# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

## Visual Exploration and Cognitive Representation during an Aesthetic Experience

What is special about an aesthetic mode of processing?

verfasst von / submitted by

Johanna Wilflingseder, BSc

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

Master of Science (MSc)

Wien, 2020 / Vienna 2020

| | |
|---|---|
| Studienkennzahl lt. Studienblatt / degree program code as it appears on the student record sheet: | UA 066 840 |
| Studienrichtung lt. Studienblatt / degree program as it appears on the student record sheet: | Masterstudium Psychologie UG2002 |
| Betreut von / Supervisor: | Univ.-Prof. Dipl.-Psych. Dr. Helmut Leder |

# Abstract

The aim of this study was to find out more about aesthetic experiences and their underlying processes. We used eye tracking as a method to clarify how art is represented in our minds and conducted a memory experiment to find out how strong the created representations are. Aesthetic processing was expected to be “deeper” and thereby create strong, detailed representations reflected in higher recognition rates in comparison to visual complexity judgements, both of which are incidental encoding tasks. The basic experimental setup consisted of an encoding phase, a distractor task, and a recognition phase. Half of the participants were asked to give liking evaluations of artworks, while the other half rated their visual complexity. Additionally, we expected participants in the aesthetic condition to show features associated with a global processing strategy during encoding such as higher fixation counts, lower fixation durations, and larger saccades. These are also associated with the apparent memory enhancing effects. Contrary to our predictions eye movement patterns did not differ significantly between the aesthetic and the complexity condition, nor did the measure of recognition accuracy. This indicates that either aesthetic judgements failed to induce a “deeper” aesthetic experience, or that these do not differ essentially from other rating tasks in the representations they create. The exploratory analysis revealed, however, that higher liking ratings were related to better recognition performance. This supports the connection between aesthetic processing and memory and may be subject of future investigations.

# Zusammenfassung

Ziel dieser Studie war es mehr über ästhetische Erfahrungen und die zugrunde liegenden Prozesse herauszufinden. Wir nutzten Eye-tracking als Methode um zu klären, wie Kunst mental repräsentiert wird und führten dazu ein Gedächtnisexperiment durch. Das Experiment bestand aus einer Enkodierungsphase, einer Distraktionsphase und einer Wiedererkennungsphase. Die Hälfte der TeilnehmnerInnen wurde gebeten, Kunstwerke zu evaluieren anhand wie gut sie ihnen gefielen, während die andere Hälfte deren visuelle Komplexität bewertete. Es wurde erwartet, dass ästhetische Verarbeitung tiefer ist und nachhaltigere Repräsentationen hervorruft, was sich in einer besseren Wiedererkennungsleistung verdeutlichen sollte im Vergleich zur Kontrollgruppe. Die bessere Gedächtnisleistung sollte außerdem mit dem für ästhetische Verarbeitung typischen globalen Augenbewegungsmuster einhergehen, das durch erhöhte Fixationen, eine niedrigere Fixationsdauer, und größere Blicksprünge gekennzeichnet ist. Entgegen unseren Erwartungen unterschieden sich weder die Augenbewegungsmuster noch die Gedächtnisleistung der beiden Gruppen voneinander signifikant. Das könnte einerseits daran liegen, dass Gefallensurteile nicht ausreichten, um die gewünschte ästhetische Erfahrung hervorzurufen. Andererseits könnte es auch bedeuten, dass sich die gebildeten Repräsentationen einer ästhetischen Verarbeitung nicht deutlich unterscheiden von anderen Arten der Verarbeitung. Nichtsdestotrotz ergab die explorative Analyse, dass positivere Gefallensurteile mit einer besseren Wiedererkennungsleistung einhergingen. Das wiederum unterstreicht den Zusammenhang zwischen ästhetischer Verarbeitung und Gedächtnis und kann künftig als Ansatz weiterverfolgt werden.

Imagine you are standing in an art gallery. The room is filled with beautiful artworks, but one of them especially catches your attention. Imagine the thoughts and feelings that strike you as soon as you start examining this piece more closely. It might activate knowledge about its style or the artist who created it, it might bring back old memories, you may feel moved, lose your sense of time and space, or you may simply appreciate its beauty. After days or even weeks have passed you still recall the artwork in great detail. It has left an irreversible impression and marked a deep memory trace.

Most of us have already encountered such an incredible sensation while looking at artworks of various kinds. But even rather mundane objects and scenes outside of the artistic context, such as photographs with high aesthetic and/or personal value, or the appreciation of beautiful naturalistic landscapes can sometimes induce a unique state of mind in the observer. This kind of emergent state is called *aesthetic experience* (Chatterjee & Vartanian, 2014; Cupchik, Vartanian, Crawley, & Mikulis, 2009; Di Dio & Gallese, 2009; Massaro et al., 2012). It is defined as a cognitive appraisal process that involves attention allocation and is accompanied by sensorimotor and emotional mechanisms (Leder, Belke, Oeberst, & Augustin, 2004). The primary focus of interest lies within the perceiver while also taking external factors into account. The perceiver's top-down processes are influenced by the person's cultural and educational background, the degree of training and interest as well as subjective taste (Leder et al., 2004; Massaro et al., 2012). Bottom-up factors that influence aesthetic experiences include contrast (Ramachandran & Hirstein, 1999), visual complexity (Nadal, Munar, Marty, & Cela-Conde, 2010; Frith & Nias, 1974), color (Zeki, 1980) and symmetry (Locher & Nodine, 1987). Considering this complex interplay of top-down and bottom-up factors, it seems reasonable to believe that there is something special and unique about the way we perceive art and it is one of the main goals of empirical aesthetics to find out more about the underlying processes. *Neuroaesthetics* is the discipline that investigates the neural bases of beauty and art perception and is especially centered around the biological substructure of aesthetic experiences (Chatterjee & Vartanian, 2014; Di Dio & Gallese, 2009). Guided by the notion that aesthetic processing diverges substantially from other types of processing, we conducted an experiment with the intention of highlighting this difference. We used eye tracking, which gave us some insight into the formation of mental representations while processing artworks under two different conditions. We expected processing to be deeper, when approaching the artworks with an aesthetic orientation compared to a non-aesthetic orientation, which was tested by a subsequent recognition test.

## Theoretical Background of Aesthetic Experiences

How exactly are aesthetic experiences formed? Answering this question was one of the main goals of Leder and colleagues (2004) by developing their model of aesthetic experience. They created a comprehensive framework within which top-down and bottom-up influences interact to generate aesthetic experiences. Specifically, they describe five successive but continuously interacting processing stages during the aesthetic appreciation of artworks that follow an initial pre-classification: (1) “perceptual analysis”, in which low-level visual features (e.g., contrast, shape) are extracted; (2) “implicit memory integration”, in which previous experiences, expertise and individual schemas are activated; (3) “explicit classification”, in which formal knowledge about content and style is integrated; (4) “cognitive mastering”, in which meaning is extrapolated by forming self- and art-related interpretations and associations; and finally (5) “evaluation”, in which the cognitive mastering process is constantly monitored and evaluated. If successful, the final outputs consist of aesthetic judgement and emotion.

In accordance with this model numerous studies have used aesthetic judgements as an indicator of aesthetic processing and experience (Jacobsen, Schubotz, Höfel, & Cramon, 2006; Locher, Krupinski, Mello-Thoms, & Nodine, 2007; Massaro et al., 2012; Nadal, Marty, & Munar, 2006; Nadal et al., 2010; Wallraven et al., 2009), because they are explicit and measurable responses that represent successful cognitive mastering of an artwork (Leder et al., 2004; Massaro et al., 2012). Aesthetic judgements have also been used in other contexts such as evaluating aesthetic preference of scenes (Choe et al., 2017; Henderson & Hayes, 2017), patterns (Jacobsen et al., 2006) and faces (Bernstein, Beig, Siegenthaler, & Grady, 2002; Marzi & Viggiano, 2010). This is linked to the idea that anything we perceive can be processed with an aesthetic orientation and therefore benefit from its unique characteristics. But what exactly are the benefits from an aesthetic mode of processing?

