals with two separate limitations in empirical aesthetics, they somehow share a meta-motivation: to reflect the real-life settings in scientific research. Of course, the dissertation only covers Japanese and German native speakers, and further replications and investigations with populations with different cultures/countries are necessary for a comprehensive understanding of art experience. Nevertheless, considering that many journals explicitly state that they welcome studies with non-WEIRD people, I do hope that future studies/models in empirical aesthetics will integrate findings from diverse populations to capture the nature of art experience.

Possibly, the findings from the dissertation as a whole can be combined in the future. For example, the results of Study 3-3 might be connected to the results of Study 2-2. Along with the past studies, Study 3-3 suggested that Japanese participants tend to focus on more broad areas in the visual field. Hence, though this is not the ‘focal’ object in an exhibition space, they might be more aware of the wall colors of the exhibition space compared to the Austrian participants. If this hypothesis holds, it might be possible that people from Japan might show the recoveries in viewing time towards artworks after the change of the wall color in an exhibition space, while this effect might disappear when Austrian participants experience the same wall color changes. This might be useful knowledge

for local museums in Japan or in countries where the participants show the similar behavioral tendency, since the decreased attention phenomenon can be avoided by not changing the order/context of the exhibits but only changing the exhibition settings.

Limitations of the present dissertation and the possible avenues for future studies

First, as introduced in the general introduction, there have been considerable amount of works which have reported that people would behave differently in museums than in a laboratory. Thus, it is still possible that the findings from Study 2-2—the novelty in the stimulus sequences could partially modulate the attention decrease at the beginning of art viewing task—would not work or work differently in actual museum exhibitions. Indeed, the participants in the present dissertation came to a lab to join the study with monetary compensation, which is completely a different setting compared to the natural museum visits. Thereby, the same method will bring different results in museum/art gallery exhibitions. For museums and galleries, it is surely important that the way to counteract the decreased attention also works in a setting with multiple confounding variables. Consequently, in order to assess if the way to counteract decreased attention is effective in the real-life settings or not, replication studies in actual museum are necessary. Actually, we did plan to conduct a replication study in a museum exhibition place. However, the COVID-19 pandemic made it quite difficult to conduct psychological studies in person in museums due to the safety issues. As one future study idea, it might be interesting to observe museum visitors' behavior in an exhibition where a variety of contents as well as the period of artworks are presented at a time. For example, in one condition, the participants are going to be handed in an introduction which shows the way to go around the museums in a chronological order as in the PC condition in Study 2-2. In another condition, the participants are going to be handed an introduction which shows the way to go around in a category-wise order as in the CC condition in Study 2-2. Though this is also slightly different from the natural experience of artworks, since we can go through museum exhibitions freely in reality, with this way, it might be possible to assess the effect of the novelty in actual museum

exhibitions, while assuring that the participants are going around the exhibits in a different context. This method will also provide an opportunity to compare the results from two distinct conditions in the same exhibition space, which might be able to control some of the confounding factors e.g., the architectural structure of the exhibitions. With the replications in actual museum exhibitions, it is expected that a multi-method research approach will be established wherein basic research in a laboratory and research in real-life situations fully complete each other's advantages and disadvantages.

Second, as one more general limitation in Study 1 and 2, one note should be made that the present dissertation only focused on the viewing times. Nevertheless, there are other proxies for attention e.g., percentage of stops and/or self-reported measure. For example, Bitgood (2009a) explicitly pointed out that viewing time could not be used as a metric representing viewers' 'interest' towards the objects. Thus, in order to examine the impact of repeated art viewing on the viewers' interest towards artworks, future research which explicitly ask the self-report measures (e.g., feeling, values) are necessary.

Third, regarding the cross-cultural components in the present dissertation, it is important to note that the findings might not be applied to other countries in. For example, in Study 3-3, the hypotheses on the cross-cultural differences in eye movement patterns when people are viewing artworks were build based on the reading/writing systems in Japan and Austria. However, in East Asian cultures, there are a wide variety of script systems, and many countries use different alphabets. For example, China does not use Japanese Hiragana and Katakana, and South Korea does not have Kanji, and instead using Hangul. These differences in each culture presumably influence our visual cultures e.g., what kind of shapes/which reading direction we are familiar with. Thereby, in order to generalize the findings to the findings to see if the tendency is similar across East Asian and Western cultures, further replication studies are necessary. This remark applies not only to this dissertation, but to all studies which include cross-cultural components in general. Since the impact of visual culture can be quite sensitive to so many aspects i.e., the form of

artworks, political situation, religious, alphabets, reading/writing directions, over-selling or -generalizations should be restricted.

Last, though this aspect was discussed in Study 3-2, the cross-cultural comparisons should be based on a clear hypothesis, focusing a specific difference in distinct cultures. Since, as mentioned, *culture* could represent so many different things, studies without clear focus might fail to add knowledge.

In the present dissertation, with the opportunity of the joint supervised program between Keio university and Vienna university, the two limitations in Empirical Aesthetics are explicitly addressed with different methodologies/foci. In order to achieve comprehensive understanding of experience of art, I do hope that the current dissertation highlights the importance of reflecting the real-life situation into the scientific work and will inspire future empirical studies which will eventually lead more comprehensive understanding of art experience.

6. References

- Alberti, L. B., & Sinisgalli, R. (2011). *On Painting: A New Translation and Critical Edition*. Cambridge University Press.
- Anderson, R. L. (2004). *Calliope's Sisters: A Comparative Study of Philosophies of Art (2nd Edition)* (2nd ed.). Pearson Prentice Hall. <https://doi.org/10.2307/431487>
- Arnheim, R. (2004). *Art and Visual Perception, Second Edition: A Psychology of the Creative Eye* (Second Edition, Fiftieth Anniversary Printing ed.). University of California Press.
- Augustin, M. D., Leder, H., Hutzler, F., & Carbon, C. (2008). Style follows content: On the microgenesis of art perception. *Acta Psychologica*, 128, 127–138. <https://doi.org/10.1016/j.actpsy.2007.11.006>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Baxandall, M. (1988). *Painting and Experience in Fifteenth-Century Italy: A Primer in the Social History of Pictorial Style (Oxford Paperbacks)* (2nd ed.). Oxford University Press.
- Berlyne, D. E. (1973). Aesthetics and psychobiology. *Journal of Aesthetics and Art Criticism*, 31(4), 553. <https://doi.org/10.2307/429334>
- Berlyne, D. E. (1975). Dimensions of perception of exotic and pre-Renaissance paintings. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 29(2), 151–173. <https://doi.org/10.1037/h0101682>

- Bertamini, M., & Blakemore, C. (2019). Seeing a Work of Art Indirectly: When a Reproduction Is Better Than an Indirect View, and a Mirror Better Than a Live Monitor. *Frontiers in Psychology, 10*. <https://doi.org/10.3389/fpsyg.2019.02033>
- Bitgood, S. (2002). Environmental psychology in museums, zoos, and other exhibition centers. In R. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 461–480). New York, NY: John Wiley & Sons.
- Bitgood, Stephen. (2009a). Museum Fatigue: A Critical Review. *Visitor Studies, 12*(2), 93–111. <https://doi.org/10.1080/10645570903203406>
- Bitgood, S. (2009b). When Is “Museum Fatigue” Not Fatigue? *Curator: The Museum Journal, 52*(2), 193–202. <https://doi.org/10.1111/j.2151-6952.2009.tb00344.x>
- Bitgood, S. (2015). Museum fatigue and related phenomena. *JVA InterpNews, 4*, 25–29.
- Bitgood, S., Devia, C., Goodwin, P., Rudin, C., Trammell, A., & Zimmerman, B. (2001). Projective focusing: Motivated attention to works of art. *Current Trends in Audience Research and Evaluation, 14*, 25–31.
- Bitgood, S., Patterson, D., & Benefield, A. (1986). *Understanding your visitors: The factors that influence visitor behavior*. American Association of Zoological Parks & Aquaria. Annual Conference, Houston, TX.
- Bode, C., Helmy, M., & Bertamini, M. (2017). A cross-cultural comparison for preference for symmetry: Comparing British and Egyptians non-experts. *Psihologija, 50*(3), 383–402. <https://doi.org/10.2298/psi1703383b>
- Boland, J. E., Chua, H. F., & Nisbett, R. E. (2008). How we see it: Culturally different eye movement patterns over visual scenes. In K. Rayner, D. Shen, X. Bai, & G. Yan (Eds.), *Cognitive and cultural influences on eye movements* (pp. 363–378). Tianjin People’s Press/Psychology Press.
- Bradley, M. M., & Lang, P. J. (2017). International Affective Picture System. *Encyclopedia of*

- Personality and Individual Differences*, 1–4. https://doi.org/10.1007/978-3-319-28099-8_42-1
- 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>
- Brinkmann, H. (2021). Kulturelle Varianzen bei der Kunstbetrachtung: Überlegungen zur empirischen Erforschung von Sehgewohnheiten. *21: Inquiries into Art, History, and the Visual*, 2(1), 159-194. <https://doi.org/10.11588/xxi.2021.1.79018>
- Broer, P. N., Juran, S., Liu, Y. J., Weichman, K., Tanna, N., Walker, M. E., Ng, R., & Persing, J. A. (2014). The Impact of Geographic, Ethnic, and Demographic Dynamics on the Perception of Beauty. *Journal of Craniofacial Surgery*, 25(2), e157–e161. <https://doi.org/10.1097/scs.0000000000000406>
- Buswell, G. T. (1935). *How People Look At Pictures: A Study Of The Psychology Of Perception In Art*. University of Chicago Press.
- Carbon, C. C. (2011). Cognitive Mechanisms for Explaining Dynamics of Aesthetic Appreciation. *I-Perception*, 2(7), 708–719. <https://doi.org/10.1068/i0463aap>
- Carbon, C. C. (2019). Empirical Approaches to Studying Art Experience. *Journal of Perceptual Imaging*, 2(1), 10501–1. <https://doi.org/10.2352/j.percept.imaging.2019.2.1.010501>
- Carbon, C. C. (2020). Ecological Art Experience: How We Can Gain Experimental Control While Preserving Ecologically Valid Settings and Contexts. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00800>
- Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, 11, 2079-2107. <https://dl.acm.org/doi/10.5555/1756006.1859921>

- 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>
- Chatterjee, A. (2003). Prospects for a cognitive neuroscience of visual aesthetics. *Bulletin of Psychology and the Arts*, 4, 55–60. <https://doi.org/10.1037/e514602010-003>
- Chatterjee, A., Widick, P., Sternschein, R., Smith, W. B., & Bromberger, B. (2010). The Assessment of Art Attributes. *Empirical Studies of the Arts*, 28(2), 207–222. <https://doi.org/10.2190/em.28.2.f>
- Che, J., Sun, X., Gallardo, V., & Nadal, M. (2018). Cross-cultural empirical aesthetics. *Progress in Brain Research*, 237, 77–103. <https://doi.org/10.1016/bs.pbr.2018.03.002>
- Che, J., Sun, X., Skov, M., Vartanian, O., Rosselló, J., & Nadal, M. (2021). The role of working memory capacity in evaluative judgments of liking and beauty. *Cognition and Emotion*, 35(7), 1407–1415. <https://doi.org/10.1080/02699931.2021.1947781>
- Chmiel, A., & Schubert, E. (2017). Back to the inverted-U for music preference: A review of the literature. *Psychology of Music*, 45(6), 886–909. <https://doi.org/10.1177/0305735617697507>
- Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). From The Cover: Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences*, 102(35), 12629–12633. <https://doi.org/10.1073/pnas.0506162102>
- Cota-Mckinley, A. (1999). An empirical investigation of workload on the visitor experience: Effects of exhibit size and time constraints on performance (Doctoral dissertation). Colorado State University, Ft. Collins, Co.
- Cronin, D. A., Hall, E. H., Goold, J. E., Hayes, T. R., & Henderson, J. M. (2020). Eye Movements in Real-World Scene Photographs: General Characteristics and Effects of Viewing Task. *Frontiers in Psychology*, 10, 2915. <https://doi.org/10.3389/fpsyg.2019.02915>
- Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C. H. (1995). “Their ideas

- of beauty are, on the whole, the same as ours”: Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, 68(2), 261–279. <https://doi.org/10.1037/0022-3514.68.2.261>
- Davey, G. (2005a). What is museum fatigue? *Visitor Studies Today*, 8(3), 17–21. https://www.researchgate.net/profile/Gareth-Davey-2/publication/292110022_What_is_museum_fatigue/links/5dd62f89458515cd48affe7c/What-is-museum-fatigue.pdf
- Davey, G. (2005b). *The influence of demographic, behavioural, socioeconomic, and environmental design factors on visitor behaviour in Britain and China* (Unpublished PhD thesis). University of Bolton, Manchester, England.
- Ensor, K. M., & Hamilton, M. (2014). Effect of Distinctive Frames on Memory for Pictures. *Empirical Studies of the Arts*, 32(2), 121–131. <https://doi.org/10.2190/em.32.2.eov.6>
- Falk, J., & Dierking, L. (2002). *Lessons Without Limit: How Free-Choice Learning is Transforming Education* (Illustrated ed.). AltaMira Press.
- Falk, J. H. (1983). Time and behavior as predictors of learning. *Science Education*, 67(2), 267–276. <https://doi.org/10.1002/sce.3730670214>
- Falk, J. H., Koran, J. J., Dierking, L. D., & Dreblow, L. (2010). Predicting Visitor Behavior. *Curator: The Museum Journal*, 28(4), 249–258. <https://doi.org/10.1111/j.2151-6952.1985.tb01753.x>
- Falk, J., & Storksdieck, M. (2005). Using the contextual model of learning to understand visitor learning from a science center exhibition. *Science Education*, 89(5), 744–778. <https://doi.org/10.1002/sce.20078>
- 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. <https://doi.org/10.3758/bf03193146>
- Fechner, G. (1876). *Vorschule der Ästhetik*. Breitkopf & Härtel.

- Fischer, T., Graupner, S. T., Velichkovsky, B. M., & Pannasch, S. (2013). Attentional dynamics during free picture viewing: Evidence from oculomotor behavior and electrocortical activity. *Frontiers in Systems Neuroscience*, 7. <https://doi.org/10.3389/fnsys.2013.00017>
- Fink, B., Neave, N., Manning, J. T., & Grammer, K. (2006). Facial symmetry and judgements of attractiveness, health and personality. *Personality and Individual Differences*, 41(3), 491–499. <https://doi.org/10.1016/j.paid.2006.01.017>
- Furnham, A., & Baguma, P. (1994). Cross-cultural differences in the evaluation of male and female body shapes. *International Journal of Eating Disorders*, 15(1), 81–89. [https://doi.org/10.1002/1098-108X\(199401\)15:1<81::AID-EAT2260150110>3.0.CO;2-D](https://doi.org/10.1002/1098-108X(199401)15:1<81::AID-EAT2260150110>3.0.CO;2-D)
- Gartus, A., & Leder, H. (2013). The Small Step toward Asymmetry: Aesthetic Judgment of Broken Symmetries. *I-Perception*, 4(5), 361–364. <https://doi.org/10.1068/i0588sas>
- Gartus, A., Völker, M., & Leder, H. (2020). What Experts Appreciate in Patterns: Art Expertise Modulates Preference for Asymmetric and Face-Like Patterns. *Symmetry*, 12(5), 707. <https://doi.org/10.3390/sym12050707>
- Geday, J., Kupers, R., & Gjedde, A. (2007). As Time Goes By: Temporal Constraints on Emotional Activation of Inferior Medial Prefrontal Cortex. *Cerebral Cortex*, 17(12), 2753–2759. <https://doi.org/10.1093/cercor/bhm002>
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- Gilman, B. I. (1916). Museum fatigue. *The Scientific Monthly*, 12, 67–74. https://www.jstor.org/stable/6127?seq=1#metadata_info_tab_contents
- Goller, J., Mitrovic, A., & Leder, H. (2019). Effects of liking on visual attention in faces and paintings. *Acta Psychologica*, 197, 115–123. <https://doi.org/10.1016/j.actpsy.2019.05.008>
- Grammer, K., & Thornhill, R. (1994). Human (*Homo sapiens*) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology*, 108(3), 233–242. <https://doi.org/10.1037/0735-7036.108.3.233>