## Aesthetic Processing and Memory

Memory can be formed without explicitly trying to memorize. This is called *incidental memory* (Choe et al., 2017; 1997; Craik, 2002; Craik & Lockhart, 1972; Hollingworth & Henderson, 2002). Previous research has shown that visual information of objects (Castelhano & Henderson, 2005) and scenes (Wolfe, Horowitz, & Michod, 2007) can be acquired and stored whether or not the observer is intentionally trying to remember that information. Of course, very often intentional memorization creates stronger memory compared to incidental encoding (Tatler & Tatler, 2013). However, there has also been a report of cases in which incidental memory outperformed intentional memorization. For example, after looking for objects in scenes participants showed better memory for the searched objects compared to the ones they were explicitly asked to memorize (Draschkow, Wolfe, & Võ, 2014; Josephs, Draschkow, Wolfe, & Võ, 2016). Similarly, aesthetic preference evaluations of faces, scenes and objects resulted in better memory than intentionally memorizing them (Bernstein et al., 2002; Choe et al., 2017; Grady, Bernstein, Beig, & Siegenthaler, 2002). This memory boost of aesthetic appreciation appears to have specific underlying neural processes. Three brain areas are often reported to show activation during the execution of aesthetic judgements: regions related to (1) reward/pleasure and emotion, (2) judgement/decision making, and (3) perception (Cela-Conde et al, 2013). The collective activation of these neural networks seems to facilitate better memory consolidation and the strength of the resulting memory traces depends on the construction and composition of representations (Nadal et al., 2006). But what exactly are those representations and what do they look like?

In ancient Greece Aristotle already noted that the human mind creates internal representations of the external world, which are used for thought (Aristotle, De Anima; $431_a$, $431_b$). Today this concept of visual imagery is described as the ability to generate mental images in the absence of retinal input (Ishai, 2011). Numerous neuroimaging studies have shown that visual imagery evokes activation in the occipito-parietal and occipito-temporal regions as well as the primary visual cortex, which are normally active during visual perception (Kosslyn et al., 1993; Le Bihan et al., 1993; Mellet et al., 1996). These mental representations, however, should not be understood as an exact depiction of the perceived visual information. In fact, mental representations have been shown to be quite fragile. Research on change blindness provides evidence that large discrepancies in scenes can go unnoticed (Bridgeman, Hendry, & Stark, 1975; Rensink, O'Regan, & Clark, 1999; Simons & Levin, 1997). These studies show that observers are very bad at recognizing substantial changes in scenes unless they happened to be attending to the elements that were changed. Clearly, internal representations are not just the reproductions of sensory information impinging the retina, nor are they equivalent to photographs. Instead it is more likely that perceivers focus their attention on specific features and thereby extract the essence of the visual input (Castelhano & Henderson, 2005; Wolfe et al., 2007). This is supported by Locher et al. (2007), who suggest that viewers are very fast and effective in obtaining the "gist" of attended objects and scenes. This information is then stored and can later be retrieved remarkably well (Wolfe et al., 2007). For example, Standing and colleagues (1970, 1973) demonstrated that humans have a vast memory capacity for photographs of natural objects. In line with that, Hollingworth and Henderson (2002) also highlighted their participants' exceptional memory performance for objects, as they were able to discriminate between previously shown stimuli and visual distractors from the same basic-level category with ease. This indicates that although some elements of the visual input are neglected during the construction of mental representations, as demonstrated by studies on change blindness (Bridgeman et al., 1975; Rensink et al., 1999; Simons & Levin, 1997), other fractions of that visual information are represented quite well in memory and can be retrieved on subsequent occasions. In our study we therefore tried to investigate what kind of information exactly is used to form representations during an aesthetic mode of processing and how well this information is stored in memory. By conducting a recognition test we were able to compare recognition performance between an aesthetic condition, in which artworks were rated according to their liking, and a non-aesthetic condition, in which visual complexity had to be rated. We expected participants in the aesthetic condition to profit from the aesthetic processing mode and thereby show memory advantages in the recognition test. To measure which elements of the visual information were extracted to form internal representations, we used eye tracking as an instrument as it provides insights into perceptual and cognitive processes, and therefore helps to understand the composition and structure of representations (Castelhano & Henderson, 2005).

## Aesthetic Processing and Eye Movements

Already in the 1930s Buswell (1935) conducted first investigations on visual processing using eye tracking. Participants viewed photographs of artworks while their eye movements where analyzed. Yarbus (1967), another pioneer in the study of eye movements, found that observers tend to focus on the most informative regions of images rather than scanning it randomly. Eye tracking is considered to be effective in detecting cognitive processes, because eye movements give information about visual attention (Fuchs, Ansorge, Redies, & Leder, 2011). Specifically, they reflect *overt attention*, which is defined as the selective processing of objects/locations at the expense of others due to a movement of the eyes (Henderson, 2017). Eye tracking therefore allows the examination of attention allocation under various conditions and contexts (Choe et al., 2017). Which perceptual elements exactly attract our attention depends on bottom-up as well as top-down processes.

**Bottom-up and top-down influences.** People's attention can be captured by visual items that stick out. Salience describes low-level image discontinuities such as bright regions, edges, color differences (Kardan, Henderson, Yourganov, & Berman, 2016). According to the saliency model of visual attention these visual features capture attention in a bottom up or stimulus-driven way (Itti & Koch, 2000). Support for this hypothesis was given by Fuchs et al. (2011), who showed that participants' eye movements were mostly located in salient regions. Similarly, Kardan et al. (2016) found that locations with higher salience captured the majority of eye movements while performing a visual search task. However, they also found that this effect was reliably smaller when participants had to perform other tasks, such as memorizing scenes or making aesthetic judgements.

That eye movements are not only influenced by visual image features but also by the viewer's task and goals was already established in the early works of Yarbus (1967) and Buswell (1935). In their studies participants showed different eye movement patterns depending on which task they had to perform. These variations in eye movements reflect different cognitive processes that direct gaze toward informative task-relevant locations (Kardan et al., 2016; Henderson, Malcolm, & Schandl, 2009). In an elaborate virtual reality experiment by Rothkopf, Ballard, and Hayhoe (2007) participants also focused primarily on items related to the instruction even with salient features present that supposedly attract attention. These studies show that human gaze is directed toward regions of the visual scene that are determined primarily by top-down cognitive control. Additional support for this finding was given by recent research suggesting that semantic informativeness plays an important role in attracting attention (Henderson & Hayes, 2017). By identifying informative features, they created so called meaning maps that allow predictions of semantic content across a scene, just as saliency maps provide predictions of salience across a scene. After contrasting both maps it was evident that participants' attention was strongly guided by cognitive factors related to the semantic features that are relevant to understanding the scene rather than by low-level image salience. Peacock, Henderson, and Hayes (2019) demonstrated in a follow-up study that even in tasks for which meaning is irrelevant and salience is relevant, meaningful features are more successful at attracting attention. These findings are consistent with earlier cognitive control theories, which prioritize semantic properties rather than image properties in attention allocation.

In sum, these studies reveal that eye movements are an index of overt attention that provide strong links between the external environment and our internal states and representations. The direction of gaze reflects ongoing cognitive processes that observers are engaged in (Kardan et al., 2016; Rothkopf et al.2007). Therefore, the analysis of eye movement patterns can help shed light on the respective contribution of bottom-up and top-down processes of aesthetic experiences.

**Specific eye movement patterns.** Visual attention is accompanied by eye movements in the form of relative immobility separated by quick jumps (Loftus, 1972). Fixations are defined as "stable eye positions directed toward a specific scene location or object used for visual information acquisition, and saccades are fast, ballistic eye movements that reorient fixations from one location to another within a scene" (Henderson, 2017, p.16). While vision is essentially suppressed during saccades, fixations serve to acquire and process visual information, which is then stored in memory (Loftus, 1972). The following three measures are frequently used in eye tracking studies to get insight into attentional processes: (1) the fixation count, which is the total number of fixations that land on a certain position; (2) the fixation duration, which is the time spent fixating on the same location; and finally (3), saccade amplitude, which describes the length of the jump from one fixation to the next. Previous research has shown that fixation durations and saccade amplitudes are governed by the same mechanism and thereby strongly connected (Pannasch et al., 2008; Unema, Pannasch, Joos, & Velichkovsky, 2005; Velichkovsky, Joos, Helmert, & Pannasch, 2005). Both reach an asymptotic level with increased viewing time (Antes, 1974; Unema et al., 2005), and both seem to complement each other: fixations of shorter durations have been found to be predominantly followed by longer saccade amplitudes, while longer fixations are mostly followed by shorter saccades (Velichkovsky, 2002). The former describes characteristics of a global processing strategy often observed in early viewing. The latter pattern is characteristic for local processing mainly present during later viewing (Pannasch et al., 2008). But not only viewing time determines whether the content is processed locally or globally. As previously mentioned, artists as well as art experts generally tend to show more global processing strategies compared to lay people (Pihko et al., 2011; Koide et al., 2015; Vogt, 1999; Zangemeister et al., 1995). Similarly, the task itself can influence which of the two processing types is predominantly active during viewing. Mills et al. (2011) found that giving the perceiver different instructions on how to visually process the image affects both spatial (e.g. saccade amplitude) and temporal characteristics (e.g. fixation duration) of fixations. Visual search tasks, for example, are generally characterized by global processing strategies with relatively short fixation durations and large saccades (Choe et al., 2017; Mills et al., 2011). Memorization tasks on the other hand tend to evoke more local processing strategies with long fixation durations and short saccades (Henderson et al., 1999; Mills et al., 2011), as well as higher numbers of fixations than visual search (Castelhano, Mack, & Henderson, 2009; Tatler & Tatler, 2013). Importantly, aesthetic preference judgements also accompanied by unique characteristics. Approaching images with an aesthetic orientation has been shown to induce global fixation patterns in the observer with high fixation counts, low fixation durations and large saccades (Choe et al., 2017; Mills et al., 2011). In order to evaluate an image aesthetically it seems useful to scan the whole piece without leaving any details unconsidered and gather as much information as possible within the given timeframe. We therefore expected our participants to show this global processing pattern in the aesthetic condition and used the non-aesthetic complexity condition as a control group to compare the eye movements with.

**Effects of expertise.** Not only task but also the individual level of expertise influences the way our eyes move. This was already established by the works of Nodine, Locher, and Krupinski (1993), who found that lay people show very different gaze patterns compared to art-trained experts. The former typically focused on central and foreground figures whereas the latter spent more time focusing on the analysis of overall compositional design and the relationships among objects. Several other studies supported this finding suggesting that art experts use a more global image processing strategy than lays (Pihko et al., 2011; Vogt, 1999; Zangemeister, Sherman, & Stark, 1995). This is in line with the model by Leder and colleagues (2004) in which knowledge and expertise highly influence aesthetic processing, especially in the late stages of explicit classification and cognitive mastering. With increasing expertise, the nature of the representation becomes more abstract as more specialized knowledge can be integrated, leading to a deeper comprehension of the artwork and resulting in greater appreciation and aesthetic experience (Augustin & Leder, 2006; Leder, Gerger, Dressler, & Schabmann, 2012).

Like art experts, artists have also been shown to be less driven by low-level features such as objects and figures, but to focus more on texture and color composition using a global processing strategy (Koide et al., 2015; Zangemeister et al., 1995). This has direct implications on memory, as artists remember pictorial features significantly better than untrained viewers (Vogt & Magnussen, 2007). Due to their domain-specific advantages artists are more efficient in visual encoding which benefits their mental representations and long-term preservation of visual detail in memory (Glazek, 2011). These studies demonstrate that artists and art-trained experts use different viewing strategies than lay persons which directly impact their eye movement patterns and cognitive processes. Art expertise was therefore an important factor that had to be considered in our study as well. We asked artists, specifically painters, not to participate in the experiment and used the Viennese Art Interest and Art Knowledge Scale (Specker et al., 2018) as a tool to differentiate between lay people and art experts in terms of art interest and art knowledge. Our focus was primarily based on the perception of art by everyday people, which is why we excluded participants categorized as art experts by the means of the VAIAK. This way we were able to eliminate the influence of art expertise on participants' eye movements as a confounding variable.

## Eye Movements and Memory

There is a strong link between eye movements and memory. Castelhano & Henderson (2005) argued that solid representations of visual information can only be created via multiple fixations on various elements of the scene, allowing the integration of fine details. The nature and strength of the representation is therefore reflected in the eye movements of the observer. Especially a high number of fixations has been shown to be an important predictor of memory performance, as each fixation retains more information for the construction of the mental representation of the stimulus (Choe et al., 2017; Loftus, 1972). The creation of a representation is also highly task dependent. Tatler and Tatler (2013) found a strategic increase in number of fixations on task-relevant objects. This suggests that shifting the focus on task-relevant elements and accumulating detailed visual information of those elements via multiple fixations leads to enhanced encoding. Similarly, the amount of time spent on each fixation is associated with ongoing cognitive processing and thereby closely linked to memory (Henderson & Pierce, 2008; Hollingworth & Henderson, 2002). The longer the fixation on the same point the more information can be extracted from that part of the visual input. This was demonstrated by Fuchs and colleagues (2011), who showed that abstract artworks gained longer fixations than depictive ones. Due to their ambiguous content abstract paintings are more difficult to analyze and require increased processing efforts reflected in longer fixation durations. Kardan and colleagues (2015) also confirmed differences in eye movement patterns across different conditions (memorization, visual search, aesthetic preference), and were additionally able to predict the task participants were engaged in, merely by analyzing eye movement data. This is a striking finding and clearly demonstrates that eye movement characteristics can be used to classify specific cognitive states as they reveal information about the composition of the representations they create. For example, during memorization tasks participants seem to spend most of their time fixating on certain image features in order to create strong memory cues, whereas in visual search tasks observers mostly scan the image with large saccades until the object of interest is found. It does not come as a surprise that usually intentional encoding, in this case memorization, leads to better recognition rates than incidental encoding (e.g. visual search) when observers are confronted with a recognition test (Tatler & Tatler, 2013). However, as already mentioned earlier, after giving aesthetic preference judgements participants have been shown to remember images better than when they were explicitly asked to memorize them (Bernstein et al., 2002; Choe et al., 2017). This is most likely associated with the unique eye movement pattern that aesthetic judgements induce in the observer. The representation created by aesthetic preference seems to be more detailed due to the high number of fixations which may facilitate subsequent memory retrieval. We therefore expected participants in the aesthetic condition to perform better at the recognition test as a result of the specific global fixation pattern of aesthetic processing (high fixation count, low fixation duration and large saccades) which is believed to produce enhanced representations of the depicted artworks. We further intended to use this fixation pattern to predict the recognition performance of each condition.