- 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. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological Review*, 77(5), 419–450. <https://doi.org/10.1037/h0029810>
- Harris, J. D. (1943). Habituated response decrement in the intact organism. *Psychological Bulletin*, 40(6), 385–422. <https://doi.org/10.1037/h0053918>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning; Data Mining, Inference and Prediction*. Springer, New York.
- Hein, G. (1998). *Learning in the museum*. Routledge.
- 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/10.2139/ssrn.1601785>
- Höfel, L., & Jacobsen, T. (2007). Electrophysiological Indices of Processing Symmetry and Aesthetics. *Journal of Psychophysiology*, 21(1), 9–21. <https://doi.org/10.1027/0269-8803.21.1.9>
- Humphrey, G. (2013). *The Nature of Learning: In Its Relation to the Living System*. Routledge.
- Ishii, K., Miyamoto, Y., Rule, N. O., & Toriyama, R. (2013). Physical Objects as Vehicles of Cultural Transmission. *Personality and Social Psychology Bulletin*, 40(2), 175–188. <https://doi.org/10.1177/0146167213508151>
- Jacobsen, T., Buchta, K., Köhler, M., & Schröger, E. (2004). The Primacy of Beauty in Judging the Aesthetics of Objects. *Psychological Reports*, 94(3), 1253–1260. <https://doi.org/10.2466/pr0.94.3c.1253-1260>
- Jacobsen, T., & Höfel, L. (2001). Aesthetics Electrified: An Analysis of Descriptive Symmetry and Evaluative Aesthetic Judgment Processes Using Event-Related Brain Potentials. *Empirical Studies of the Arts*, 19(2), 177–190. <https://doi.org/10.2190/p7w1-5f1f-njk9-x05b>
- Jacobsen, T., & Höfel, L. (2002). Aesthetic Judgments of Novel Graphic Patterns: Analyses of

- Individual Judgments. *Perceptual and Motor Skills*, 95(3), 755–766.
<https://doi.org/10.2466/pms.2002.95.3.755>
- Jellema, T., & Perrett, D. I. (2003). Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia*, 41(13), 1728–1737. [https://doi.org/10.1016/s0028-3932\(03\)00175-1](https://doi.org/10.1016/s0028-3932(03)00175-1)
- Johnston, R. J. (1998). Exogenous Factors and Visitor Behavior. *Environment and Behavior*, 30(3), 322–347. <https://doi.org/10.1177/001391659803000304>
- Kazandjian, S., & Chokron, S. (2008). Paying attention to reading direction. *Nature Reviews Neuroscience*, 9(12), 965. <https://doi.org/10.1038/nrn2456-c1>
- Kim, H. S., & Sherman, D. K. (2008). What Do We See in a Tilted Square? A Validation of the Figure Independence Scale. *Personality and Social Psychology Bulletin*, 34(1), 47–60. <https://doi.org/10.1177/0146167207309198>
- Kirk, U., Skov, M., Hulme, O., Christensen, M. S., & Zeki, S. (2009). Modulation of aesthetic value by semantic context: An fMRI study. *NeuroImage*, 44(3), 1125–1132. <https://doi.org/10.1016/j.neuroimage.2008.10.009>
- Kliegl, R., Laubrock, J., & Köstler, A. (2015). *Augenblicke bei der Bildbetrachtung. Eine kognitionswissenschaftliche Spekulation zum Links und Rechts im Bild*. Räume-Bilder-Kulturen; Lepper, V, M., Deuflhard, P., & Marksches, C., (Eds), 77-90.
- Klorman, R. (1974). Habituation of Fear: Effects of Intensity and Stimulus Order. *Psychophysiology*, 11(1), 15–26. <https://doi.org/10.1111/j.1469-8986.1974.tb00817.x>
- Koutstaal, W. (1998). Memory for Picture Frames. *Empirical Studies of the Arts*, 16(1), 47–57. <https://doi.org/10.2190/4cch-9030-3thm-ltmh>
- Kubovy, M. (2020). Lives as Collections of Strands: An Essay in Descriptive Psychology. *Perspectives on Psychological Science*, 15(2), 497–515. <https://doi.org/10.1177/1745691619887145>

- Kwart, D. G., Foulsham, T., & Kingstone, A. (2012). Age and Beauty are in the Eye of the Beholder. *Perception*, 41(8), 925–938. <https://doi.org/10.1068/p7136>
- 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., Carbon, C. C., & Ripsas, A. L. (2006). Entitling art: Influence of title information on understanding and appreciation of paintings. *Acta Psychologica*, 121(2), 176–198. <https://doi.org/10.1016/j.actpsy.2005.08.005>
- Leder, H., Mitrovic, A., & Goller, J. (2016). How Beauty Determines Gaze! Facial Attractiveness and Gaze Duration in Images of Real World Scenes. *I-Perception*, 7(4), 204166951666435. <https://doi.org/10.1177/2041669516664355>
- 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. <https://doi.org/10.1111/bjop.12084>
- Leder, H., Tinio, P. P. L., Brieber, D., Kröner, T., Jacobsen, T., & Rosenberg, R. (2018). Symmetry Is Not a Universal Law of Beauty. *Empirical Studies of the Arts*, 37(1), 104–114. <https://doi.org/10.1177/0276237418777941>
- Leder, H., Tinio, P. P. L., Fuchs, I. M., & Bohrn, I. (2010). When attractiveness demands longer looks: The effects of situation and gender. *Quarterly Journal of Experimental Psychology*, 63(9), 1858–1871. <https://doi.org/10.1080/17470211003605142>
- Beholder. *Perception*, 41(8), 925–938. <https://doi.org/10.1068/p7136>
- Leventhal, A. M., Martin, R. L., Seals, R. W., Tapia, E., & Rehm, L. P. (2007). Investigating the dynamics of affect: Psychological mechanisms of affective habituation to pleasurable stimuli. *Motivation and Emotion*, 31(2), 145–157. <https://doi.org/10.1007/s11031-007-9059-8>
- Levinson, J. (2003). Philosophical aesthetics: An overview. In J. Levinson (Ed.), *The Oxford*

- handbook of aesthetics (pp. 3–24). Oxford: Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199279456.003.0001>
- Lindell, A. K., & Mueller, J. (2011). Can science account for taste? Psychological insights into art appreciation. *Journal of Cognitive Psychology*, 23(4), 453–475.
<https://doi.org/10.1080/20445911.2011.539556>
- Little, A. (2014). Domain Specificity in Human Symmetry Preferences: Symmetry is Most Pleasant When Looking at Human Faces. *Symmetry*, 6(2), 222–233.
<https://doi.org/10.3390/sym6020222>
- Liu, Z., Zheng, X. S., Wu, M., Dong, R., & Peng, K. (2013). Culture influence on aesthetic perception of Chinese and western paintings. *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction - VINCI '13*.
<https://doi.org/10.1145/2493102.2493111>
- Locher, P. (2011). Contemporary Experimental Aesthetics: State of the Art Technology. *I-Perception*, 2(7), 697–707. <https://doi.org/10.1068/i0449aap>
- 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(4), 449–465. <https://doi.org/10.1068/p3008>
- 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
- 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(2), 121–129. <https://doi.org/10.2190/r1wn-taf2-376d-efuh>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.
- Makin, A. D. (2017). The Gap Between Aesthetic Science and Aesthetic Experience. *Journal of Consciousness Studies*, 1(2), 184–213. chrome-

extension://efaidnbmnnnibpcajpcgglefindmkaj/viewer.html?pdfurl=https%3A%2F%2Fwww.imprint.co.uk%2Fwpcontent%2Fuploads%2F2019%2F03%2FMakin_Open_Access.pdf&clen=2318369&chunk=true

- Makin, A. D., Helmy, M., & Bertamini, M. (2018). Visual cortex activation predicts visual preference: Evidence from Britain and Egypt. *Quarterly Journal of Experimental Psychology*, 71(8), 1771–1780. <https://doi.org/10.1080/17470218.2017.1350870>
- Makin, A. D. J., Pecchinenda, A., & Bertamini, M. (2012). Implicit affective evaluation of visual symmetry. *Emotion*, 12(5), 1021–1030. <https://doi.org/10.1037/a0026924>
- Mangelsdorff, A. D., & Zuckerman, M. (1975). Habituation to Scenes of Violence. *Psychophysiology*, 12(2), 124–129. <https://doi.org/10.1111/j.1469-8986.1975.tb01261.x>
- Marcellini, D. L., & Jenssen, T. A. (1988). Visitor behavior in the National Zoo's reptile house. *Zoo Biology*, 7(4), 329–338. <https://doi.org/10.1002/zoo.1430070404>
- Marin, M. M., Lampatz, A., Wandl, M., & Leder, H. (2016). Berlyne Revisited: Evidence for the Multifaceted Nature of Hedonic Tone in the Appreciation of Paintings and Music. *Frontiers in Human Neuroscience*, 10, 536. <https://doi.org/10.3389/fnhum.2016.00536>
- Markovic, S., Stevanovic, V., Simonovic, S., & Stevanov, J. (2016). Subjective experience of architectural objects: A cross-cultural study. *Psihologija*, 49(2), 149–167. <https://doi.org/10.2298/psi1602149m>
- Martindale, C. (2007). Recent Trends in the Psychological Study of Aesthetics, Creativity, and the Arts. *Empirical Studies of the Arts*, 25(2), 121–141. <https://doi.org/10.2190/b637-1041-2635-16nn>
- Mastandrea, S., & Crano, W. D. (2018). Peripheral Factors Affecting the Evaluation of Artworks. *Empirical Studies of the Arts*, 37(1), 82–91. <https://doi.org/10.1177/0276237418790916>
- Mastandrea, S., Tinio, P. P. L., & Smith, J. K. (2021). Editorial: Environment, Art, and Museums: The Aesthetic Experience in Different Contexts. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.675165>

- Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, 81(5), 922–934. <https://doi.org/10.1037/0022-3514.81.5.922>
- Matamoros, M. (1986). *Information overload*. University of Massachusetts, Boston.
- Mcsweeney, F. K., & Roll, J. M. (1998). Do animals satiate or habituate to repeatedly presented reinforcers? *Psychonomic Bulletin & Review*, 5(3), 428–442. <https://doi.org/10.3758/bf03208818>
- Melton, A. (1935). *Problems of installation in museum of art*. American Association of Museums.
- Molnar, F. (1981). About the Role of Visual Exploration in Aesthetics. *Advances in Intrinsic Motivation and Aesthetics*, 385–413. https://doi.org/10.1007/978-1-4613-3195-7_16
- Morii, M., Sakagami, T., Masuda, S., Okubo, S., & Tamari, Y. (2017). Correction to: How does response bias emerge in lengthy sequential preference judgments? *Behaviormetrika*, 45(1), 221–222. <https://doi.org/10.1007/s41237-017-0044-6>
- Muth, C., Raab, M. H., & Carbon, C. C. (2015). The stream of experience when watching artistic movies. Dynamic aesthetic effects revealed by the Continuous Evaluation Procedure (CEP). *Frontiers in Psychology*, 6, 365. <https://doi.org/10.3389/fpsyg.2015.00365>
- Muth, C., Raab, M. H., & Carbon, C. C. (2017). Expecting the Unexpected: How Gallery Visitors Experience Semantic Instability in Art. *Art and Perception*, 5(2), 121–142. <https://doi.org/10.1163/22134913-00002062>
- Nadal, M., & Ureña, E. (2021). One hundred years of Empirical Aesthetics: Fechner to Berlyne (1876 – 1976). *PsyArXiv Preprints*. <https://doi.org/10.31234/osf.io/c92y7>
- Nadal, M., & Vartanian, O. (2021). *Empirical Aesthetics: An overview*. M. Nadal & O. Vartanian (Eds.), *The Oxford Handbook of Empirical Aesthetics*. New York: Oxford University Press.
- Nadeau, C. (2003). Inference for the Generalization Error. *Machine Learning*, 52(3), 239–281. <https://doi.org/10.1023/a:1024068626366>

- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108(2), 291–310.
<https://doi.org/10.1037/0033-295x.108.2.291>
- Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual Aesthetics and Human Preference. *Annual Review of Psychology*, 64(1), 77–107. <https://doi.org/10.1146/annurev-psych-120710-100504>
- Pannasch, S., Helmert, J. R., Roth, K., Herbold, A. K., & Walter, H. (2008). Visual Fixation Durations and Saccade Amplitudes: Shifting Relationship in a Variety of Conditions. *Journal of Eye Movement Research*, 2(2). <https://doi.org/10.16910/jemr.2.2.4>
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining Nonoverlap and Trend for Single-Case Research: Tau-U. *Behavior Therapy*, 42(2), 284–299.
<https://doi.org/10.1016/j.beth.2010.08.006>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825–2830. Available from: <http://arxiv.org/abs/1201.0490>
- 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.
<https://doi.org/10.1016/j.newideapsych.2010.04.001>
- 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., 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>
- 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. <https://doi.org/10.3389/fnhum.2016.00160>
- Porter, M. (1938). *Behavior of the average visitor in the Peabody Museum of Natural History Yale University*. American Association of Museums.
- R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, <https://www.R-project.org/>
- Rayner, K., Castelano, M. S., & Yang, J. (2009). Eye movements when looking at unusual/weird scenes: Are there cultural differences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 254–259. <https://doi.org/10.1037/a0013508>
- Rhee, S. C. (2017). Differences between Caucasian and Asian attractive faces. *Skin Research and Technology*, 24(1), 73–79. <https://doi.org/10.1111/srt.12392>
- Rhodes, G. (2006). The Evolutionary Psychology of Facial Beauty. *Annual Review of Psychology*, 57(1), 199–226. <https://doi.org/10.1146/annurev.psych.57.102904.190208>
- Robinson, E. S. (1928). The behavior of the museum visitor. American Association of Museums. <https://doi.org/ED419696>
- Serrell, B. (1997). Paying Attention: The Duration and Allocation of Visitors' Time in Museum Exhibitions. *Curator: The Museum Journal*, 40(2), 108–125. <https://doi.org/10.1111/j.2151-6952.1997.tb01292.x>
- Shelton, B., & Okayama, E. (2006). Between script and pictures in Japan. *Visible Language*, 40(2), 157. <https://www.proquest.com/docview/232920059?pq-origsite=gscholar&fromopenview=true>
- Silver, H. R. (1979). Ethnoart. *Annual Review of Anthropology*, 8(1), 267–307.
<https://doi.org/10.1146/annurev.an.08.100179.001411>

- Smets, G. (1975). Pleasingness vs Interestingness of Visual Stimuli with Controlled Complexity: Their Relationship to Looking Time as a Function of Exposure Time. *Perceptual and Motor Skills*, 40(1), 3–7. <https://doi.org/10.2466/pms.1975.40.1.3>
- Smith, J. K., & Smith, L. F. (2001). Spending Time on Art. *Empirical Studies of the Arts*, 19(2), 229–236. <https://doi.org/10.2190/5mqm-59jh-x21r-jn5j>
- 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(1), 77–85. <https://doi.org/10.1037/aca0000049>
- Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H. (2020). The Vienna Art Interest and Art Knowledge Questionnaire (VAIAK): A unified and validated measure of art interest and art knowledge. *Psychology of Aesthetics, Creativity, and the Arts*, 14(2), 172–185. <https://doi.org/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. <https://doi.org/10.1037/aca0000107>
- Stein, R. (2018). *Museums as Economic Engines*. American Alliance of Museums. <https://www.aam-us.org/2018/01/19/museums-as-economic-engines/>
- Sutherland, C. A. M., Rhodes, G., & Young, A. W. (2017). Facial Image Manipulation. *Social Psychological and Personality Science*, 8(5), 538–551. <https://doi.org/10.1177/1948550617697176>
- Thausing, M. (1983). Die Stellung der Kunstgeschichte als Wissenschaft. *Wiener Jahrbuch Für Kunstgeschichte*, 36(1), 140–150. <https://doi.org/10.7767/wjk.1983.36.1.140>
- Thompson, R. F., & Spencer, W. A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychological Review*, 73(1), 16–43. <https://doi.org/10.1037/h0022681>

- Thornhill, R., & Møller, A. P. (2007). DEVELOPMENTAL STABILITY, DISEASE AND MEDICINE. *Biological Reviews*, 72(4), 497–548. <https://doi.org/10.1111/j.1469-185x.1997.tb00022.x>
- Tinio, P. P., Gerger, G., & Leder, H. (2013). Birds of a feather. . . Generalization of facial structures following massive familiarization. *Acta Psychologica*, 144(3), 463–471. <https://doi.org/10.1016/j.actpsy.2013.08.003>
- 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>
- Tinio, P. P., & Leder, H. (2009). Just how stable are stable aesthetic features? Symmetry, complexity, and the jaws of massive familiarization. *Acta Psychologica*, 130(3), 241–250. <https://doi.org/10.1016/j.actpsy.2009.01.001>
- Treiman, R., & Allaith, Z. (2013). Do reading habits influence aesthetic preferences? *Reading and Writing*, 26(8), 1381–1386. <https://doi.org/10.1007/s11145-012-9424-1>
- Tschacher, W., Greenwood, S., Kirchberg, V., Wintzerith, S., van den Berg, K., & Tröndle, M. (2012). Physiological correlates of aesthetic perception of artworks in a museum. *Psychology of Aesthetics, Creativity, and the Arts*, 6(1), 96–103. <https://doi.org/10.1037/a0023845>
- Uduehi, J. (1995). A cross-cultural assessment of the Maitland Graves Design Judgment Test using U.S. and Nigerian subjects. *Visual Arts Research*, 21(2), 11–18.
- van Paasschen, J., Bacci, F., & Melcher, D. P. (2015). The Influence of Art Expertise and Training on Emotion and Preference Ratings for Representational and Abstract Artworks. *PLOS ONE*, 10(8), e0134241. <https://doi.org/10.1371/journal.pone.0134241>
- Võ, 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. <https://doi.org/10.3758/brm.41.2.534>