## A Gap in Past Research

Although previous research on aesthetic processing has already compared different kinds of incidental encoding tasks (Choe et al., 2017; Kardan et al., 2015; Mills et al., 2011), there has not been much variation within this category. Most studies used visual search and aesthetic preference tasks. But when it comes to focusing on aesthetic experience and its unique characteristics, visual search is hardly a suitable incidental encoding task to draw inferences for aesthetics. Already intuitively it seems obvious that searching for objects in images is accompanied by different processing mechanisms than rating that image on a given scale. Empirical evidence was given by Murty, DuBrow, and Davachi (2015) who showed that the simple act of choosing resulted in enhancements in declarative memory via the striatum, an area involved in subjective evaluation and decision making (Barta, McGuire, & Kable, 2013), indicating that the very act of making a rating may induce an unintentional memory boost. Studies on aesthetic experience should therefore consider comparing aesthetic preference ratings to other types of rating tasks instead of using visual search. This is also highlighted by the depth of processing theory which proposes a continuum of processing levels, in which deeper levels of processing (e.g. semantic judgements) produce stronger, longer lasting, and more elaborate memory traces than shallow levels (Craik, 2002; Craik & Lockhart, 1972). Applied to art, lower-level cognitive judgements include, for example, aspects of subject matter or labeling style while aesthetic judgements would fall into the category of deeper processing, as they not only encompass cognitive elements, but also affective and evaluative processes (Leder et al., 2004; Wang, Cant, & Cupchik, 2016). In line with the depth of processing theory Jacobsen and colleagues (2006) proposed that aesthetic judgements are a subset of higher-level evaluative judgements whereas judgements of symmetry are an example of lower-level descriptive judgements. They attempted to identify the cortical network of aesthetic judgements and conducted an fMRI study comparing the brain activation participants show during aesthetic preference ratings and symmetry ratings. The results supported their claims as aesthetic judgements lead to activation in fronto-median areas (Broadman area 9/10 and inferior precuneus) and the bilateral ventral prefrontal cortex (BA 45/47), both of which have been previously reported to be active during social and moral evaluative judgements on persons and actions (Cunningham et al., 2003; Greene et al., 2004; Moll, de Oliveira-Souza, Bramati, & Grafman, 2002; Zysset, Huber, Ferstl, & von Cramon, 2002). Symmetry judgements on the other hand, elicited activations primarily in several areas related to visuospatial analysis, including superior parietal lobule, intraparietal sulcus, as well as dorsal premotor cortex. This difference in activation might be an indicator for why aesthetic judgements facilitate memory consolidation in contrast to other cognitive tasks. However, another study by Wallraven and colleagues (2009) revealed contradictory results. They also compared a lower-level cognitive task (judging visual complexity in artworks) to aesthetic appeal judgements, but unlike Jacobsen et al. (2006) they did not find any differences in their processing strategies. Using eye tracking they showed that none of the compared eye movement parameters between the two conditions differed significantly. In fact, they found both viewing patterns to be highly similar. These results do not support the depth of processing hypothesis and indicate that aesthetic processing does not differ substantially from other, lower-level cognitive processing. To resolve this apparent contradiction, we conducted this study comparing aesthetic preference ratings to lower-level complexity ratings with the intention of clarifying how special aesthetic processing really is.

## The Present Study

The general aim of our study was to identify unique characteristics of an aesthetic experience via eye movement parameters and to highlight the depth of aesthetic processing with its memory enhancing effects. Similar to Wallraven et al. (2009) we compared higher-level aesthetic judgements to lower-level visual complexity judgements. In line with the depth of processing theory we wanted to show that aesthetic evaluative judgements induce deeper processing compared to descriptive visual complexity judgements. Both tasks require cognitive processing, but aesthetic evaluations have been shown to be more elaborate by also integrating affective elements and inducing aesthetic emotion (Leder et al., 2004). We therefore expected aesthetic processing to produce better memory representations, which we examined by conducting a recognition test. These representations were anticipated to be accompanied by specific eye movement patterns previously shown to be predictive of better memory performance (Kardan et al., 2015). Specifically, we expected aesthetic judgements to induce global processing strategies during encoding reflected in high fixation counts, low fixation durations, and high saccade amplitudes. This pattern has been shown to produce strong memory cues incidentally (Choe et al., 2017; Mills et al., 2011) and is predictive of better recognition (Kardan et al., 2015).

To test our hypotheses the current study was designed as a memory experiment with an encoding phase, a distractor task and a recognition phase. In the encoding phase participants viewed selected artworks from Nadal and colleagues (2010) and rated them either on visual complexity or they evaluated how much they liked them. Unlike Wallraven et al. (2009), who asked participants to rate the image's aesthetic appeal, we chose liking ratings instead. Rating aesthetic appeal may lead participants to evaluate how aesthetically pleasing the image objectively is, whereas the concept of liking may be more suitable in reflecting the subjective experience of the observer. That way we wanted to get as close as possible in capturing the underlying aesthetic experience. While participants performed the rating task their eye movements were recorded. We expected to find the typical global processing strategy in the aesthetic condition and posited that it differs significantly from eye movement patterns in the visual complexity condition. Participants then performed a distractor task to make sure that the encoding phase is not directly followed by the recognition test. The test was planned as a surprise to prevent participants from using memorization strategies during encoding. We expected the representations formed by aesthetic processing to be stronger and more elaborate, which should be reflected in higher recognition rates compared to the visual complexity condition. This would bring further support for the depth of processing theory. It would also build on previous research comparing incidental encoding tasks, highlighting the memory advantages of aesthetic processing. Exploratively we examined within participant idiosyncrasies, participants' behavioral rating data, reaction times during encoding, and eye movement patterns during the recognition phase.

# Method

## Participants

Prior to data collection a power analysis was conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). As we intended to compare the aesthetic condition to the complexity condition, we computed the sample size using the difference between two independent means. With an expected medium effect size of Cohen's *d* = 0.2, a *p* value of .05, and a statistical power of .80 we arrived at a total sample size of *N* = 102. This was increased to *N* = 104, as it facilitated the counterbalancing procedure described below. After the experiment 18 participants had to be excluded for various reasons: the experiment was aborted three times due to calibration errors, eight participants had missing trials, two were categorized as art experts by the means of the VAIAK, one participant misunderstood the task, and finally, four had already expected a recognition test prior to the experiment. This resulted in a remaining sample size of *N* = 86. Aesthetic judgements were given by 46 subjects, the other 40 performed the visual complexity rating task. Overall mean age was 21.37 (*SD* = 3.09) with 72 female and 14 male participants; all participants received course credit for their participation.

## Materials

**Apparatus.** The eye tracking experiment was constructed using the software program Experiment Builder 1.6.121 and Eye Data Viewer 1.10.123 (SR research Ltd., Ontario, Canada). It was then conducted on a Windows XP computer using a SR research Ltd. EyeLink 1000 desktop mounted eye tracker (http://www.sr-research.com/EL_1000.html). Eye movement data were recorded by a second computer running ROMDOS. A head- and chinrest was placed on the participant's desk at a distance of 57cm from the screen and the chair was put into a fixed position, allowing participants to merely adjust its height. After the assessment of each participant's ocular dominance, recording of eye movements was limited to the dominant eye only. The experiment started with a calibration and validation procedure ensuring eye position errors of less than 1°. Instructions and stimuli were presented at the center of the screen with a resolution of 1024×768 pixels and a consistent light gray background color. Participants' responses were given on a standard computer keyboard placed easily accessible in front of them.