- Võ, M. L. H., Jacobs, A. M., & Conrad, M. (2006). Cross-validating the Berlin Affective Word List. *Behavior Research Methods*, 38(4), 606–609. <https://doi.org/10.3758/bf03193892>
- Weichselbaum, H., Leder, H., & Ansorge, U. (2018). Implicit and Explicit Evaluation of Visual Symmetry as a Function of Art Expertise. *I-Perception*, 9(2), 204166951876146. <https://doi.org/10.1177/2041669518761464>
- Weigand, R., & Jacobsen, T. (2021). Beauty and the busy mind: Occupied working memory resources impair aesthetic experiences in everyday life. *PLOS ONE*, 16(3), e0248529. <https://doi.org/10.1371/journal.pone.0248529>
- Weyl, H. (1952). *Symmetry*. Princeton science library. Princeton Uni.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9(2, Pt.2), 1–27. <https://doi.org/10.1037/h0025848>
- Zajonc, R. B., Crandall, R., Kail, R. V., & Swap, W. (1974). Effect of Extreme Exposure Frequencies on Different Affective Ratings of Stimuli. *Perceptual and Motor Skills*, 38(2), 667–678. <https://doi.org/10.2466/pms.1974.38.2.667>
- Zangemeister, W. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia*, 33(8), 1009–1025. [https://doi.org/10.1016/0028-3932\(95\)00014-t](https://doi.org/10.1016/0028-3932(95)00014-t)

7. English Abstract

This dissertation aims to tackle two general limitations in Empirical Aesthetics: (1) A gap in the art experience between real-life and lab-based settings and (2) The skewed participants' demographics in past research. To this aim, three main sections are presented: (a) Translating well-documented attention decrease phenomenon towards artworks from museum studies into lab-based settings, (b) Empirical investigation on the possible causes and how to counteract the attention decrease phenomenon in lab-based setting, and (c) Cross-cultural similarity/difference in aesthetic evaluation and art viewing behavior. The work in (a) and (b) assessed one marching topic, the decreased attention/appraisal towards artworks due to the repeated art viewing. Through the two sections, we tested whether and how this phenomenon in museums can be examined in a lab and how to detect possible causes as well as to counteract it. The work in (c) assessed the impact of cultural background of art beholders on art experience by comparing native Japanese and German speakers through three empirical studies. Specifically, we tested 1. the beauty appraisals for symmetry/asymmetry objects from native Japanese speakers, 2. art dimensional features constituting the beauty appraisals between native Japanese and German speakers, and 3. eye movement patterns while native Japanese and German speakers view non-art and art images. All studies are motivated to assess how *visual culture* would have impacts on art experience. Though three sections deal with different topics, each study shares a common goal to tackle either one of the two general limitations in empirical aesthetics.

8. German Abstract

In der vorliegenden Dissertation wurden mögliche Lösungen für zwei grundlegende Limitationen der empirischen Ästhetik gesucht: (1) Die Kluft zwischen Studien unter Laborbedingungen und solchen im natürlichen Umfeld bei der Untersuchung von Kunstinteraktionen und (2) die verzerrte demographische Zusammensetzung von Studienteilnehmern vorheriger Studien. Der Fokus liegt dabei auf drei Aspekten: (a) Die Übertragung des in Museen gut dokumentierten Effekts eines „Aufmerksamkeitsverlustes gegenüber Kunstwerken“ auf Laborbedingungen, (b) eine empirische Ursachenuntersuchung desselben Phänomens sowie Möglichkeiten, dieses zu reduzieren, und (c) interkulturelle Unterschiede/Gemeinsamkeiten in der Kunstwahrnehmung und im Sehverhalten. Die Arbeit in den Kapiteln (a) und (b) behandelt das Thema des bekannten Effekts, dass die Aufmerksamkeit gegenüber Kunstwerken sowie die Beurteilung dieser nachlassen, wenn sie wiederholt besichtigt werden. Wir untersuchten, ob und wie dieser aus Museen bekannte Effekt unter Laborbedingungen überprüft werden kann, seine Ursachen identifiziert und Konsequenzen gemindert werden können. Kapitel (c) untersucht, welchen Einfluss die unterschiedlichen kulturellen Hintergründe japanischer und deutschsprachiger Kunstbetrachter auf deren Kunsterlebnis haben. Im Rahmen von drei empirischen Studien untersuchten wir: 1. Wie japanische Muttersprachler symmetrische und asymmetrische Objekte hinsichtlich ihrer Schönheit beurteilen, 2. analysierten die zugrundeliegenden Faktoren, die die Beurteilung von Kunstobjekten durch japanische und deutsche Muttersprachler ausmachen, und 3. analysierten die Muster der Augenbewegungen japanischer und deutscher Muttersprachler, während sie Kunst und nicht-künstlerische Bilder betrachteten. Die Motivation hinter diesen Studien ist, zu untersuchen, welchen Einfluss kulturelle Hintergründe auf das Kunsterlebnis haben. Obwohl die Schwerpunkte in allen drei Themenbereichen unterschiedlich sind, haben sämtliche Studien in dieser Arbeit das Ziel, eine der existierenden Limitationen der empirischen Ästhetik zu überwinden.

9. Additional Achievements

Due to the restriction of the structure for a doctoral thesis from Keio University, the scientific work where the Ph.D. candidate acted as one of the co-authors has to be separated from the main body of the doctoral thesis. Consequently, in this section, the scientific work where the Ph.D. candidate acted as one of the co-authors are going to be presented in a cumulative form. Before presenting each work, I present the relationships between each work to the present thesis. In this section, three scientific works will be presented.

In the first study, 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, p. 47-53., we examined if brightness of colors is associated with positivity and if this association could be universality found, using Implicit Association Test (IAT, Greenwald et al., 1998) including participants from Japan and Austria. As a result, we repeatedly showed that the association between brightness and positivity is present regardless of participants' cultures. Hence, this study added another layer of knowledge on the universality in aesthetics, and thereby, is in line with Study 3 in the present dissertation, where I assessed the cross-cultural similarity/differences in aesthetics experiences. The idea to test the ever-postulated relation between brightness and positivity and its universality eventually led us to test another ever-presented preference/values for symmetry in Study 3-1 in the present thesis.

In the second study, Pelowski, M., Cabbai, G., Brinkmann, H., **Mikuni, J.**, Hegelmaier, L. M., Forster, M., Rosenberg, R., & Leder, H. (2020). The kitsch switch—or (when) do experts dislike Thomas Kinkade art? A study of time-based evaluation changes in top-down versus bottom-up assessment. *Psychology of Aesthetics, Creativity, and the Arts.*, we assessed when the art viewers show the potential for modulations in art assessment e.g., when the individual might show changes in their art assessment such as increased or decreased preference for the same works of art. Though

this study is not explicitly examining the decreased attention/beauty appraisals, the changes in the art assessment can be seen as another variant of testing the effect of repeated art viewing, which is the core topic of Study 1 and 2 in the present dissertation.

9.1. 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, p. 47-53.

My Contribution:

I acted as the 6th author of the paper and contributed to recruit participants together with ES and HB, and writing up the draft with other co-authors.



Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

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; Mogensen & 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 Schreiblemayr 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

Available online 23 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, Võ 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, -112.2
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, -80.2
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, $r = -0.140$, $p = .207$.

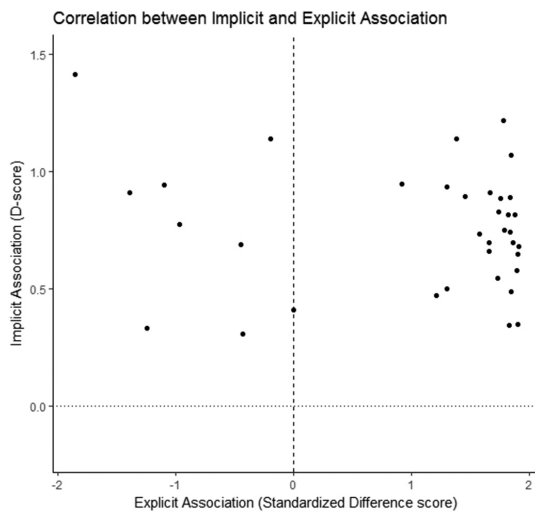


Fig. 1. Scatterplot of the correlation between Explicit and Implicit (D score of the IAT) rating. To aid interpretation lines are drawn on the 0 point of both axis. Because both measures represent a difference score, a positive score, thus above the 0 point on the horizontal axis and right from the 0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

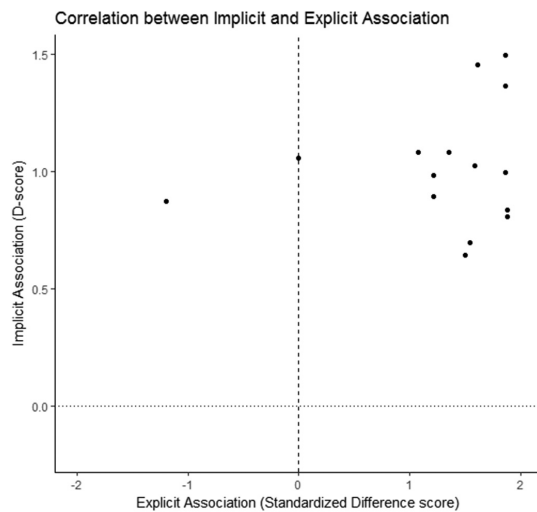


Fig. 2. Scatterplot of the correlation between Explicit and Implicit (D score of the IAT) rating. To aid interpretation lines are drawn on the 0 point of both axis. Because both measures represent a difference score, a positive score, thus above the 0 point on the horizontal axis and right from the 0 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., Persegue, 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 Geistige in der Kunst. Insbesondere in der Malerei*. München: Piper and 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>.

9.2. Pelowski, M., Cabbai, G., Brinkmann, H., Mikuni, J., Hegelmaier, L. M., Forster, M., Rosenberg, R., & Leder, H. (2020). The kitsch switch—or (when) do experts dislike Thomas Kinkade art? A study of time-based evaluation changes in top-down versus bottom-up assessment. *Psychology of Aesthetics, Creativity, and the Arts*.

My Contribution:

I acted as the fourth author of the paper. I was responsible for analysing the data and contributed to make a draft with other authors.

The Kitsch Switch—or (*When*) Do Experts Dislike Thomas Kinkade Art? A Study of Time-Based Evaluation Changes in Top-Down Versus Bottom-Up Assessment

Matthew Pelowski, Giulia Cabbai,
and Hanna Brinkmann
University of Vienna

Jan Mikuni
Keio University

Lisa M. Hegelmaier, Michael Forster, Raphael Rosenberg, and Helmut Leder
University of Vienna

We investigate the potential for modulations in art assessment, involving either “bottom-up” artwork-derived visual features or more overt “top-down” considerations based on personal history or taste. Such changes—whereby individuals might come to relatively increase or decrease their liking of the same works of art—have been suggested in recent art processing models and may potentially occur in addition to, or even in spite of, preframing (i.e., predictive coding) that may come before viewing experience. To assess this, we showed both novice and expert (art history graduate student) viewers abstract and representational paintings, including a specific variety (kitsch, represented by the artist Thomas Kinkade) suggested to provide a particularly salient example of divergent art, and manipulated exposure times to 100 ms, argued to afford only a bottom-up assessment of low-level visual features, 500 ms, allowing awareness of style/content without conscious reflection, and 6,000 and 12,000 ms, potentially allowing more top-down modulations. We hypothesized that by comparing groups and time durations we might identify a point of inflection—a “kitsch switch”—whereby especially the experts come to dislike the kitsch art. Viewers with relatively higher art interest did show lower liking ratings of kitsch paintings at 500 and 6,000 ms when compared to 100 ms. Those with higher interest also showed an opposite pattern of higher appraisals for abstract paintings at the same longer durations. At the same time, we also found robust evidence for a main effect on rating differences between lays and experts even at 100 ms, suggesting both a top-down and bottom-up modulation of ratings and numerous targets for future research.

Keywords: top-down and bottom-up processing, art appreciation, viewing duration, kitsch, appraisal modulation

Supplemental materials: <http://dx.doi.org/10.1037/aca0000302.supp>

“Kitsch passes itself off as artistic communication. But since the fundamental project is . . . simply to overpower . . . or promote a special effect . . . kitsch acts as an intrinsic lie.”

—Umberto Eco (in Nerdrum, 2001, p. 18)

The perception of visual art presents an intriguing focus for psychology because of the complex mix of processes and factors under-

lying its enjoyment. Artworks are expected to involve a blend of information channels as well as aspects of the viewer, the cultural background, and the physical setting. All of which may combine to create a range of appraisals, meanings, and affective responses, which often differ widely between individual viewers or even between individual moments, for the same viewer, when interacting with the same work of art.

Matthew Pelowski, Department of Basic Psychological Research and Research Methods, Faculty of Psychology and Vienna Cognitive Science Hub, University of Vienna; Giulia Cabbai, Department of Psychology, Humboldt-University and Middle European Interdisciplinary Programme in Cognitive Science, University of Vienna; Hanna Brinkmann, Department of Art History, University of Vienna; Jan Mikuni, Department of Psychology, Keio University; Lisa M. Hegelmaier and Michael Forster, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna; Raphael Rosenberg, Department of Art History and Vienna Cognitive Science Hub, University of Vienna; Helmut Leder, Department of Basic Psychological Research and Research

Methods, Faculty of Psychology and Vienna Cognitive Science Hub, University of Vienna.

Giulia Cabbai is now at School of Psychology and Sussex Neuroscience, University of Sussex.

This work was supported by a grant to Matthew Pelowski and Helmut Leder by Marie Skłodowska-Curie MSCA-IF-2014-EF: Individual Fellowships [655379].

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

One particularly intriguing aspect of our processing of art, and a target of recent theoretical emphasis, is the meeting of more or less “top-down” and “bottom-up” factors and the question of how these may each contribute to an emerging experience. According to several processing models (e.g., Chatterjee, 2003; Leder, Belke, Oeberst, & Augustin, 2004; Locher, Krupinski, Mello-Thoms, & Nodine, 2008; see Pelowski, Markey, Forster, Gerger, & Leder, 2017 for review and comparison), a typical interaction with visual art may involve a progression from initial attention to low-level features of the formal object. Viewers may process and respond to certain colors, symmetry, balance, as well as, slightly later, gestalt appearance, style, or mimetic content. These features may be processed relatively automatically (i.e., in a generally “bottom-up” fashion),¹ and themselves may lead to certain appraisals or affective—fluent/disfluent, hedonically pleasing/displeasing, etcetera—responses.

At the same time, these factors may also be integrated with processing tied to specific aspects—personal history, memories, tastes, expectations, training—of the viewer and of the viewing context. These aspects could also influence the initial framing or expectations that we bring into an encounter, helping, for example, to direct attention even when first faced with a stimulus (see e.g., Clark, 2013; Friston, 2010 for recent discussion with such a predictive coding argument). Perhaps even more salient, however, theoretical art processing models suggest that these influences may especially become relevant in later, more overt and cognitively involved (“top down”) processing stages, shaping and often modulating *how* we conceptualize, classify, appreciate, and finally emotionally and evaluatively respond to a stimulus or a work of art (Pelowski, Markey, et al., 2017). The interplay of these broadly top-down and bottom-up processing sequences marks an important topic for empirical study, offering a window into our understanding of basic perception and appraisals (e.g., Corbetta & Shulman, 2002) as well as regarding the refinement of art processing models and our understanding of aesthetic experience.

When considering such top-down/bottom-up integrations, and returning to the particular empirical utility of exploring interactions with stimuli often considered to be art, it may be that these are particularly salient in conjunction with certain types of art and with certain types of art viewers. The dislike, by many, of conceptual, minimalistic, or abstract art may relate to these art forms not rewarding initial viewing or attention to formal features, but rather requiring appreciation related to context and training or contextualization within the broader collection of already produced art (Ortlieb & Carbon, 2018). Similarly, disgusting or challenging images may require certain conceptual approaches (such as a switch to a detached aesthetic mode) allowing individuals to “look past” the appearance to appreciate the wider meaning or significance of a work (Cooper & Silvia, 2009). Even more, this progression could lead to changes or conflict *within* an individual’s processing experience. It may be that we automatically respond to formal features, whether pleasurable or negative, and only later do some individuals reclassify or contextualize art. This could also work in the opposite direction. Certain art styles, although visually pleasing, may be eventually seen by some as ultimately not rewarding, rapid, unhealthy, or even may be ultimately dismissed as “bad” or even “not” art.