**Stimuli of the eye tracking experiment.** For the current study we chose 60 images of the stimulus set from the third phase of Nadal and colleagues' (2010). These images had already undergone a careful selection process. Out of a pool of over 1,500 digitalized images Nadal et al. (2010) selected only relatively unknown pieces of art to avoid the impact of familiarity. They excluded images depicting close-up human figures and faces, as well as scenes that might potentially elicit strong emotional responses. To control for undesired psychophysiological influences, they then homogenized the images removing all signatures, using the same size of 9×12 cm, a resolution of 150 ppi, standardizing the color spectrum, and adjusting luminance to 370-390 lx. All images that weren’t modifiable according to these standards were discarded. After adjusting them into a consistent format they made sure that images varied on three dimensions: complexity, artistry, and style. In terms of complexity they differentiated between three levels (high, low, and medium). High in complexity means that these images may contain a high amount and variety of elements that may be disorganized and/or asymmetric. In terms of artistry they selected images that were reproductions of catalogued pieces created by renowned artists and hence classified as artistic. Non-artistic pictures included postcards, photographs of landscapes, artifacts, urban scenes, and so on. Finally, in terms of style a variety of schools such as realism, cubism, impressionism, and more, was selected to differentiate between abstract and representational images. Images were classified as representational in the presence of explicit content, and abstract in its absence. This resulted in a total of 800 images out of which we selected 60 artistic images varying in complexity and style. We then searched for 60 images more that matched the 60 originals in content and style. Our intention was to make the recognition test more challenging by pairing up each painting with a very similar looking twin image from the same artist and time period (see Figure 1). A pre-study was conducted to evaluate whether the 60 pictures from Nadal et al. (2010) and their corresponding twin images were sufficiently similar and yet, not impossible to distinguish. It was concluded that the selected stimulus set was well suited for application in a recognition test. The test was neither too difficult for participants to handle, nor too easy for them to score all points. The final set therefore consisted of 120 artworks with varying style and complexity.

*Figure 1.* Example images with their corresponding pairs

Because each image had its counterpart it was not possible to simply randomize all stimuli. We thus applied a counter-balancing procedure to make sure each participant received 30 images in the recognition test that they had already seen during encoding, but also 30 new distractor images that replaced the other half of the original images from T1, while additionally making sure to prevent sequencing effects. We therefore divided the stimulus set of 120 images into 2 groups (A = original images, B = twin images) which were further divided into 4 subgroups of 30 stimuli each: A_1, A_2, B_1, and B_2. Images were randomized once within each subgroup and then maintained a fixed position, which is necessary to counterbalance the image sets. Figure 2 illustrates the randomization and counterbalancing procedure.

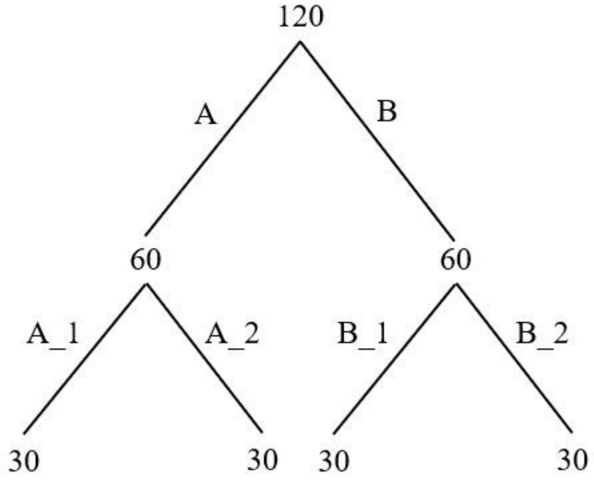

*Figure 2.* The tree diagram illustrates the distribution of images used in the experiment. (A) stands for the original images, (B) stands for the corresponding twin images, (A_1) stands for the first half of the original images, (A_2) for the second half of the original images, (B_1) for the first half of the twin images corresponding to the originals in A_1, and (B_2) for the second half of the twin images corresponding to the originals in A_2.

The application of the procedure is illustrated in this example: during the first part of the experiment Participant X received 30 images of the original images in Group A and 30 twin images from Group B. It is important to keep in mind that it would not be possible for the participant to receive A_1 and B_1, because those are 30 of the originals from Group A and 30 of their corresponding twin pairs. Only the following four combinations are therefore possible: (A_1+B_2), (B_2+A_1), (A_2+B_1), (B_1+A_2). In our example Participant X viewed A_1 and B_2. In the second part of the experiment Participant X received 30 “old” images that they had already been given, but also 30 “new” images that they had not seen in the first part yet. That means that 15 images out of each image pool A_1 and B_2 were presented for a second time, along with 30 “new” images resulting in the required 60 images. In this example 15 images were drawn from B_1 and 15 from A_2, because these are the two image sets that the participant has not been given in the first part but the corresponding counterparts to the ones that Participant X has already seen. This procedure results in a recognition test in which the participant is as asked to indicate whether they recognize the images. In total, each participant saw 120 images (60 in part one and 60 in part two). In part two, however, half of the pictures repeat themselves, which means that each participant only viewed 90 images of the total stimulus set. Further information about the recognition test is given in the following sections.

**Stimuli distractor task.** The distractor task was performed on the computer right after Part one of the eye tracking experiment and was then directly followed by Part two. Images for the task were created using the same method as Gartus and Leder (2013) who constructed design patterns with varying levels of symmetry. Figure 3 shows example images of a total of 105 stimuli.

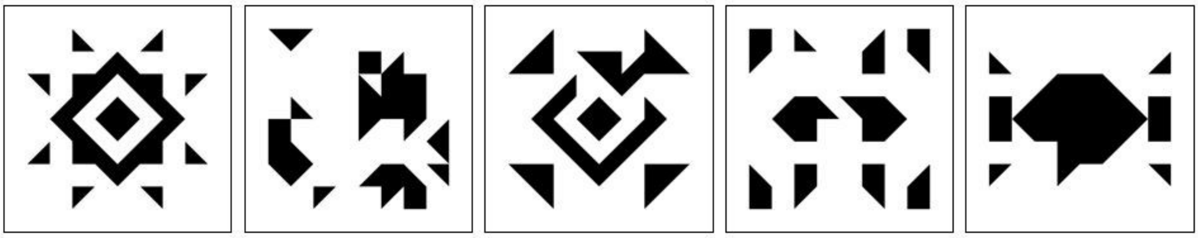

*Figure 3.* Example images of the distractor task

## Design and Measures

The current study consists of an eye tracking experiment followed by the Viennese Art Interest and Art Knowledge Scale (VAIAK; Specker et al., 2018). The experiment comprised two main parts: an encoding phase at the beginning and a recognition phase at the end. Between these parts, participants were given a distractor task which serves merely as a buffer between the two main phases.

In the first phase there are two different sets of instructions for the same task (viewing artworks), which were used as independent measures. Participants were asked to either give liking or complexity evaluations on artworks. These two groups were compared in their performance on a subsequent recognition test. Essentially, this recognition test resulted in four outcomes: correct recognition of the a previously seen image (hit), correct rejection of a distractor (correct rejection), false recognition of one of the distractors (false alarm), and false dismissal of a previously shown image (miss). As a measure of recognition performance *dprime* or *d'* from Signal Detection Theory (Stanislaw & Todorov, 1999) was used. It is calculated by subtracting the *z*-score of false-alarm rates (FA) from the *z* score of hit rates (H):

$$d' = z(H) - z(FA)$$

The higher the score the better the performance. A more detailed description of *d'* and its derivation can be found in Stanislaw and Todorov (1999). The recognition test resulted in behavioral data used to measure the outcome of the manipulation in phase one. The security rating that was given during the recognition test and the liking rating during encoding were also used for exploratory analysis.