This phenomenon suggests an intriguing aspect of the art processing sequence, heretofore not often empirically explored. It may be that there is a certain point of inflection at which top-down

aspects “come online” or become more salient, and thus from which we might find a point of demarcation between those who do and do not respond in certain ways to certain art. This is the target of this article. We consider this topic by an empirical study using a particular form of artworks—kitsch—that is argued to reveal such a conflict between bottom-up and top-down aspects (e.g., Greenberg, 1939), noted, as it is, to be both specifically designed as immediately visually pleasing or easy to process and also often conceptually rejected by critics. We attempt to locate whether, and, if so, when, there is such a point in time of inflection—a “kitsch switch,” so to speak—whereby some (i.e., art experts) change their evaluations. We do this by using a sample of both lay (psychology student) and relatively more expert (art history master student) viewers and experimentally manipulating exposure time to paintings based on current models of aesthetic processing stages. We consider works by Thomas Kinkade (1958–2012), who is perhaps the 21st century’s most notorious representative of kitsch (Perl, 2011), and match these to a set of paintings of similar style but of higher critical acclaim, as well as abstract paintings, which may show an opposite inflection in appreciation especially among a lay population.

Kitsch, Expert Versus Lay Viewers, and Top-Down Versus Bottom-Up Processing

Kitsch seems to be the ideal material to test a time-dependent change of aesthetic valuation, based on different levels of processing. Although it is difficult to place a precise definition on what is and is not kitsch,² if we follow, as do many writers (e.g., see Greenberg, 1939 and below), a basic emphasis on an inherent tension between visual appearance or formal features as they manifest, unreflectively, to a viewer and contextual or conceptual importance and meaning, it becomes clear why it may be a useful

¹ Note that this discussion does involve a certain ambiguity regarding what is meant by “top-down.” Certainly, even from early viewing stages (i.e., intermediate or even early vision, see Chatterjee, 2003), our identification of objects, style, or even a sense of familiarity or novelty is tied to our stored schema and past experience. This stored information is also important when making even an initial assessment. However, we would argue that such a reaction is still largely automatic and unreflective in terms of personal implications or context. Only later does this become more overt and we can talk of deliberate top-down integration.

² We acknowledge that others have attempted to provide a more concrete psychological definition of kitsch involving sets of necessary processing dynamics or appraisal factors. Ortlieb and Carbon (2018), for example, suggested that kitsch might be thought of a stimuli that evokes a fluent, positive, immediate, unreflective emotional experience without encoding of new information (as opposed to stimuli that evoke more disfluent processing, cognitive elaboration, but can also lead to insights). We would generally agree, but would stress that it is more productive to not think of these responses as a product of the stimuli themselves, but a result of the interaction between stimuli, viewer, context and schema/expectations, and which could very well change even within one processing experience. On the other hand, for arguments such as Kulka (2002), who suggested that kitsch is a stimuli that evokes positive emotions, has an immediately identifiable subject matter, and uses a conventional manner of representation, we would not agree. It may precisely be because a work evokes very negative visceral reactions, despite a design which might be positively perceived by many others, that leads an expert to call something kitsch as a pejorative. We also see no reason why kitsch would have to be constrained to only representational depictions, although given present attitudes towards abstract art, probably usually is.

object of study. In fact, kitsch might be marked precisely by its ability to be perceived as both “good,” positive, “art,” etcetera and as “bad,” “not-art” by different individuals or in different moments, depending on one’s way of framing or perceiving, and thus specifically requiring, in order to fit these reactions together, for one to consider the processing experience.

“Kitsch” originated as a German designation for a type of media that was and still is considered poor in taste and associated with decorative, mass commodities, or cheap entertainment (see Ortlieb & Carbon, 2018 for recent review). As put by Greenberg (1939, p. 9) in his treatise on the topic, kitsch can refer to any object of mass culture, designed for wide and immediate enjoyment—“magazine covers, pulp fiction, comics, Hollywood movies, etc.” The term is also particularly used in regard to paintings or pictures rich in superficial sentimentality, presenting the observer with exaggerated imagery that triggers automatic, unreflective dramatic emotional reactions, or bright colors aiming to create immediate positive responses. “We all know the . . . little boy with the tear. The grandmother with the child on her lap and . . . the moose by the lake” recounts Nerdrum (2001, p. 11). This is often indeed reflected in mass appeal, with kitsch enjoyed and purchased among a good deal of the general public.

At the same time, kitsch is connected, especially in its evaluation, to specific expertise- or background-related viewer differences. In the writings of art critics, historians, or other art-interested individuals, kitsch is often considered as a “forbidden area for the educated taste” (Nerdrum, 2001, p. 11; see also Greenberg, 1939), inferior, or “not art” (Karpfen, 1925). Although there may be many reasons *why* certain viewers may not like kitsch, a review of such reasons generally suggests an essential top-down, memory-, or context-based modulation. This includes again the feeling that, although the object might be visually or conceptually pleasing, it is too simple or nonchallenging. One may dislike it, as put by Gombrich (1985, p. 62) “not because he has failed ‘to pick up the message’ but because he understands it all too well.” As also noted by Greenberg (1939, p. 14), because “identifications are self-evident immediately and without any effort on the part of the spectator,” it thus becomes quickly boring or unrewarding (Szpunar, Schellenberg, & Pliner, 2004). This argument can also be framed in psychological discussions of fluency—which generally lead to higher initial liking and security but at the expense of change or new perceptions (Ortlieb & Carbon, 2018, 2019).

Kitsch may also be seen as too beautiful or directly “pretty” (Mekjan, as cited in Nerdrum, 2001), using “old-fashioned” tropes of artistic quality or identification. The visual appearance may also become “too much,” or be considered “bad for” a perceiver—much as with foods that are sweet or fatty.³ Although perhaps initially pleasing, kitsch could become considered “syrupy, saccharined, cloying . . . offer[ing] too obvious, too childish, gratification” thus “invi[t] regression” in one’s taste (Gombrich, 1985, p. 39).⁴ This may also take on a social, class-based, or even moralistic dimension (Lauring et al., 2016).⁵ Kitsch, returning again to the topic of its processing, is described as a poor copy or a simulacrum of “real” culture, allowing the masses to partake in a vaulted institution, but in fact acting as a façade, a “lie,” or a “special effect” (Eco, as cited in Nerdrum, 2001; see also Lorand, 1994; Kulka, 2002), mechanically and immediately evoking certain biological reactions, but ultimately empty—“demand[ing]

nothing of its customers except their money—not even their time” as would be required for later, more complex, reflective processing (Greenberg, 1939, p. 10, emphasis added).⁶ On the other hand, as Greenberg (1939, p. 15) suggests for what might be considered good (nonkitsch) art, the “ultimate values which the cultivated spectator derives . . . are derived at a second remove, as the result of reflection upon the immediate impression left by the plastic values,” they “must be projected into it by the spectator” and “belong to [a] ‘reflected’ effect.”

As framed especially by the above final quote, this then again raises an intriguing empirical question. It could of course be that viewer personality and expectations do modulate art taste and perception from the very beginning, presumably tied to stored schema, predictions, or preclassifications (see below). People may train themselves or naturally develop in such a way as to not prefer, or to even to have an immediate visceral aversion against, certain types of art, even despite initial positive visual appearance (Berridge, Ho, Richard, & DiFeliceantonio, 2010; Garcia, Lasiter, Bermudez-Rattoni, & Deems, 1985; Reilly & Schachtman, 2009). This may also tie to recent arguments for predictive coding whereby our prior schema may play a vital role even in previously thought to be “bottom-up” perception, partially determining what we look for in an encounter with a stimulus (e.g., Clark, 2013). However, we might also find a change within the act of viewing. It may well be that even experts are originally drawn to kitsch because of its visual appeal. At some point, however, they may override this response and change their opinion, or at the very least use more of a relative emphasis on contextual, historical, or art-critical factors, at the expense of attention to formal appearance. It is such a later “reflected effect” (Greenberg, 1939, p. 15) that would again specifically suggest the need for an online switch in processing and relate to the final evaluation and thus interindividual differences.

Kitsch, Exposure Time, and Processing Models

To tease out changes in assessment or between-participants differences, a particularly useful empirical strategy might indeed involve time afforded to viewing. This follows current models of

³ As noted by Berridge et al. (2010), “a once-‘liked’ sweet” might “become unpleasant” as we develop our palate (see also Garcia et al., 1985; Reilly & Schachtman, 2009). Similar arguments are made for music and art (e.g., Gombrich, 1985). However, see Berridge et al. (2010) and Lemmens, Valkenburg, and Peter (2009) for suggestion against changes from actual “liking” (as opposed to simply neutral ratings) to active “dislike” in food research.

⁴ Note, the specific progression from initial pleasantness to eventual cloyingness is not explicitly stated by Gombrich. Rather it is implied in conjunction with his discussion of sweet foods and stated explicitly here by ourselves.

⁵ Perhaps one of the most over-the-top moralistic claims is the essay on kitsch by Hermann Broch (as cited in Nerdrum, 2001, p. 15): “A person who produces kitsch is not someone who produces inferior art. He is not short on skills and can definitely not be judged by esthetic standards. He is quite simply a bad human being . . .”

⁶ Similarly, such a line of argumentation is also used to explain the works/actions of artists who make art that may visually/mimetically “resemble” kitsch but who are still accepted by critics as nonkitsch examples. For example, Jeff Koons (born 1955), an artist broadly acclaimed in the “serious” art world, is seen as making kitsch with an ironic, self-aware perspective in which, the critic believes, the artist shares the joke that the works can only be directly enjoyed by the naïve or uneducated.

art processing that suggest a basic temporal progression between initial low-level attention to formal aspects and then later processing stages in which more top-down features come online or become more salient in modulating responses. As noted previously (see Pelowski, Markey, et al., 2017; Pelowski, Markey, Laturing, & Leder, 2016 for reviews), most art-processing models (Chatterjee, 2003; Leder et al., 2004; Locher et al., 2008) currently follow a basis in models of basic vision (Marr, 1982; see also bottom, Figure 1), which propose a sequence through early, largely automatic, attention to bottom-up visual features, followed by “intermediate” and “later” processing stage in which aspects are matched to viewer schema and context.

For example, as shown in the Figure 1, a standard model for empirically addressing especially the initial processing of art (Leder et al., 2004), proposes five main information-processing stages: (a) preclassification, followed by (b) perceptual analysis that involves initial processing of low-level visual features (e.g., shape, contrast); followed by (c) memory integration and (d) explicit classification, during which individuals may combine basic visual features to identify specific objects as well as style or even artist, and finally (e) cognitive mastering, in which individuals combine elements and create meaning, culminating in aesthetic judgment and affective response. The viewer may end at this point, with an assessment largely tied to the bottom-up content, or may also engage in further assessment cycling again through stages, as might occur if factors draw specific interest and reflecting or integrating more information based on context. This broad

division between top-down and bottom-up aspects is also mirrored in neurobiological models of attention (e.g., Corbetta & Shulman, 2002) which also posit largely distinct—but also interrelated—processing components focused on initial stimulus-driven assessments and later modulation via top down reflection and selective attention to specific elements (see also Graf & Landwehr, 2015, for other, related cognitive fluency/reflection arguments).

As also suggested in Figure 1, these stages and events may be linked to certain time periods (see Pelowski, Markey, et al., 2017, for review), allowing a means for empirical assessment. It has also been shown that by truncating viewing based on seconds or milliseconds, a researcher can at least partially determine the stages involved in a processing experience and thus the relative involvement of the above aspects (Augustin, Leder, Hutzler, & Carbon, 2008; Bachmann & Vipper, 1983; Brieber, Forster, & Leder, 2018; Jakesch & Leder, 2009; Leder & Nadal, 2014; Marin & Leder, 2016). To this end, this study used four exposures, with two points where we expected to potentially find a top-down/bottom-up demarcation:

First, 100 ms was suggested as a minimum exposure needed for impression formation based on a largely preconscious, bottom-up assessment of visual information. Several authors (Cupchik, 2011; Kuchinke, Trapp, Jacobs, & Leder, 2009; Locher et al., 2008; Locher, Overbeek, & Wensveen, 2010) report that this time allows individuals to detect and group artwork characteristics such as color, order, or complexity, and even make assessments regarding pleasantness or interestingness, based on a single fixation.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

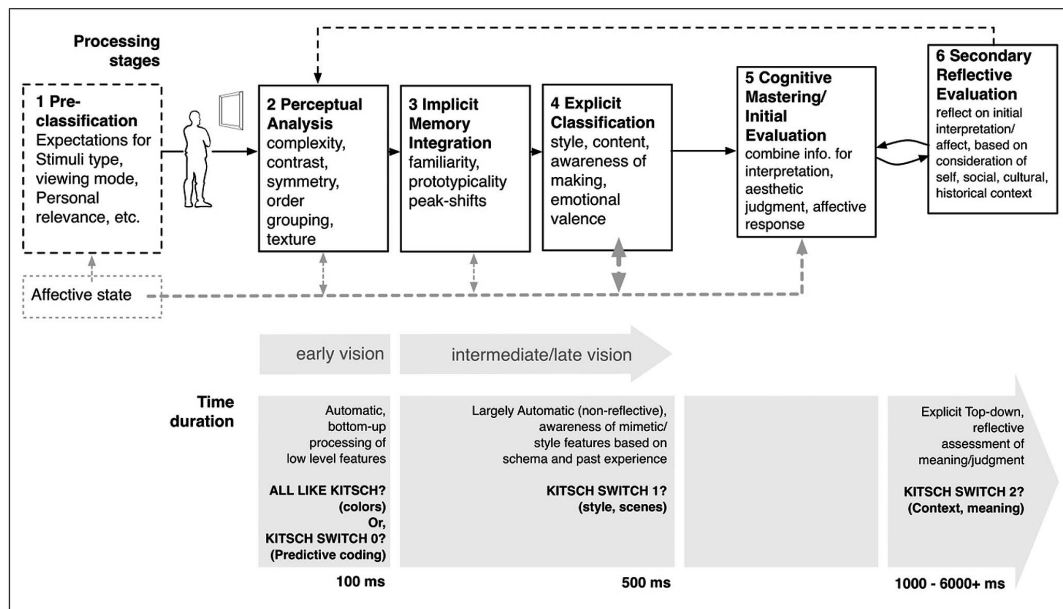


Figure 1. Processing stages of art viewing/response (based on model by Leder et al. (2004) and integrated with time durations and automatic (bottom-up) and more overt cognitive (top-down) stages from Pelowski, Markey, et al. (2017)).

Five hundred milliseconds was suggested as allowing more conscious perception of visual features as well as some identification of style or even artist (Goffaux et al., 2005; Torralba & Oliva, 2003). For example, an EEG study by Noguchi and Murota (2013) suggested that sensory and contextual (title, artist) integration can occur within 200–300 ms of seeing an artwork, while a MEG study by Munar et al. (2012) showed potentially related activation in the frontal brain regions after 300–400 ms. This time duration is thus argued to essentially allow an individual to fully move through the posited processing sequence of viewing stages above (Pelowski, Markey, et al., 2017), and obviously might involve the use of some top-down aspects (i.e., recruitment of stored schema or personal memory in order to make an identification of style/artist or even familiarity). However, this duration and judgment would also be argued to still be based primarily on a largely automatic processing of visual information without overt reflective judgments involving the work's wider context or social/personal importance.

On the other hand, 6,000 ms may provide a minimum time needed for individuals to both form an initial impression from bottom-up formal elements and to make a more overtly top-down assessment, modifying their first impressions based on conceptualization or connection with context, history, taste, culture, or personally relevant implications (Graf & Landwehr, 2015; Leder & Nadal, 2014; Meskin, Phelan, Moore, & Kieran, 2013; Pelowski, Markey, et al., 2017; Wilson, 1979). This is also supported by brain studies (e.g., Cela-Conde et al., 2004), which have shown that it may take at least 1–1.5+ s to connect more overt cognitive appraisal to a basic hedonic response. Geday, Kupers, and Gjedde (2007) also find that 3–4 s marks a major threshold for emotion beyond basic arousal and valence, which may lag visual appraisal and be tied to reflection or consideration of context. In empirical studies with art, it has also been the case that truncating viewing to less than around 6 s might create relatively more unstable judgments (e.g., see Bachmann & Vipper, 1983; Brieber et al., 2018; Graf & Landwehr, 2015; Leder & Nadal, 2014), with results (i.e., artwork liking) tied to the specific time durations, presumably due to limits on the type of information that can be assessed and the omission of more reflective, deliberative processing.

Last, we included a duration of 12,000 ms. This also followed present research. For example, museum studies have shown that individuals, left to freely view art, may tend to have a mean time of assessment of roughly 10+ s, where they may further attenuate or explore their initial responses (Brieber et al., 2018; Brieber, Nadal, Leder, & Rosenberg, 2014; see also Lengger, Fischmeister, Leder, & Bauer, 2007, for similar lab results).