Psychophysiological data was provided by the recorded eye movements in both phases. In this study, we focused on three eye movement parameters: (a) fixation count (the number of fixations), (b) fixation duration (the time spent looking at the same spot) and (c) saccade amplitude (the length of the jump from one fixation to the next). Differences in eye movements of the two groups during the encoding phase were intended to be used as predictors of performance in Phase two. Recorded eye movements from Phase two were additional outcome variables and were used for exploratory comparisons of eye movements from both phases across participants.

Finally, the knowledge and interest scale of the VAIAK were used as an exclusion criterion. Participants who scored over 50% on the art knowledge scale or 85% on the art interest scale were excluded from analysis, as art experts focus more on style and hence process art differently (Augustin & Leder, 2006; Leder et al., 2012; Nodine et al., 1993) leading to different eye-movement patterns (Pihko, 2011; Vogt, 1999; Zangemeister, Sherman, & Stark et al., 1995).

Another exclusion criterion was the debriefing question at the end of the experiment. Participants were asked whether they had expected to perform a recognition test *before* taking part in the experiment. This was crucial to asses as prior expectation of a potential recognition test may alter eye movements during encoding of images. Instead of rating visual complexity and aesthetic preference, participants may change their strategy to memorization which would highly bias their eye movements as has been shown by studies comparing these tasks (Choe et al., 2017; Kardan et al., 2015; Mills et al., 2011). Therefore, if the answer to the final debriefing question was "yes", they had to be excluded from analysis.

## Procedure

After initially giving informed consent, participants provided answers about their gender and age and their ocular dominance was assessed. Then the light was dimmed, and they were seated in front of the computer. To ensure minimal motion, participants were asked to place their heads on a head/chin rest and told to try to move as little as possible throughout the experiment. A brief explanation of the eye tracking system and the procedure was given. The eye-tracking device was then adjusted to measure the dominant eye only. At the start of the experiment participants' eye movements were calibrated and validated. This process could potentially be repeated during the experiment before each trial, in case the eye tracking system was not able to detect participants' pupils properly due to unexpected movements or other reasons. Finally, the experiment was started in the version that was indicated on protocol.

**Experiment.** The eye tracking experiment started with an encoding phase (T1), directly followed by the distractor task and concluded with the recognition phase (T2). Before the start of the experiment participants were unaware of having to perform a recognition test later in T2. Figure 4 illustrates the structure of the eye tracking experiment which is described in detail below.

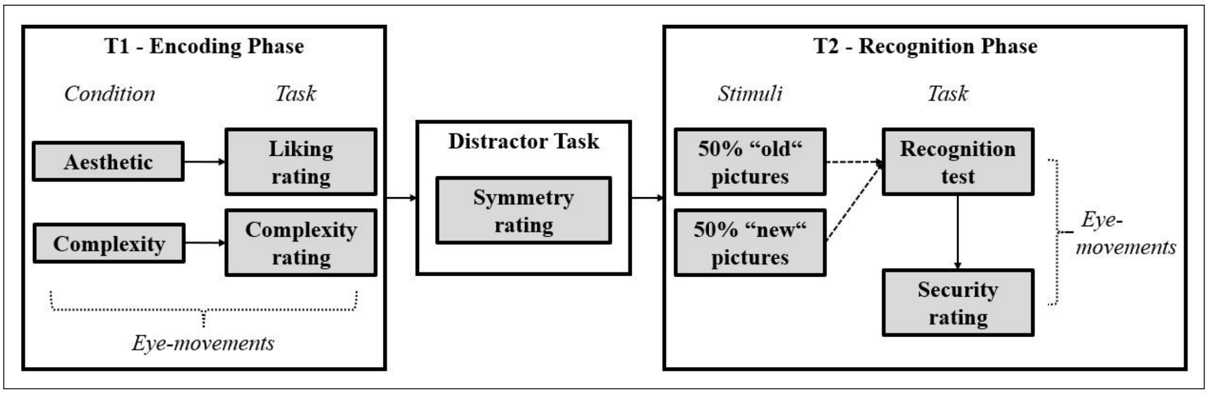

*Figure 4.* Structure of the eye tracking experiment

***T1 – encoding phase.*** At the very beginning of the experiment participants were randomly assigned to one of two groups and it was made sure that both groups were equal in size. Both groups viewed 60 images of artworks for five seconds after which a 7-point-Likert-scale appeared on screen, instructing them to rate the image (see Figure 5). The viewing time was limited to five seconds as empirical evidence suggests that two seconds already suffice to capture the essence of an image (Locher et al., 2007) and because eye movements tend to be relatively stable within five seconds (Castelhano et al., 2009; Cela-Conde et al., 2013; Mills et al., 2011). In Version A participants evaluated how much they liked them (ranging from 1 = dislike very much to 7 = like very much) which is illustrated in Figure 2. In Version B participants rated how complex they thought the image was (ranging from 1 = not complex at all to 7 = very complex). Here, they were additionally told to focus on the image's visual complexity rather than the complexity of its creation. Before each trial participants performed a fixation check. Throughout this phase eye movements were recorded.

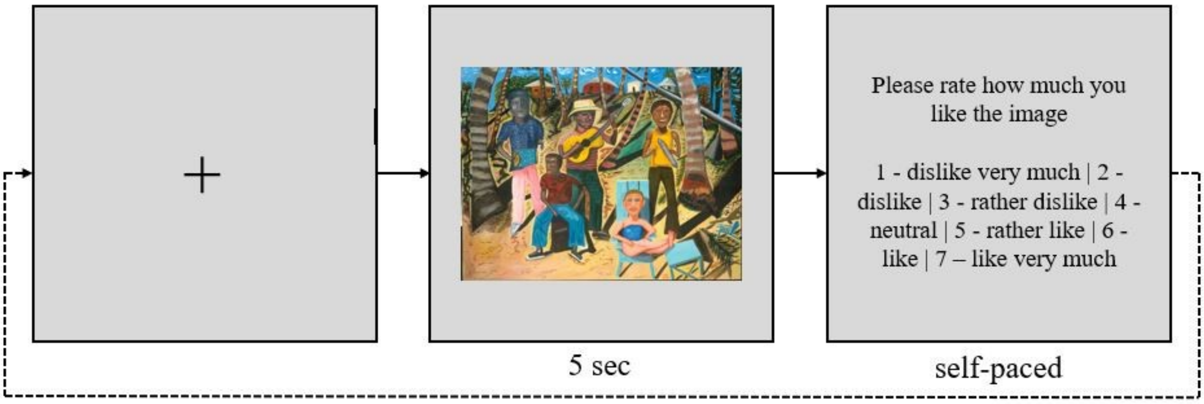

*Figure 5.* Example trial of the encoding task

***Distractor task.*** Participants viewed 105 images of design patterns and were asked to rate their symmetry on a 5-point-Likert-scale (1 = asymmetric, 5 = symmetric). Both, the image and the scale appeared simultaneously on screen as shown in Figure 6. Trials were self-paced and ended after the rating was given. This task was necessary so that the encoding phase was not directly followed by the recognition phase, thereby preventing participants from consciously or unconsciously repeating and memorizing the images from T1. The recording of eye movements was not necessary during this part.

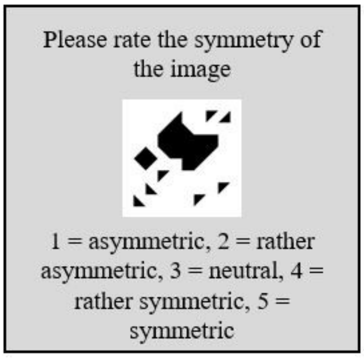

*Figure 6.* Example distractor task

***T2 – recognition phase.*** In the last phase of the experiment participants performed a surprise recognition test (see Figure 7). As in the encoding phase they once again saw images of artworks on screen for five seconds. They were given a total of 60 images: 30 “old” images that were part of the original stimulus set in T1, and 30 “new” images that looked very similar to the stimuli from T1. Participants were now asked to indicate for each image whether they had seen it before in T1 or not, using a Yes/No answer format. This was followed by a certainty rating in which participants rated how sure they were about their answer on a 7-point-Likert scale (ranging from 1 = not sure at all to 7 = very sure). Both ratings were self-paced. After the rating was given, a new trial started with a fixation cross at the beginning. Images were randomized and counterbalanced as explained in the Materials section of this paper. Like in the first part, eye movements were recorded in this phase as well.