It should of course be noted that the above times are only rough suggestions, with the present literature on time and art viewing still producing results that are rather inconsistent (see e.g., Augustin et al., 2008; or Verhaver, Wagemans, & Augustin, 2018 for arguments that some style or content processing may happen even more quickly). However, we hypothesized that, if there is a kitsch switch, it should coincide to a change in ratings (in this case focusing on visual artistic paintings) between the earliest 100 ms and the later durations. Specifically, viewers may generally like relatively more kitsch paintings at the 100 ms viewing duration, because they are designed to be immediately appealing. However, with longer durations some individuals, especially experts, may

adjust their initial response and show lowered liking. Further, these changes may require either reflections and placement of the art within wider context and personal taste—what we would argue to be “overtly top-down” processing as labeled in Figure 1—and coinciding with 6 or even 12 s (Kitsch Switch 2, Figure 1).

Alternatively, basic awareness of style and content may be enough to evoke such changes (Kitsch Switch 1, Figure 1). This modulation, tied to 500 ms, would suggest that a switching may be primarily tied to utilization of past schema or training, even without need to overtly reflect on wider implications or importance. These hypotheses can also be tested against an alternative hypothesis that art taste and perception could be modulated from the very beginning (Kitsch Switch 0 in Figure 1). As suggested by discussions of predictive coding, certain framing or expectations (i.e., from the preclassification stage) could drive our attention or assessments, representing an almost continuous aspect of art experience. Thus, we would expect a rating difference between kitsch and other art examples for certain viewers even at 100 ms, with, or especially without, further change across the subsequent time periods.

Note as well, despite theoretical arguments for key viewing-time-based differences, actual empirical studies, specifically those considering behavioral impact on ratings, are quite sparse. Thus, this study served as also a general assessment of such previous models and arguments that ratings might change especially in the initial interaction with art.

Three Artwork Types and the 21st Century's Most Notorious Kitsch Artist

To assess this integration within processing experience, we used paintings by three artists, led by one particularly notorious creator of kitsch. Thomas Kinkade (1958–2012) was, until his death in 2012, arguably one of the world's most successful contemporary artists. Known for a loose, painterly style using idyllic nature scenes, cottages with illuminated windows and fall or summer foliage, or pictures of quaint main streets—and in tandem with a very complex multilevel sales and marketing mechanism—he was known as “America's most collected living artist” with some type of work (canvas, print, mug, etc.) owned in 5% of American households (Perl, 2011). The reasons given for his success—familiar, visually pleasing scenes with highly saturated/pastel colors—align with the visual or semantic features tied to basic visual art processing above. On the other hand, his art is almost universally hated by critics. Described as “awful” (Heath & Potter, 2005, p. 124), and dismissed as a “kitsch crime against aesthetics” (Burkeman, 2006), the reasons for this are often tied to a more context-based response to the same factors above. These include an unironic focus on settings—“Hollyhocked and morning-gloried cottage porn” (Miller, 2002; see also Perl, 2011)—offering non-threatening images, which are too easy and too fluent. Critics also note overly bright colors, exaggeration of reflections, which are too affecting or are melodramatic (Didion, 2003).

Kinkade has also recently been considered in a study that suggested—although did not directly assess—evidence for such a top down modulation of response. Meskin et al. (2013, following an even earlier exploratory study by Cutting (2003) showed participants a series of Kinkade paintings (deemed by the authors as “bad art”). These were shown one, two, or five times (exact same

painting repeated) over seven weeks as part of an unrelated lecture. The Kinkade set was accompanied by a set of landscape paintings with similar themes but which were more appreciated by critics (“pretty good art” by John Everett Millais) and shown in a similar single or repeated fashion. In an initial control rating, paintings by Kinkade had higher mean liking ratings than paintings by Millais. After one repeat viewing, the same Kinkade paintings showed decreased liking and the Millais works showed increased liking. After five viewings, both sets showed decreased liking from the previous exposures. The authors explained this result as suggesting that the Kinkades—although again initially more pleasing—did not reward repeat interactions. Potentially, this may have been due to the above-discussed evolving taste or conceptualization factors; it was also argued to be contrary to a typical finding that repeated exposures can often increase liking ratings due to increased fluency (however, see, e.g., Bornstein, 1989 for a discussion of similar contrary effects with art in general). Thus, in the present study, using a similar set of paintings (although with some key adjustments regarding methodological issues in the Meskin et al. paper, see footnote)⁷ we considered more specifically potential changes as they may occur within the individual viewing sequence.

For a direct comparison to the Kinkades, we used a set of landscape paintings by John Constable. Constable (1776–1837), an English Romantic painter, also produced realistic paintings of cottages and forest scenes and thus represented a very similar thematic and, to some extent, stylistic approach as the Kinkades. However, for presumably complex art historical and conceptual reasons, as well as potentially higher mastery in brushwork, his paintings are also critically respected. These were therefore expected—unlike the Kinkades and following critical argument—to better reward longer viewing durations, as individuals might come to appreciate the mastery of brushwork or bucolic scenery. Thus, they were expected to represent a neutral set of paintings that may or may not become progressively more liked.

Finally, we included abstract paintings by Gerhard Richter (born 1932). These were selected because of a previously documented general dislike by lay viewers of abstract art (e.g., Pelowski, Gerger, Chetouani, Markey, & Leder, 2017, for review), potentially again involving such a point of inflection between artwork-derived and context-derived processing. Notably, abstract art may demand interpretation via top-down connections to context and medium and was specifically suggested as the antithesis of kitsch by Greenberg (1939). Thus, we hypothesized that these paintings may elicit an inverse pattern from the kitsch examples, with a general increase in liking among experts. Richter is extremely well respected among those with knowledge of contemporary art. He was the recipient of the Golden Lion of the 47th Venice Biennale in 1997 and has set several auction records for a living artist. On the other hand, if individuals do not appreciate or use such top-down factors in their experience, we might expect a liking decrease. For this study, we selected a very uniform set from his abstract works with only differing colors spread across the canvas. These also contained similar bright colors to those of the Kinkades (as if one had wiped a squeegee across a wet Kinkade painting) allowing for a more controlled consideration of basic color-based response.

Study 1

Method

This project was designed from the beginning as a general comparison of relatively more lay and expert art viewers and viewing durations. However, the two (lay and expert) samples were collected in separate studies and thus will be discussed in sequence before a final combined analysis. Study 1 involved an initial assessment of lay art viewers’ liking ratings and varied between subjects for exposure times to the art. This was followed by Study 2 using the same methodology but with art experts—the group expected to show the most pronounced changes in ratings for kitsch.

Participants. Study 1 was administered to a final sample of 154 participants. All were psychology students at the University of Vienna and completed the study for course credit. All were also lay viewers with no special training in Fine Arts or Art History (as assessed in posttest survey, see below). The participants were divided into four subgroups, based on the time allowed for viewing the art: 100 ms ($n = 40$; $M_{\text{age}} = 21.9$, $SD = 3.8$; 34 female); 500 ms ($n = 38$; $M_{\text{age}} = 22.3$, $SD = 3.9$; 25 female); 6,000 ms ($n = 38$; $M_{\text{age}} = 21.6$, $SD = 4.9$; 27 female); 12,000 ms ($n = 38$; $M_{\text{age}} = 20.4$, $SD = 2.1$; 30 female; no sig. demographic differences between groups). The final sample was reduced from an original 160 participants, with six individuals excluded due to not completing all study components. The study was approved by the Ethics Committee of the University of Vienna.

Stimuli. The stimuli set consisted of 75 digitized paintings, divided equally into the three types (25 each; see Supplementary Table S1 in the online supplemental material for full list): (a) a set of Thomas Kinkade (kitsch) paintings based on the original set used in Meskin et al. (2013) as listed in their paper. This set was however revised to provide a more uniform collection, retaining only paintings that depicted similar scenes of cottages with lake/river and Summer/Autumn foliage or flowers, and notably removing paintings that depicted seascapes or identifiable landmarks/characters (see also Footnote 6). All artworks in this revised set were also without watermarks. (b) We also included a set of landscape paintings by John Constable and (c) a set of abstract paintings by Gerhard Richter. All stimuli were displayed on a monitor (19" Iiyama ProLite B1906S, 1280x1024, 60Hz), centered on the screen with a white background, and with the longest painting dimension set at 1,000 pixels.

⁷ Note that the Meskin et al. (2013) study, in our opinion, included some methodology decisions that could have conflated the results. Notably, although the majority of the selected artworks were “typical” Kinkade paintings of nostalgic house and forest scenes, they were intermingled with paintings depicting the ocean, religious figures such as Jesus, Santa Claus, Disney characters, or well-known landmarks/cities (London, Paris/Eiffel tower). In addition, the control stimuli, which were generally not liked by the viewers, used a set of paintings with a quite dark and gloomy color palette. This alone may have driven responses (although a case can be made that the potential negative emotional valence itself distinguishes this art from kitsch; see e.g., Ortlieb & Carbon, 2018). Artworks were also displayed to individual participants while in a large group of other participants and shown as PowerPoint files via projector at the front of the classroom, raising issues with social pressure and the uniformity of presentation. These issues are further taken into account in the Methods below.

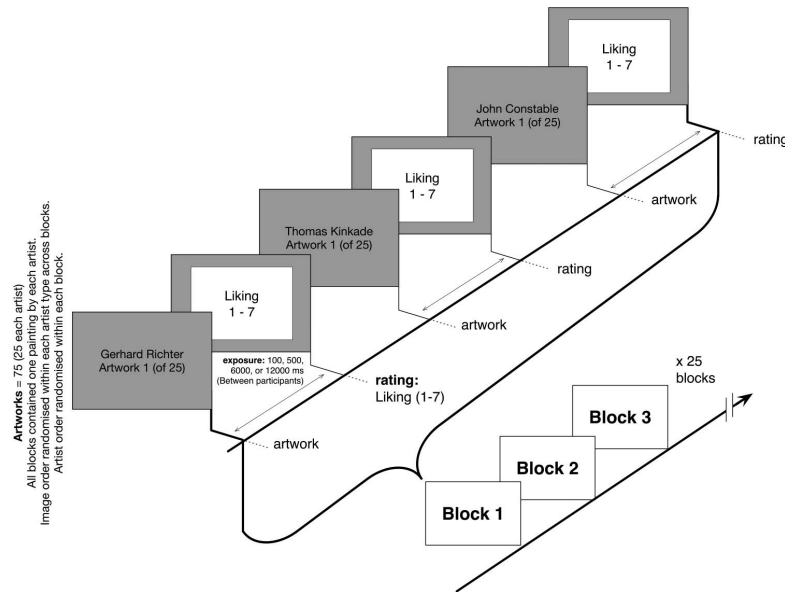


Figure 2. Study design. Exposure time (100, 500, 6000, 12,000 ms) counterbalanced between participants.

Procedure. Participants were first administered a consent form explaining that they would be viewing some works of art and would be asked to make a rating for liking using a 7-point scale ranging from 1 (*not at all*) to 7 (*very much*). The study was designed with E-Prime software (2.0, Psychology Software Tools, Inc.). Procedure consisted of a between-subjects design, in which participants were divided based on time of artwork exposure (100, 500, 6,000, or 12,000 ms). Participants from each group viewed the full set of paintings in a series of 25 blocks (see Figure 2). Each block (see Figure 2) consisted of one example of each of the three painting types (i.e., Kinkade > Constable > Richter), with participants presented with a cue cross, followed by one painting, shown individually for the chosen duration, and then followed immediately by a screen presenting the rating scale answered via keyboard. Following a participant's keypress, the procedure was repeated for the next painting (by one of the two remaining artists) in the block.

After completing one block of three paintings (one by each of the three artists), the procedure was repeated with a new block of three additional paintings until all 75 had been viewed and rated once. The selection of specific paintings by each artist for each block was randomized between participants, as was the order of artists within each block (e.g., K-C-R; C-R-K; etc.). This paradigm insured that each participant would see artworks of each type with a generally consistent duration between repeated exposures to the same artist (minimum of zero to a maximum of four paintings) while at the same time avoiding a monotonous 1-2-3, 1-2-3 structure in which participants could always anticipate the next painting type.⁸ Before the main study, participants were given three practice trials to confirm their comfort with the procedure

and were instructed to record their ratings as quickly as possible. Response times were recorded by the E-Prime software.

After completing the main study, participants answered a number of questions (administered via online survey, Limesurvey, v. 1.92, Limesurvey.org; accessed via the same lab computer) to add further information about the basis of their appraisal decisions. First, participants were asked to think back to how they had made the liking ratings of the art and to consider 17 factors (e.g., "colors/contrast," "technical quality"; see Table 5) for how much emphasis they had been given when making decisions on a 7-point scale, ranging from 1 (*not at all*) to 7 (*very much*). This was based on a previous study (Pelowski et al., 2017), which showed this to be a productive means of providing more context regarding the approach individuals take with typical scale-based art appraisals. The included factors could also be generally divided into those relatively more related to bottom-up or top-down processing.

The posttest also included a number of questions to assess expertise or art interest (e.g., frequency of visits to museums and galleries, or reading of art literature, from Leder, Gerger, Brieber, & Schwarz, 2014), as well as previous art training (following Chatterjee, Widick, Sternschein, Smith, & Bromberger, 2010), art style preferences (i.e., relative liking for kitsch, abstract, or repre-

⁸ This blocked design was also selected to test an unrelated research question involving general declines or increases in liking over the repeated viewing of several paintings by the same artist or of the same style (e.g., assessing a generally positive repeat exposure or a negative 'museum fatigue' effect). However, this had no bearing on the present results and will not be discussed here.

sentational paintings, etc.), and basic attitudes about art or relative art involvement.

Results

Inspection of the data from the included participants revealed no issues that might stem from monotonous (e.g., using all same ratings) or inattentive (answering faster than 500 ms) answering. Comparison of the individual ratings for each artwork within the three categories also revealed high consistency, with mean ratings of all paintings falling within $\frac{1}{2}$ standard deviation of the group means. Thus, all participants and artworks were kept for analysis. Posttest interviews revealed that no participants were aware of our block-wise repeated exposure paradigm, but rather were only cognizant that they had been shown artworks by a selection of a few artists. As expected, results of the posttest survey regarding art interest and training suggested that the participants were lay viewers with low professed art knowledge, almost no experience taking art history ($M = 0.9$, $SD = 1.4$, score of 0 = "none" on scale) or art making courses ($M = 0.9$, $SD = 1.6$), and rarely visited museums ($M = 3.4$, "two or less times per year").

Descriptive statistics of ratings and response times. The liking ratings of the three types of art, comparing between the four exposure times, are shown in Figure 3. Overall, participants tended to prefer the Kinkades (M liking rating = 4.11, $SD = 1.90$), followed by the Constables ($M = 3.57$, $SD = 1.54$), and finally the Richter abstract paintings ($M = 3.17$, $SD = 1.77$). As a general consideration of agreement regarding ratings across individual artworks but within the artist types, we derived McDonald's omega scores (ω_h , see Zinbarg, Revelle, Yovel, & Li, 2005) for each artist at each time duration (these are reported in Table A1 of the Appendix). This suggested that the participants were quite

consistent in their ratings for all types of art ($\omega_h = .6$ to $.9+$). They were also generally consistent at all viewing durations.

Response times for recording liking ratings (following the disappearance of the paintings and appearance of the scale on screen) are also shown in Figure 3 (bottom). Repeated measures analysis of variance (ANOVA), Artist (3) $\times$ Viewing Duration (4), with response time as a dependent variable revealed a significant main effect for viewing duration, $F(3, 150) = 22.70$, $p < .001$, $\eta_p^2 = .88$, with a notable decrease in response times between the shorter 100 ($M = 1,902$ ms, $SD = 628$) and 500 ms ($M = 1,654$, $SD = 711$) durations and the longer 6 ($M = 1,162$, $SD = 499$) and 12 s ($M = 1,014$, $SD = 430$). No significant effects were found for artist, $F(2, 300) = 2.00$, $p = .14$, $\eta_p^2 = .01$, or Artist $\times$ Duration, $F(6, 300) = 1.12$, $p = .35$, $\eta_p^2 = .02$.

Impact of viewing duration on liking ratings—lay viewers.

To assess our main question regarding the impact of exposure duration on the ratings of the three painting types, we conducted a linear mixed model (LMM), using the *lmer()* function of the 'lme4' package (Version: 1.1–18-1, Bates, Mächler, Bolker, & Walker, 2015) in combination with the 'lmerTest' package (Version: 3.0–1, Kuznetsova, Brockhoff, & Christensen, 2017) in R (Version: 3.5.1, R Core Team, 2014). "Liking" was set as the dependent variable with the two fixed factors of time and artist (both treatment coded [dummy coded]). Given our hypotheses, we compared time with 100 ms (baseline) against the progressively longer durations. To fully evaluate all comparisons, we reran the model with the relevelled factor artist twice. We also included random intercepts for participants and for artworks, and random slopes of artist for participants. The random intercepts allowed us to account for the natural variation that can occur between participants and images. The random slopes allowed us to account for variations in rating differences between artists among the participants. The random slope of time for artworks was omitted due to convergence issues.

The main results of the above analyses are reported in Table 1. Because of our hypotheses regarding which factor-level combinations of time and artist should differ from each other, we discuss the results using an uncorrected alpha ($p \leq .05$), and in fact follow the original arguments of Fisher in exploratory hypothesis testing in regards to not putting overemphasis on significance thresholds but rather considering effect sizes in light of the overall patterns of data (e.g., Gigerenzer, 2004). However, we also include information regarding adjusted alpha thresholds following Bonferroni correction, reported in the table notes with full explanation, for the interested reader.

As can be seen in the table, when considering potential viewing duration-based changes, the only significant effects involved a decrease in liking ratings from 100ms to 6s and from 100 ms to 12 s for paintings by Constable. On the other hand, both Kinkade and Richter did not show any changes. We also detected significant differences between the ratings made for the different artists within specific viewing durations. Notably, this included 100 ms, where differences were found between Kinkade (highest ratings) and Constable, and Richter, respectively (lowest). Similar trends or significant differences were found for all other viewing times.

Art ratings between viewing durations when separating lay viewers based on lower and higher art interest. The relatively stability of liking ratings for the kitsch paintings did not appear to provide evidence of a viewing-duration-based switch in evalua-

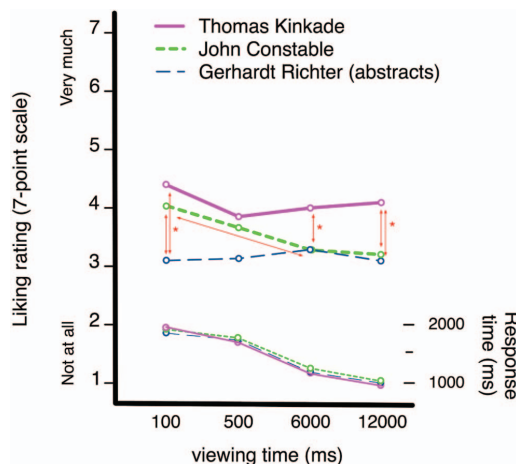


Figure 3. Liking evaluations and response time of Kinkade (kitsch), Constable, and Richter (abstract) paintings by lay (psychology student) participants following different viewing durations. See the online article for the color version of this figure.

Table 1
Linear Mixed Model (LMM) of Liking Ratings for Paintings Comparing Between Viewing Durations (100, 500, 6,000, 12,000 ms, Varied Between Participants) and Artists (Kinkade, Constable, Richter, Varied Within Participants) for Lay (Psychology Student) Viewers

Variable	Estimate	SE	<i>t</i>	<i>p</i>
Within artwork, between exposure times				
Kinkade				
500 ms–100 ms	–0.52	0.36	–1.44	.15
6 s–100 ms	–0.38	0.36	–1.06	.29
12 s–100 ms	–0.29	0.36	–0.82	.41
Constable				
500 ms–100 ms	–0.36	0.23	–1.59	.11
6 s–100 ms	–0.74	0.23	–3.28	.001*
12 s–100 ms	–0.83	0.23	–3.70	.0003*
Richter				
500 ms–100 ms	0.03	0.30	0.11	.91
6 s–100 ms	0.20	0.30	0.66	.51
12 s–100 ms	–0.02	0.30	–0.06	.95
Within exposure times, between artworks				
100 ms				
Artist C–K	–0.35	0.26	–1.36	.17
Artist R–K	–1.29	0.40	–3.22	.002*
Artist R–C	–0.93	0.129	–3.21	.002*
500 ms				
Artist C–K	–0.20	0.27	–0.74	.46
Artist R–K	–0.74	0.41	–1.80	.07
Artist R–C	–0.54	0.30	–1.82	.07
6 s				
Artist C–K	–0.71	0.27	–2.69	.008*
Artist R–K	–0.70	0.41	–1.72	.09
Artist R–C	0.01	0.30	0.04	.97
12 s				
Artist C–K	–0.89	0.27	–3.37	.001*
Artist R–K	–1.01	0.41	–2.47	.01*
Artist R–C	–0.11	0.30	–0.39	.70

Note. Results based on LMM with two factors Time and Artist and “liking” ratings (7-point scale) as dependent variable; treatment coded contrast for time, 100 versus 500, and 12,000 ms, and for each artist. Random intercepts for participants and images and random slopes of artist for participants. Results based on 154 lay art viewers (psychology students). Bold text = significant following a Bonferroni correction, adjusted critical *p*-value = .0024 (21 pairwise comparisons; nine comparisons for time, comparing each longer duration against 100 ms separately for each of the three artists; 12 comparisons for ratings between artists at each of the four time points).

* *p* < .05 (uncorrected for multiple comparison).

tion. This finding was also in tandem with a pronounced difference in relative liking between artists, which was maintained from even 100 ms. The lack of time-based changes with kitsch, although probably not the latter artist-level differences, could also be interpreted, however, to fit the overall hypothesis that such modulations may not be discovered with lay viewers as it was this group who was expected to continue liking kitsch even with longer times afforded to viewing. Therefore, before moving to the study of art experts, we conducted a post hoc assessment—taking into account the relative art interest of individuals—to see if we might detect at least some indication for time-related differences.

To conduct this analysis, an “art interest” score was calculated for each individual based on their averaged answers to the five questions: “I am comfortable looking at and discussing art,” “I think I have a good knowledge about art,” “art is important,” “I enjoy viewing art,” and “I am interested in art.” These had previously been shown to provide a good general assessment of individuals’ relative art involvement and are typically answered more positively by experts (Pelowski, 2015;

Pelowski, Markey, et al., 2017). The mean score was 4.6 (*SD* = 1.3, *Mdn* = 4.8, range = 1.4 to 7; intercorrelations between items *r* = .40 to .70, all *ps* < .001).

Because of the rather large range in participant interest scores (again 1.4 to 7 on a 7-point scale), and to facilitate a general descriptive assessment, we first conducted a median split of our sample. This produced two relatively equally sized groups with lower interest (*N* = 74, *M*_{score} = 3.48, *SD* = 0.87) and higher interest (*N* = 80, *M*_{score} = 5.58, *SD* = 0.55). A *t* test comparing the interest score of both groups indeed revealed significant difference, *t*(152) = 18.07, *p* < .001, *g* = 2.91 (note also, the groups were relatively evenly split within the viewing times as well).

The relative liking ratings, differentiating between the Three Artists × the Four Viewing Durations and for the higher and lower art interest groups of lay viewers are shown in Figure 4. As can be seen, this split did appear to provide some indication of a notable divergence in ratings between the groups over the progressively longer viewing durations, and especially within the Kinkade and the Richter art. The Kinkades elicited a general consensus in

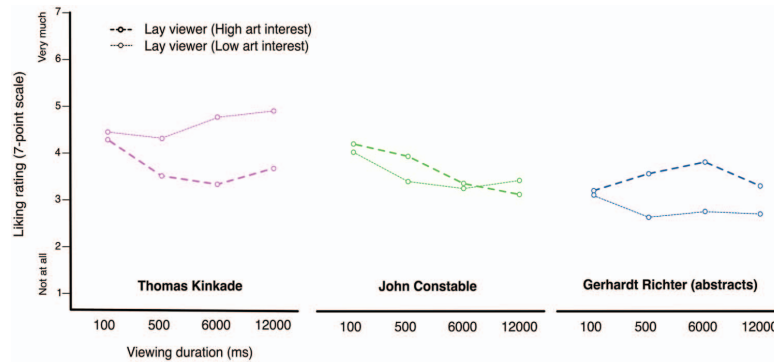


Figure 4. Liking evaluations and response time of Kinkade (kitsch), Constable, and Richter (Abstract) paintings by lay (Psychology student) participants following different viewing durations and when divided between relatively higher and lower art interest (median split). See the online article for the color version of this figure.

ratings among all viewers at 100 ms and then showed a decrease among those with relatively higher interest (bottoming at 6 s) and a general rise over time among those with low art interest. The Richter showed the opposite pattern of changes, especially when comparing between 100 ms and 6 s. On the other hand, the Constable tended to show a decrease across all viewers. Interestingly, only those who had low art interest, and who were rating the Kinkades, tended to use the positive “like” side of the rating scale. In all other cases of artist and rater, the average assessments were on the negative half.

To statistically assess the relationships, we conducted another set of LMMs again with “liking” as the dependent variable and considering the two factors of time (treatment coded contrast comparing against 100 ms) and art interest. Note, due to our anticipation of assessing the entire sample of both lay and expert participants in one final model following Study 2, below, art interest was coded as a continuous variable. Models were again conducted separately for each artist (with random intercepts for participants and for images, random slopes of artist for participants, and random slopes of art interest for images). Results of the main effects and interactions with art interest are shown in Table 2.

Although again, this did not detect a significant interaction of art interest and time within evaluations of particular artists (see top part, Table 2), this did suggest a trend for the divergence in ratings as participants with differing interest levels had the opportunity to view the paintings for longer durations. Notably, although no differences were found between participants of differing interest scores for any of the three artists at 100 ms, differences were detected for ratings of both Kinkades and Richters at 6 s, and significant or nearly significant differences were detected for the same artists at 500 ms. Ratings at 12 s also showed differences for Kinkades and Constables. (Note, a similar series of models with art interest using the median split and a contrast coding returned similar results. This is provided in the online supplementary materials).

Study 2—Art Experts

To further consider the above patterns, we conducted a second study with individuals who identified as having even higher expertise in art (graduate level art history students) and who were expected to show a more pronounced change in ratings between the short and longer viewing durations.

Method

Participants for Study 2 consisted of 90 students at the Master level or higher in Art History at the University of Vienna ($M_{\text{age}} = 23.1$, $SD = 3.4$; 66 female). The procedure and stimuli were the same as in Study 1, with participants split equally into groups ($n = 30$ each), for each time durations of 100, 500, and 6,000 ms. Because of a relatively higher difficulty in finding art historians to participate, and based on our hypotheses and Study 1 findings, we omitted the 12,000-ms duration.

Results: Study 2

Once again, all participants completed all of the portions of the study. As expected, the historians showed a relatively higher interest in art, with mean answers to the composite “art interest” score of 6.3 on the 7-point scale ($SD = 0.7$, $Mdn = 6.4$, range 4.2 to 7), or roughly $\frac{1}{2}$ point higher than even the highest art interest individuals in Study 1. Similar results were also found for the scales relating to visits to art museums and training.

Reaction times and rating consistency. Reaction times (Figure 5, bottom) were again analyzed via mixed ANOVA, first for the expert group exclusively (Viewing Duration $\times$ Artist). As in Study 1, this revealed a significant difference for viewing duration, $F(2, 87) = 12.24$, $p < .001$, $\eta_p^2 = .22$, and for artist, $F(1.86, 161.72) = 10.40$, $p < .001$, $\eta_p^2 = .11$. Similar to the previous study the interaction between Artist $\times$ Duration was not significant ($p = .63$). Mixed ANOVA comparing between lay and expert Groups $\times$ Artists $\times$ Viewing Durations (100, 500, and 6,000 ms durations

Table 2

Linear Mixed Model (LMM) of Liking Ratings of Paintings Comparing Between Viewing Durations (100, 500, 6000, 12,000 ms, Varied Between Participants) and Art Types (Kinkade, Constable, Richter, Within Participant) for Lay (Psychology Student) Viewers, in Regards to Viewer art Interest (Continuous Variable)

Variable	Estimate	SE	<i>t</i> value	Pr(> <i>t</i>)
Within artwork, between exposure times and art interest				
Kinkade				
Art Interest × Time (500–100 ms)	−0.15	0.27	−0.56	.58
Art Interest × Time (6,000–100 ms)	−0.34	0.29	−1.17	.24
Art Interest × Time (12,000–100 ms)	−0.42	0.29	−1.45	.15
Constable				
Art Interest × Time (500–100 ms)	0.09	0.18	0.52	.61
Art Interest × Time (6,000–100 ms)	−0.13	0.19	−0.66	.51
Art Interest × Time (12,000–100 ms)	−0.25	0.19	−1.33	.18
Richter				
Art Interest × Time (500–100 ms)	0.29	0.23	1.28	.20
Art Interest × Time (6,000–100 ms)	0.35	0.25	1.39	.17
Art Interest × Time (12,000–100 ms)	0.08	0.25	0.31	.76
Within exposure times and artworks, between art interest				
100 ms				
Kinkade: Art interest (continuous)	−0.18	0.18	−1.02	.31
Constable: Art interest (continuous)	0.03	0.16	0.21	.84
Richter: Art interest (continuous)	0.11	0.20	0.55	.59
500 ms				
Kinkade: Art interest (continuous)	−0.33	0.17	−1.95	.06 [‡]
Constable: Art interest (continuous)	0.12	0.10	1.29	.20
Richter: Art interest (continuous)	0.40	0.14	2.96	.005 ^{***}
6 s				
Kinkade: Art interest (continuous)	−0.53	0.21	−2.47	.02 [*]
Constable: Art interest (continuous)	−0.10	0.13	−0.73	.47
Richter: Art interest (continuous)	0.45	0.17	2.69	.01 [*]
12 s				
Kinkade: Art interest (continuous)	−0.60	0.22	−2.76	.009 [*]
Constable: Art interest (continuous)	−0.22	0.13	−1.70	.09 [‡]
Richter: Art interest (continuous)	0.18	0.17	1.10	.28
Slope of art interest for each artist				
Kinkade: Art interest (continuous)	−0.18	0.21	−0.89	.38
Constable: Art interest (continuous)	0.03	0.14	0.24	.81
Richter: Art interest (continuous)	0.11	0.18	0.61	.54

Note. Results based on LMM with two factors Time and Artist, “art interest” entered as continuous variable (7-point scale), grand mean centered, and “liking” ratings (7-point scale) as dependent variable; treatment coded contrast for time, 100 versus 500 and 6,000 ms, and for each artist. Random intercepts for participants and images and random slopes of artist for participants and of art interest for images. Results based on 154 lay art viewers (psychology students). “Art interest” calculated via average score of five questions (7-point scales): “I am comfortable looking at and discussing art,” “I think I have a good knowledge about art,” “art is important,” “I enjoy viewing art,” and “I am interested in art.” Bold text = significant following a Bonferroni correction, adjusted critical *p* value = .0024 (21 pairwise comparisons; nine comparisons for the interaction between art interest and time; nine comparisons for ratings between art interest for each artist and time point, three comparisons for art interest for each of the artists).

‡ *p* < .10. * *p* < .05. ** *p* < .01 (uncorrected for multiple comparison).

only) showed no significant main effect for lay/art historian groups, $F(1, 200) = 3.59$, $p = .06$, $\eta_p^2 = .02$, as well as no significant Group × Viewing Duration, $F(2, 200) = 1.10$, $p = .33$, $\eta_p^2 = .01$, or three-way interactions, $F(4, 400) = 0.16$, $p = .96$, $\eta_p^2 = .002$. We did detect a significant Group × Artists interaction, $F(2, 400) = 7.88$, $p < .001$, $\eta_p^2 = .04$, with post hoc pairwise comparisons suggesting that this was driven by a significant difference between the RTs for the Kinkade ratings (relatively faster) versus both Richters and Constables within the Art expert group ($ps < .001$) and by a difference between the experts’ (generally slower) ratings for Constables ($p = .01$) and Richters ($p = .02$) when compared to the lay viewers (all other pairwise comparisons, n.s.).

Assessment of ratings consistency across the differing paintings within specific artist sets (see Appendix Table A1) also showed similar, generally high levels of consistency as that of the lay participants.

Liking ratings between viewing durations—art experts. Figure 5 displays the mean liking ratings for the expert viewer population, again dividing between the three painting types and between the 100-, 500-, and 6,000-ms viewing durations. For comparison, this also displays the results from Study 1 (with the lay participants split the into higher/lower art interest groups). As with the relatively art interested lay viewers in Study 1, the experts’ results did suggest a change in the kitsch art ratings, with a generally linear decrease in liking from short to longer durations.