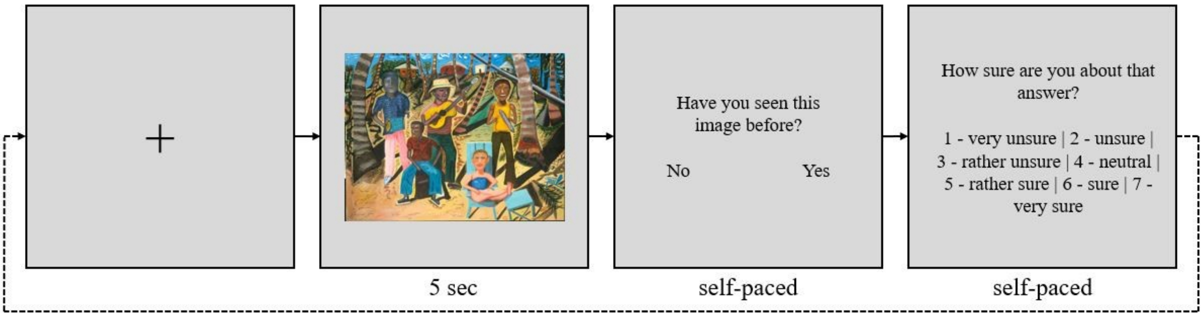

*Figure 7.* Example trial of the recognition test

**VAIAK.** Following the eye tracking experiment, participants were given the VAIAK booklet containing the images and the questionnaire which was filled out in paper-pencil format. It included questions about the participants' art interest and knowledge which was used to differentiate between art experts and lays.

Finally, participants had to answer the final debriefing question whether they had already expected to perform a recognition test before taking part in the experiment. Their answer was noted on the protocol. All in all, the experiment lasted 70-80 minutes. Before leaving, participants were informed about the background of the study and were thanked for their participation.

## Data Processing

Before analysis raw eye movement data were preprocessed in MS Excel. As previously explained, eye movements were recorded during the encoding and the recognition phase. Only eye movements during image viewing trials were used, the rest was discarded. Additionally, all eye movements after 5.000 milliseconds on each trial were excluded. To calculate median fixation durations and fixation counts, the first fixation of each trial was excluded. This was due to the fixation cross prior to each trial guiding participants' first fixation inevitably to the center of the stimulus and therefore biasing all first trial fixations. To calculate saccade amplitudes, exclusion of the first fixation was not required as a standardized starting point has no influence on the amplitude of the saccade. However, as viewing time was limited to five seconds with all eye movements exceeding that limit being discarded, it was necessary to disregard the final saccade of each trial, because it may have been interrupted and therefore the actual length of the final saccade could have been biased.

In order to conduct statistical analyses, the remaining eye movements were used to compute fixation count, fixation duration, and saccade amplitude. First, the median of fixation durations and saccade amplitudes were calculated for each trial in the encoding- and recognition phase, respectively. Then the median was calculated for each person in each phase. The reason to use the median here instead of the mean is that fixation durations and saccade amplitudes usually show rather right-skewed distributions (Antes, 1974; Pannasch et al., 2008; Unema et al., 2005). In terms of fixation count, the total number of fixations on each trial was counted. Finally, the mean was computed for each person in each phase. The resulting mean fixation count, median fixation duration, and saccade amplitude were now ready to be subjected to calculations and analyses.

## Hypotheses and Analyses

All calculations and statistical analyses were conducted using the software program IBM SPSS Statistics (Version 25). Prior to data collection all three main hypotheses were specified, and an exact layout of the data analysis was outlined:

**H1.** *Aesthetic processing is expected to induce global processing strategies during encoding with higher fixation counts, lower fixation durations, and larger saccades in the aesthetic condition compared to the complexity condition.*

The first hypothesis concerns eye movements in the encoding phase, as we expected the two groups to show different patterns. Specifically, a global processing strategy with an increased mean fixation count, lower median fixation duration and larger median saccade amplitudes was expected in the aesthetic condition. To compare the fixation count among the two groups we used a one-sided independent *t*-test. As mentioned before, we expected an asymptotic distribution of fixation duration and saccade amplitude. Therefore, we compared the aesthetic and complexity group using the non-parametric Mann-Whitney test for these parameters.

**H2**. *The differences in eye movements associated with an aesthetic orientation lead to an enhanced representation reflected in higher recognition rates.*

As outlined in the introduction, we expected liking judgements to produce better representations of the artworks, which should lead to better recognition accuracy in Part two of the experiment. To test this hypothesis, *d'* was used as a measure of recognition accuracy and was expected to be higher in the aesthetic condition. After calculating *d'* for the liking evaluation group and the complexity rating group respectively, a one-sided independent *t*-test was conducted.

**H3**. *Recognition accuracy can be predicted by certain characteristics during encoding.*

Finally, showing the described characteristics during the encoding phase was expected to be related to the performance in the recognition test. The characteristics are condition, fixation count, fixation duration, and saccade amplitude. With these predictors we wanted to compute a multiple linear regression with *d'* as outcome variable. If the fixation count is higher, the fixation duration lower, the saccade amplitude higher, and the condition is to make liking evaluations during the encoding phase, then higher recognition rates (*d'*) were expected.

**Exploratory analysis.** As eye movements were recorded in the encoding and recognition phase respectively, we wanted to explore if and how parameters changed from T1 to T2. Additionally, we intended to inspect idiosyncrasies within participants. Previous research indicates that there is high inter-subject variability, but also high intra-subject consistency when it comes to viewing patterns (Foulsham et al., 2012; Quian Quiroga & Pedreira, 2011), which we therefore expected from our sample as well. Another psychophysiological measure that we wanted to inspect were the reaction times in T1. Consistent with the depth of processing hypothesis Wang et al. (2016) showed that aesthetic judgements involved "deeper" processing compared to non-aesthetic judgements reflected in longer reaction times during encoding of artworks. However, due to the restricted validity of reaction times in our study resulting from the imposed viewing time of 5 seconds for each trial, we could not make a clear prediction about the outcome and examined the results exploratively instead.

Participants' behavioral data was also analyzed more closely. We were interested in how much participants liked or disliked the images in general. Then, we investigated whether high and/or low liking ratings were connected to recognition accuracy. Finally, we examined the security ratings in T2. We wanted to know how sure participants were in general about their answers during the recognition test and whether high security was associated with better performance.

# Results

First, we inspected how the spatial and temporal eye movement parameters developed throughout the 5 second viewing period in the encoding phase. Figure 8 shows median fixation duration (a) and median saccade amplitude (b) over the course of 5s. We used the median of overall fixation durations and saccade amplitudes separated by condition. The figure illustrates quite well the expected skewed distributions. In line with previous studies we can also observe the asymptotic trend of fixation durations and saccade amplitudes which first increase and then stabilize after about two seconds of viewing time. These graphs serve descriptive purposes only and were not used for statistical analyses.