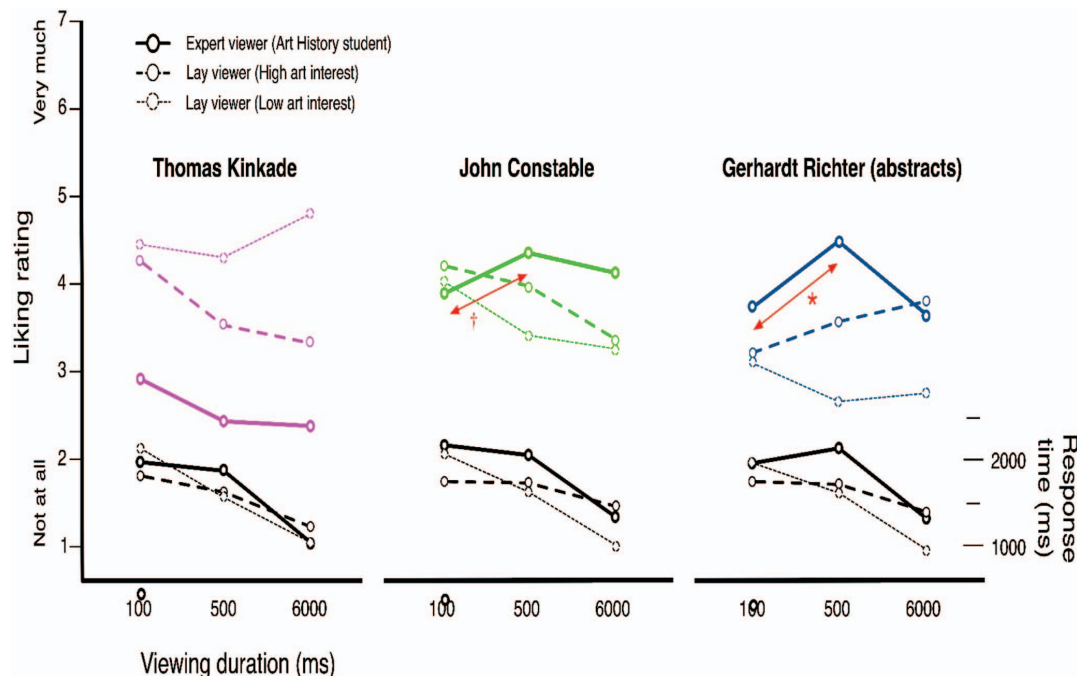


Figure 5. Liking evaluations of Kinkade (kitsch), Constable, and Richter (abstract) paintings by expert (art history master student) following different viewing durations. Lay (psychology student) participants with relatively higher and lower art interest (Median split) also shown for comparative purposes. * indicates significant difference, $p < .05$, for experts based on linear mixed models, see Table 3. † indicates trend $p < .10$. See the online article for the color version of this figure.

On the other hand, we found some positive change with the Richters, especially at 500 ms, and—unlike both higher and lower interest lay viewers—a general increase in ratings for the Constables. Note also that the Kinkade liking ratings, at all time durations, were markedly lower for experts than for those by all lay viewers and for both other types of art.

To assess these relationships, we once again conducted a series of LMMs looking at the expert group (painting “liking” = dependent variable; fixed factors of time and artist, both treatment coded; random intercepts for participants and artworks; random slopes of artist for participants). Results are shown in Table 3. Although the trends did follow our hypothesized directions, we largely did not detect significant changes as a function of time, even for the art experts. One significant change was found for the Richter paintings between 100 and 500 ms, with the latter time leading to higher ratings. Similarly, the Constables showed a trend for higher ratings between 100 and 500 ms ($p = .10$). On the other hand, significant main effects regarding differences between the ratings of the different artists’ paintings were detected for all three artists: At each time point Kinkade paintings were liked less than Constables. In addition, at 500 ms, Kinkade paintings were liked less than Richters (at 6,000 ms this differences came close to our adjusted significance criterion).

Combined analysis of lay and expert viewers. Last, all participants from both studies were combined into a joint analysis. For this, we again used LMMs considering the factors of time (treatment coded contrast comparing against 100 ms) and artist, as well as art interest (coded as a continuous variable, grand mean centered; with random intercepts for participants and images and random slopes of artist for participants and of art interest for images). As can be seen in Table 4, which reports the main results, we again found some suggestion for a difference in ratings between the times, as these interacted with differing relative art interest. Notably, time-related changes involved Constable paintings, which showed a significant interaction between art interest and time in the comparison between 100 and 500 ms conditions. The Richter paintings showed a similar, albeit not significant, trend ($p = .07$) also for the interaction between art interest and the comparison between 100 and 500 ms. The Kinkades, interestingly, did not show significant interactions or notable trends here. Although somewhat a stretch of the potential inferences, it can also be noted that the effect sizes did tend to increase especially between 100 and the later conditions, peaking at 500 ms. At the same time, once again, significant differences were also found between the artists at each time point.

Table 3
*Linear Mixed Model (LMM) of Liking Ratings for Artworks
 Comparing Between Viewing Durations (100, 500, 6,000,
 Between Participants) and Artists (Kinkade, Constable, Richter,
 Within Participants) for "Art Expert" (Master Level Art History
 Student) Viewers*

Variable	Estimate	SE	t/z value	Pr(> t)
Within artwork, between exposure times				
Kinkade				
500 ms–100 ms	–0.45	0.38	–1.18	.24
6 s–100 ms	–0.52	0.38	–1.36	.18
Constable				
500 ms–100 ms	0.47	0.29	1.65	.10 [‡]
6 s–100 ms	0.25	0.29	0.89	.38
Richter				
500 ms–100 ms	0.87	0.37	2.39	.02 [*]
6 s–100 ms	–0.15	0.37	–0.40	.69
Within exposure times, between artworks				
100 ms				
Artist C–K	1.00	0.28	3.61	.0004[*]
Artist R–K	0.88	0.42	2.08	.04 [*]
Artist R–C	–0.12	0.35	–0.35	.73
500 ms				
Artist C–K	1.92	0.28	6.95	<.0001[*]
Artist R–K	2.20	0.42	5.21	<.0001[*]
Artist R–C	0.28	0.35	0.81	.42
6 s				
Artist C–K	1.77	0.28	6.41	<.0001[*]
Artist R–K	1.25	0.42	2.96	.004 [*]
Artist R–C	–0.52	0.35	–1.50	.14

Note. Results based on LMM with two factors Time and Artist and “liking” ratings (7-point scale) as dependent variable; treatment coded contrast for time, 100 versus 500 and 6,000 ms, and for each artist. Random intercepts for participants and images and random slopes of artist for participants. Results based on 90 expert art viewers (art history students). Bold text = significant following a Bonferroni correction, adjusted critical p -value = .0033 (15 pairwise comparisons; six comparisons for time, comparing each longer duration against 100 ms separately for each of the three artists; nine comparisons for ratings between artists at each of the three time points).

[‡] $p < .10$. ^{*} $p < .05$ (uncorrected for multiple comparison).

Artwork ratings, viewing duration, and strategies for appraising the art. Finally, to provide more context to the above findings, we considered the results of the posttest survey in which individuals were provided a list of potential strategies or factors that might be used in making their ratings of the art. This was exploratory, however was thought to potentially provide more information regarding the lay/expert differences and the relative emphasis on top-down or bottom-up approaches.

The strategies are shown in Table 5 with their correlation to the liking ratings of the three types of art (combining viewers from all time durations). The table also displays means for the lay and expert samples, as well as results of t test comparisons between these groups (Bonferroni adjusted for familywise error with specific adjusted alpha information provided in the table notes). In general, those who liked the Kinkades gave more emphasis to an artwork’s beauty, content, ability to make one feel safe or comfortable, and to evoke nostalgia. With the exception of the last item, all of these also showed significant differences in usage between lays (higher mean) and experts (lower). On the other hand, not liking Kinkade correlated to focusing more on style (also

showing significant difference in usage between lay/experts), materials, visual interest/novelty, ability to evoke challenge, and thinking of expert opinions. Note also, in general, these same factors showed the opposite relation to liking the abstract (Richter) art, whereas the relation to Constables tended to be similar to that of the Kinkades.

The usage of these factors (shown as mean responses regarding their importance on the provided scale) between the expert/lay individuals and, within these groups, for individuals answering following a 100 or 6,000 ms viewing duration, are shown in Figure 6. The strategies are further organized (based on the opinion of the authors) into those expected to pertain more to bottom-up and to top-down, later assessments. The time durations were selected based on our theoretical discussion and because these tended to show the most salient differences for both the lay and the expert groups.

Keeping in mind that these are only descriptive suggestions, there did appear to be some evidence for a hypothesized switch in emphasis. Several bottom-up features showed relatively higher importance at 100 compared to 6,000 ms for both lay and expert groups. This includes visual interest, style, and beauty/visual appeal. At the same time, at 6,000 versus 100 ms, several top-down factors (content, technical quality, evoked idea of deeper meaning, comfort/discomfort, evoked specific emotions, and thoughts of experts’ opinions) showed relatively higher emphasis, however primarily only among the expert viewers. These factors did not show the same increase, or in some cases showed opposite trends, among lay viewers, even at the longer viewing duration.

Conclusion

This study investigated the potential for changes in our art assessments based on the relative emphasis of bottom-up, artwork-derived visual features and—potentially coming into play later within a processing experience—more overt top-down consideration based on the personal tastes, history, or other context of the viewer. This integration itself is a key and largely underexamined aspect of current art viewing models. Even more, by considering such models’ theoretical arguments for stages in art processing, and by looking to differences between novice viewers and experts while manipulating the time afforded to viewing a selection of abstract, representational, and kitsch paintings, it was argued that we might identify specific points of conflict or inflection in individuals’ ratings. Such kitsch switches, as we rather pithily coined them for this article, or other, say, art-historical or context-related changes, may reveal important new information about the dynamic nature of our processing experience.

Looking to the results, we had a mixture of findings that both support—albeit quite tentatively—the above hypotheses and also, perhaps more robustly, highlight other (i.e., predictive coding) arguments, which themselves are in need of much future research. First, and somewhat in contradiction to what we had expected, lay viewers and art experts both showed a rather pronounced consistency in their ratings with only minor adjustments between different viewing duration conditions for the different types of paintings. What is more, even from 100 ms, individuals showed quite pronounced differences in ratings for the different art types—with the kitsch paintings generally disliked by experts, relatively more liked by lays, and with the opposite pattern emerging for the abstract

Table 4
Linear Mixed Model (LMM) of Liking Ratings of Art Comparing Between Viewing Durations (100, 500, 6,000 ms, Between Participants) and Art Types (Kinkade, Constable, Richter, Within Participants) for Lay (Psychology Student) and "Art Expert" (Master Level Art History Student) Viewers, in Regards to Viewer Art Interest (Continuous Variable)

Variable	Estimate	SE	<i>t</i> value	Pr(> <i>t</i>)
Within artwork, between exposure times and art interest				
Kinkade				
Art Interest × Time (500–100 ms)	0.01	0.18	0.06	.95
Art Interest × Time (6,000–100 ms)	−0.18	0.19	−0.91	.36
Constable				
Art Interest × Time (500–100 ms)	0.29	0.13	2.23	.03*
Art Interest × Time (6,000–100 ms)	0.22	0.14	1.55	.12
Richter				
Art Interest × Time (500–100 ms)	0.30	0.17	1.81	.07‡
Art Interest × Time (6,000–100 ms)	−0.001	0.18	−0.01	.99
Within Exposure times and artworks, between art interest				
100 ms				
Kinkade: Art interest (continuous)	−0.54	0.13	−4.09	<.001*
Constable: Art interest (continuous)	−0.07	0.10	−0.72	.47
Richter: Art interest (continuous)	0.27	0.12	2.28	.02*
500 ms				
Kinkade: Art interest (continuous)	−0.53	0.13	−4.03	<.001*
Constable: Art interest (continuous)	0.23	0.09	2.38	.02*
Richter: Art interest (continuous)	0.58	0.12	4.83	<.001*
6 s				
Kinkade: Art interest (continuous)	−0.72	0.15	−4.93	<.001*
Constable: Art interest (continuous)	0.15	0.11	1.41	.16
Richter: Art interest (continuous)	0.27	0.13	2.05	.04*
Slope of art interest for each artist				
Kinkade: Art interest (continuous)	−0.59	0.08	−7.46	<.001*
Constable: Art interest (continuous)	0.10	0.06	1.75	.08‡
Richter: Art interest (continuous)	0.37	0.07	5.13	<.001*

Note. Results based on LMM with two factors Time and Artist, "art interest" entered as continuous variable (7-point scale), grand mean centered, and "liking" ratings (7-point scale) as dependent variable; treatment coded contrast for time, 100 versus 500 and 6,000 ms, and for each artist. Random intercepts for participants and images and random slopes of artist for participants and of art interest for images. Results based on 154 lay art viewers (psychology students) and 90 art history students (Masters level+). "Art interest" calculated via average score of five questions (7-point scales): "I am comfortable looking at and discussing art," "I think I have a good knowledge about art," "art is important," "I enjoy viewing art," and "I am interested in art." Bold text = significant following a Bonferroni correction, adjusted critical *p* value = .0028 (18 pairwise comparisons; six comparisons for the interaction between art interest and time; nine comparisons for ratings between art interest for each artist and time point, three comparisons for art interest for each of the artists).
‡ *p* < .10. * *p* < .05 (uncorrected for multiple comparison).

Richters and to a lesser extent for the Constables. This was also in spite of the generally similar colors or content, suggesting that some identifications or meaningful judgments can be made even with an extremely short exposure to art.

As recently argued for the predictive coding ideas that would be supported by this evidence, individuals may bring certain framing or expectations into an encounter (i.e., from the preclassification stage), which could modulate attention or assessments even from the very beginning of visual awareness. For example, it may be that a major distinction between lays and experts could be explained by fluency or the presence of identifiable content, which is most probably identifiable at 100 ms. That is, abstract art might be identified by experts and relatively liked whereas lay participants might dislike the Richters even at short durations because they are less comfortable with being unable identify what is "in" the artwork. Although the depicted scenes were chosen to be relatively the same for the Kinkades and Constables, the latter too may have also included subtle style choices making them less clear, and may

provide enough information to the viewer that the stimuli is something they would or would not like. Such a focus on fluency and even representational depictions have been given as a key component of kitsch versus nonkitsch art (see, e.g., Ortlieb & Carbon, 2018), and the above argument is supported by the fact that, in the present study, "content (represented images)" also correlated with liking for Kinkade and Richter in Table 4 and was significantly different between art experts and lay persons. Similarly, there may be rather ingrained reactions among lays and experts to aspects such as relative harmony, complexity, color and content relations, etcetera, which may also be detected quite early and inform initial impressions of paintings (see Mullin, Hayn-Leichsenring, Redies, & Wagemans, 2017; Schwabe, Menzel, Mullin, Wagemans, & Redies, 2018; Verhaver et al., 2018).

At the same time, when comparing directions of ratings changes between participants based on their background or expertise, we did also find some evidence for modulations or widening in ratings differences following our above switch hypotheses. These were

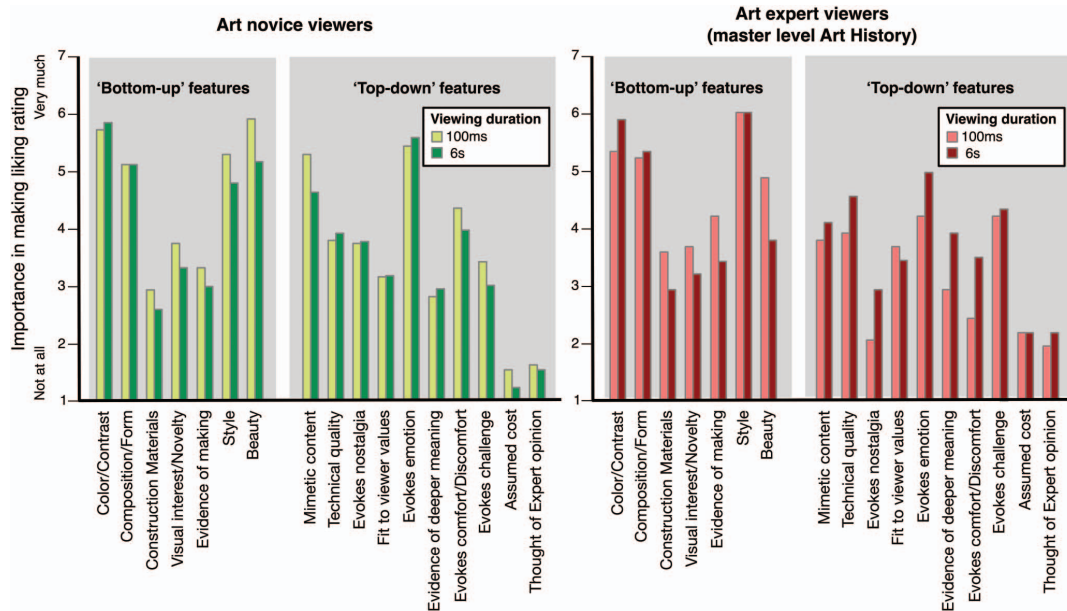


Figure 6. Differences in importance of factors when making liking appraisals of art between lay/expert viewers and 100-ms/6-s viewing durations. See the online article for the color version of this figure.

not as pronounced as the above main effect regarding different ratings assigned to artists. However, when looking at the collection of results and the comparison between viewers, we do see at least some notable emerging trends. In the population of lay viewers, when considering relative art interest, we found that, while ratings were quite similar at 100 ms, as viewing duration increased to 500 ms and especially 6 s, ratings of Kinkade tended to decrease among those with a relatively higher interest in art, whereas ratings of Richter increased. Among those with lower art interest, we found the opposite trend.

These time-related differences do add nuance to the above main finding of nearly immediate ratings differences and support the hypotheses for time-duration and processing stages. Interestingly, it seems that both of the latter (500 ms and 6 s) durations allow for modulation of ratings, suggesting that they may be rough points of such switching, or that basic style may be enough to consistently judge between, say, Kinkade and similarly themed Constable art. This would be in line with the theoretical argument that about 500 ms would allow a basic identification of content and style, whereas, 6 s was hypothesized to allow a more overt top-down modulation. These arguments were also supported by the consistency in ratings (as measured by McDonald's omega) for the individual paintings in an artist set. Similarly, tentative evidence was found in regards to viewer attention to different factors (e.g., style or color vs. meaning and context) when making ratings, which also showed some modulation (especially more attention to top-down aspects among experts) at longer viewing durations.

This article's results—both for and against a kitsch switch—thus have implications for our general understanding of perception

and the appraisal process. It does appear that, although the expectations we may bring into an encounter obviously appear to play an important role in guiding even the earliest attention and assessments, top-down context may also come online or become more salient at a certain point within the processing sequence. That is, despite the differing starting points, individuals also appear to use such information once again at the later stage to make further modifications. This was also supported in the subjective reports of the aspects viewers used to make their assessments, which also showed some changes based on time from more bottom-up to top-down features, suggesting also that viewers may have even had overt awareness. This moment, and the wider top-down and bottom-up interaction, should be a key target for future assessment, allowing us to better understand and anticipate how this is accomplished and what specific factors or cognitive processes are important. This raises the issue of these initial, preclassificatory and later switch stages' relative importance and integration within a processing sequence. This also has implications for the critical discussion of art, confirming the arguments for the need for a "reflective" approach, or even use of extra time and effort, in relation to responding to art or explaining why one may not like a painting.

This study of course comes with caveats and other demands for future research. First, it is important to stress that this study was largely exploratory and should be considered as only a step toward uncovering the temporal nature of our art assessments. In this line, a major caveat involves the general lack of robust changes in ratings for specific types of art when comparing between the time durations. Although again, in the collective consideration of results

Table 5
Correlation Between Relative Importance of Rating-Factors and Liking of Kitsch, Classical, and Abstract Paintings (Study 1 + 2 Viewers), and Mean Scores for Lay and Expert Viewers

Variable	Correlation with liking			Mean scores		
	Kinkade (Kitsch)	Constable (Classic)	Richter (Abstract)	M_{Lay}	M_{Expert}	t (Exp–Lay)
Beauty/visual appeal	.408**	.066	–.215*	5.6	4.2	4.2*
Color/contrast	.017	–.143	.049	5.8	5.5	1.0
Materials/media	–.244**	–.060	.183	3.0	3.5	1.3
Visual interest/novelty	–.158*	–.125	.257**	3.6	3.5	0.4
Evidence of making	.120	.148*	–.192**	3.5	3.6	0.3
Composition	.124	.179*	.037	5.1	5.3	0.9
Style	–.293**	.014	.209**	5.1	4.0	3.2*
Content (represented images)	.377**	.145*	–.264**	5.1	4.0	3.4*
Technical quality	–.078	.115	–.085	4.0	4.4	1.4
Assumed cost	–.087	.069	–.009	1.4	1.9	1.5
Evidence of deeper meaning	–.089	–.061	.010	3.2	3.5	1.0
Evokes security/comfort	.524**	.142**	–.401**	4.0	2.9	3.2*
Evokes challenge	–.367**	–.079	.375**	3.5	4.3	2.1
Evokes nostalgia	.336**	.156*	–.243*	3.7	2.8	2.8
Fits my values or beliefs	.006	–.036	–.004	3.4	3.8	1.0
Evokes emotion	.144	.009	–.097	5.5	4.8	1.9
I thought of expert opinions	–.308**	.038	.188*	1.7	2.0	1.1

Note. Correlations based on Pearson product moment analyses (two-tailed). Analyses of correlations are not corrected for multiple comparisons. The reader should be mindful of this when making inferences. The comparisons of means between experts and lay viewers are corrected for multiple comparisons (familywise correction, adjusted alpha, 17 parallel analyses = .0029).

* Significant at $p < .05$. ** Significant at $p < .01$.

and evidence, a pattern can be found and an argument made for a switch in ratings, contrary to hypotheses this was not pronounced. It may of course be that such an effect is quite small or that our measures or sample were not robust enough to detect this. An immediate target would also involve a within-participant paradigm, which may be able to more directly test this. Although we have not done so here, we suggest future follow-up studies might consider this using a similar paradigm and sets of artworks.

Another next step would be to further zero in on the exact temporal location of a switch in perception or ratings. It would also be useful to further refine the present analysis with even higher temporal resolution or using better continuous behavioral or physiological measures, such as facial EMG, for example looking even between 100 and 500 or incrementally before 6,000 ms. It would also be quite interesting if we can identify where such an online top-down integration might occur in the brain, as was one of the future intended extensions of the present behavioral study.

It could also be that our chosen times—even our shortest duration of 100 ms—were too “late” to detect certain processing. For example, Augustin and colleagues (2008) have suggested that even 50 ms might allow some content and even style awareness. Verhaver et al. (2018) have similarly reported that aspects such as specialness, impressiveness, as well as beauty, may be detected at 30–50 ms (see also Mullin et al., 2017, for gist perceptions of scenes or Schwabe et al., 2018, for gist perceptions of abstract artworks at 50 ms). As noted above, such basic information—balance, color, a gist of a scene—might be enough for an individual to decide that they like or do not like especially certain types of art. Alternatively, it may of course be that, even in the absence/presence of initial rating differences, individuals need much more time for initial ratings to significantly change as a result of top-down self-related processing. For example, museum and laboratory studies in which viewers are allowed to view art at their own

pace, show mean durations of 10–12 but also often closer to 40+ s (Pelowski et al., 2017 for a review), whereas more profound/self-involved or revisionist reactions might be measured on the scale of minutes (see also Pelowski, 2015; Smith & Smith, 2001). Looking beyond initial viewing—at the level of 10+ to 30 or even more seconds—may provide a window into other aspects of top-down modulation that are equally intriguing when considering art processing experience.

That said, indeed, individuals may routinely switch their evaluations as different artwork or context-related factors become more salient. We hope that this article can serve as a basis for such perception and evaluation research.

References

- Augustin, M. D., Leder, H., Hutzler, F., & Carbon, C.-C. (2008). Style follows content: On the microgenesis of art perception. *Acta Psychologica*, 128, 127–138. <http://dx.doi.org/10.1016/j.actpsy.2007.11.006>
- Bachmann, T., & Vipper, K. (1983). Perceptual rating of paintings from different artistic styles as a function of semantic differential scales and exposure time. *Archiv für Psychologie*, 135, 149–161.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Berridge, K. C., Ho, C.-Y., Richard, J. M., & DiFeliceantonio, A. G. (2010). The tempted brain eats: Pleasure and desire circuits in obesity and eating disorders. *Brain Research*, 1350, 43–64. <http://dx.doi.org/10.1016/j.brainres.2010.04.003>
- Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106, 265–289. <http://dx.doi.org/10.1037/0033-2909.106.2.265>
- Brieber, D., Forster, M., & Leder, H. (2018). On the mutual relation between art experience and viewing time. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. <http://dx.doi.org/10.1037/aca0000210>

- 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. <http://dx.doi.org/10.1371/journal.pone.0099019>
- Burkeman, O. (2006, March 25). Dark clouds gather over "Painter of Light". *Guardian*. Retrieved from <https://www.theguardian.com/world/2006/mar/25/arts.artsnews>
- Cela-Conde, C. J., Marty, G., Maestú, F., Ortiz, T., Munar, E., Fernández, A., . . . Quesney, F. (2004). Activation of the prefrontal cortex in the human visual aesthetic perception. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 101, 6321–6325. <http://dx.doi.org/10.1073/pnas.0401427101>
- Chatterjee, A. (2003). Prospects for a cognitive neuroscience of visual aesthetics. *Bulletin of Psychology and the Arts*, 4, 55–59.
- 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>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181–204. <http://dx.doi.org/10.1017/S0140525X12000477>
- Cooper, J. M., & Silvia, P. J. (2009). Opposing art: Rejection as an action tendency of hostile aesthetic emotions. *Empirical Studies of the Arts*, 27, 109–126. <http://dx.doi.org/10.2190/EM.27.1.f>
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3, 201–215. <http://dx.doi.org/10.1038/nrn755>
- Cupchik, G. C. (2011). The role of feeling in the Entertainment = Emotion formula. *Journal of Media Psychology: Theories, Methods, and Applications*, 23, 6–11. <http://dx.doi.org/10.1027/1864-1105/a000025>
- Cutting, J. E. (2003). Gustave Caillebotte, French impressionism, and mere exposure. *Psychonomic Bulletin & Review*, 10, 319–343.
- Didion, J. (2003). *Where I was from*. Westminster, MD: Knopf.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138. <http://dx.doi.org/10.1038/nrn2787>
- Garcia, J., Lasiter, P. S., Bermudez-Rattoni, F., & Deems, D. A. (1985). A general theory of aversion learning. *Annals of the New York Academy of Sciences*, 443, 8–21. <http://dx.doi.org/10.1111/j.1749-6632.1985.tb27060.x>
- Geday, J., Kupers, R., & Gjedde, A. (2007). As time goes by: Temporal constraints on emotional activation of inferior medial prefrontal cortex. *Cerebral Cortex*, 17, 2753–2759. <http://dx.doi.org/10.1093/cercor/bhm002>
- Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*, 33, 587–606. <http://dx.doi.org/10.1016/j.soc.2004.09.033>
- Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, 12, 878–892. <http://dx.doi.org/10.1080/13506280444000562>
- Gombrich, E. H. (1985). *Meditations on a hobby horse and other essays on the theory of art*. London, UK: Phaidon Press Ltd.
- 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. <http://dx.doi.org/10.1177/1088868315574978>
- Greenberg, C. (1939). The avant-garde and kitsch. *Partisan Review*, 6, 34–49.
- Heath, J., & Potter, A. (2005). *The rebel sell: Why the culture can't be jammed*. Chichester, UK: Capstone.
- Jakesch, M., & Leder, H. (2009). Finding meaning in art: Preferred levels of ambiguity in art appreciation. *The Quarterly Journal of Experimental Psychology*, 62, 2105–2112. <http://dx.doi.org/10.1080/17470210903038974>
- Karpen, F. (1925). *Der Kitsch. Eine Studie über die Entartung der Kunst*. Hamburg, Germany: Weltbund Verlag.
- Kuchinke, L., Trapp, S., Jacobs, A. M., & Leder, H. (2009). Pupillary responses in art appreciation: Effects of aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, 3, 156–163. <http://dx.doi.org/10.1037/a0014464>
- Kulka, T. (2002). *Kitsch and art*. University Park, PA: Penn State Press.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82, 82. <http://dx.doi.org/10.18637/jss.v082.i13>
- Lauring, J. O., Pelowski, M., Forster, M., Gondan, M., Pfitz, 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>
- 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. <http://dx.doi.org/10.1111/bjop.12084>
- Lemmens, J. S., Valkenburg, P. M., & Peter, J. (2009). Development and validation of a game addiction scale for adolescents. *Media Psychology*, 12, 77–95. <http://dx.doi.org/10.1080/15213260802669458>
- Lengger, P. G., Fischmeister, F. P. S., Leder, H., & Bauer, H. (2007). Functional neuroanatomy of the perception of modern art: A DC-EEG study on the influence of stylistic information on aesthetic experience. *Brain Research*, 1158, 93–102. <http://dx.doi.org/10.1016/j.brainres.2007.05.001>
- Locher, P., Krupinski, E. A., Mello-Thoms, C., & Nodine, C. F. (2008). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, 21, 55–77.
- Locher, P., Overbeeke, K., & Wensveen, S. (2010). Aesthetic interaction: A framework. *Design Issues*, 26, 70–79. http://dx.doi.org/10.1162/DESL_a_00017
- Lorand, R. (1994). Beauty and its opposites. *The Journal of Aesthetics and Art Criticism*, 52, 399–406.
- Marin, M. M., & Leder, H. (2016). Effects of presentation duration on measures of complexity in affective environmental scenes and representational paintings. *Acta Psychologica*, 163, 38–58. <http://dx.doi.org/10.1016/j.actpsy.2015.10.002>
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Henry Holt and Co.
- Meskin, A., Phelan, M., Moore, M., & Kieran, M. (2013). Mere exposure to bad art. *British Journal of Aesthetics*, 53, 139–164. <http://dx.doi.org/10.1093/aesth/ays060>
- Miller, L. (2002, March 18). The writer of dreck TM. *Salon.com*. Retrieved from <https://www.salon.com>
- Mullin, C., Hayn-Leichsenring, G., Redies, C., & Wagemans, J. (2017). The gist of beauty: An investigation of aesthetic perception in rapidly presented images. *Electronic Imaging*, 2017, 248–256. <http://dx.doi.org/10.2352/ISSN.2470-1173.2017.14.HVEL-152>
- Munar, E., Nadal, M., Rosselló, J., Flexas, A., Moratti, S., Maestú, F., . . . Cela-Conde, C. J. (2012). Lateral orbitofrontal cortex involvement in initial negative aesthetic impression formation. *PLoS ONE*, 7(6), e38152. <http://dx.doi.org/10.1371/journal.pone.0038152>
- Nerdrum, O. (2001). *On kitsch*. Oslo, Norway: Kagge Forlag AS.

- 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*, 1077–1084.
- Ortlieb, S. A., & Carbon, C. C. (2018). Kitsch and perception: Towards a new “Aesthetic from Below.” *Art & Perception*, *1*, 1–26.
- Ortlieb, S. A., & Carbon, C. C. (2019). A functional model of kitsch and art: Linking aesthetic appreciation to the dynamics of social motivation. *Frontiers in Psychology*, *9*, 2437. <http://dx.doi.org/10.3389/fpsyg.2018.02437>
- 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., 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. <http://dx.doi.org/10.3389/fpsyg.2017.01729>
- 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>
- 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>
- Perl, J. (2011, July 14). Bullshit Heaven, review of Alexis L. Boylan (Ed.), *Thomas Kinkade: The artist in the mall* (Durham, NC: Duke University Press, 2011). *New Republic*. Retrieved from <https://newrepublic.com/article/91615/thomas-kinkade>
- R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, Vienna, Austria.
- Reilly, S., & Schachtman, T. R. (2009). Introduction. In S. Reilly & T. R. Schachtman (Eds.), *Conditioned taste aversion: Behavioral and neural processes*. New York, NY: Oxford University Press.
- Schwabe, K., Menzel, C., Mullin, C., Wagemans, J., & Redies, C. (2018). Gist perception of image composition in abstract artworks. *i-Perception*, *9*, 2041669518780797. <http://dx.doi.org/10.1177/2041669518780797>
- Smith, J. K., & Smith, L. F. (2001). Spending time on art. *Empirical Studies of the Arts*, *19*, 229–236. <http://dx.doi.org/10.2190/5MQM-59JH-X21R-JN5J>
- Szpunar, K. K., Schellenberg, E. G., & Pliner, P. (2004). Liking and memory for musical stimuli as a function of exposure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 370–381. <http://dx.doi.org/10.1037/0278-7393.30.2.370>
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network*, *14*, 391–412. http://dx.doi.org/10.1088/0954-898X_14_3_302
- Verhaver, S., Wagemans, J., & Augustin, M. D. (2018). Beauty in the blink of an eye: The time course of aesthetic experiences. *British Journal of Psychology*, *109*, 63–84. <http://dx.doi.org/10.1111/bjop.12258>
- Wilson, W. R. (1979). Feeling more than we can know: Exposure effects without learning. *Journal of Personality and Social Psychology*, *37*, 811–821. <http://dx.doi.org/10.1037/0022-3514.37.6.811>
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach’s α , Revelle’s β , and McDonald’s ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*, 123–133. <http://dx.doi.org/10.1007/s11336-003-0974-7>

(Appendix follows)

Appendix
Supporting Tables

Table A1
McDonald's Omega (ω_{μ}) Scores (Interstimuli Consistency in Ratings) for Differing Viewing Durations and Types of Art

Variable	Kinkade	Constable	Richter
Lay			
100 ms	0.72	0.89	0.82
500 ms	0.86	0.57	0.80
6 s	0.90	0.62	0.77
100 ms	0.93	0.65	0.78
Experts			
100 ms	0.89	0.86	0.81
500 ms	0.92	0.80	0.79
6 s	0.83	0.79	0.67

Received April 3, 2019

Revision received August 30, 2019

Accepted December 19, 2019 ■