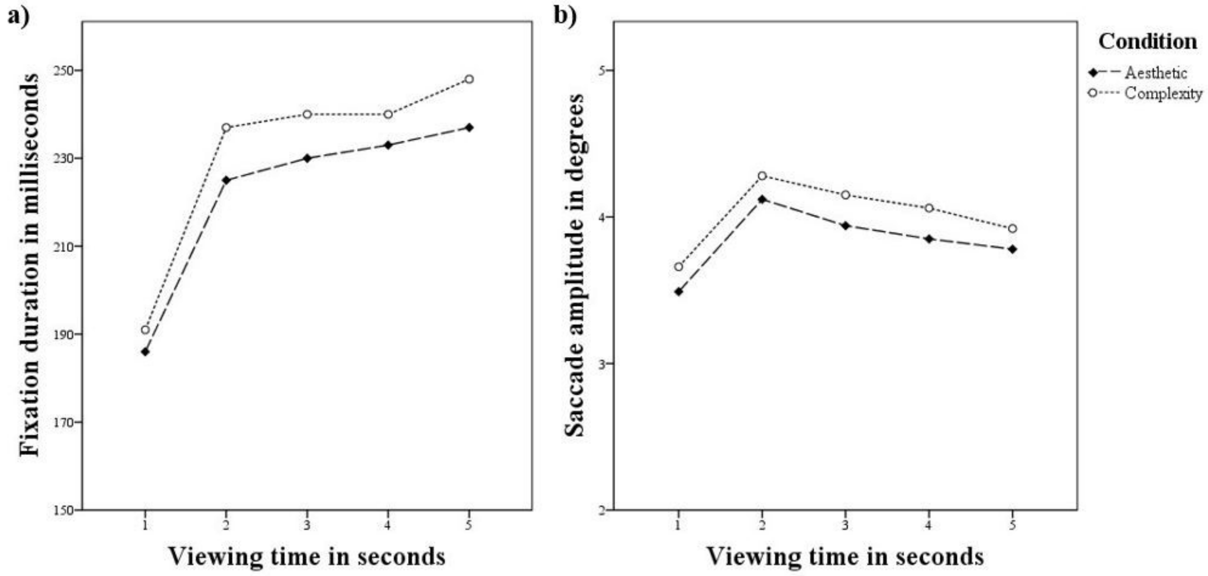

*Figure 8.* Graph a) shows median fixation durations, graph b) median saccade amplitudes across overall fixation durations and saccade amplitudes over the time span of 5 seconds separated by condition.

## Confirmatory Analysis

**H1.** We tested whether eye movement parameters differed significantly between the aesthetic and the complexity condition. Table 1 contains overall mean and median eye movement measures separated by T1 and T2 and for each condition respectively. Mean fixation count, median fixation duration, and saccade amplitude were calculated for each trial, then averaged across participants for T1 and T2, and finally, aggregated across condition. Although at first glance the data seem to fit to the hypothesis with fixation count to be higher, fixation duration to be lower and saccade amplitude to be higher in the aesthetic condition during encoding, the respective standard deviations and interquartile ranges cast doubt on the significance of these differences.

Table 1

*Eye Movement Parameters in T1 and T2*

| | Aesthetic | | Complexity | |
|---|---|---|---|---|
| Features | T1 | T2 | T1 | T2 |
| Fixation count | 16.13 (1.58) | 16.04 (1.67) | 15.75 (2.12) | 15.68 (2.13) |
| Fixation duration | 228.5 (34.88) | 226.5 (43.13) | 235.25 (47.5) | 226.5 (28.5) |
| Saccade Amplitude | 4.01 (0.79) | 4.25 (0.83) | 3.88 (0.95) | 4.32 (1.2) |

*Note*. Fixation durations were measured in milliseconds, saccade amplitudes were measured in degrees of visual angle. The sample size for the aesthetic condition was 46, and 40 for the complexity condition. Means and standard deviations are given for the fixation count. Fixation durations and saccade amplitudes are depicted in medians and interquartile range.

The one-sided independent *t*-test comparing the fixation count between the aesthetic ($n = 46$) and the complexity condition ($n = 40$) was not significant, $t(84) = 0.93$, $p = .353$, $d = .19$. Fixation durations and saccade amplitudes were compared using a Mann-Whitney test. Fixation durations did not differ significantly between the aesthetic and the complexity group, $U = 746.5$, $z = -1.5$, $p = .133$, $r = -.16$. Similarly, saccade amplitudes in the aesthetic condition were not significantly different from saccade amplitudes in the complexity condition, $U = 855.5$, $z = -0.56$, $p = .577$, $r = .06$. Our results do not support our first hypothesis, as none of the measured eye movement parameters differ between the aesthetic and the complexity group.

**H2.** We then checked whether processing was deeper in the aesthetic condition, which should be reflected in higher recognition rates. Although mean *d'* rates are slightly higher in the aesthetic ($M = 1.32$, $SD = 0.35$, $n = 46$) compared to the complexity condition ($M = 1.24$, $SD = 0.33$, $n = 40$), this difference was not significant, $t(84) = 0.99$, $p = .321$, $d = .21$. This shows that participants who rated how much they liked images were not better at recognizing them in a subsequent recognition test compared to participants rating them in terms of visual complexity.

**H3.** As all our previously set hypotheses turned out counter our predictions, the calculation of a multiple linear regression predicting recognition accuracy with group affiliation and eye movement parameters was redundant. We therefore moved on to examining the data exploratively in detail.

### Exploratory Analysis

**Eye tracking data.** First, we wanted to know how eye movement parameters changed between the first presentation of images and the second exposure when participants saw the stimuli again during the recognition test. Means and medians of the measured eye movement features in each condition separated by T1 and T2 are depicted in Table 1. We compared each eye movement parameter in T1 with the same parameter in T2 for each condition respectively. Paired-sample *t*-tests comparing mean fixation count in T1 and T2 resulted non-significant, $t(45) = 0.68$, $p = .499$, $d = .05$ for the aesthetic condition, and $t(23) = 0.48$, $p = .633$, $d = .03$ for the complexity condition. Due to the skewed distribution of fixation duration and saccade amplitude, we compared these features using the non-parametric Wilcoxon signed ranks test for paired samples. Results and effect sizes are depicted in Table 2. Fixation durations decreased from T1 to T2 in both the aesthetic and the complexity condition, although this was only significant for the complexity condition. Saccade amplitudes significantly increased from T1 to T2 in both conditions.

Table 2

*Statistics of Eye Movement Comparisons between T1 and T2*

| Comparison | *z* | *p* | *r* |
|---|---|---|---|
| Aesthetic (*n* = 46) | | | |
| fixation duration T1 vs. T2 | -2.39 | .016 | -.35 |
| saccade amplitude T1 vs. T2 | -3.51 | <.001* | -.52 |
| Complexity (*n* = 40) | | | |
| fixation duration T1 vs. T2 | -3.73 | <.001* | -.59 |
| saccade amplitude T1 vs. T2 | -3.18 | .001* | -.50 |

*Note*. Fixation duration and saccade amplitude were compared with a Wilcoxon signed ranks test. Effect sizes are given in *r*. As six pairwise comparisons were conducted, *p* values were compared with the Bonferroni corrected α = .008.

$^{*}p < \alpha_{\text{corrected}}$

Then we examined potential idiosyncrasies within participants. To find out how consistent participant were in their eye movement patterns, we correlated each parameter from T1 with the same parameter from T2. All features were highly correlated, with fixation count $r = .89$, fixation duration $r = .90$, and saccade amplitude $r = .82$, indicating that participants were consistent in the way they looked at the images (Figure 9). This means that subjects tended to show a very similar pattern in T1 and T2. For example, if a participant made many fixations with short durations and large saccades during encoding, it is very likely that he/she showed the same pattern during recognition. Like previous research that reported high inter-subject variability but also high within-subject similarity of eye movement patterns, our data support these findings.

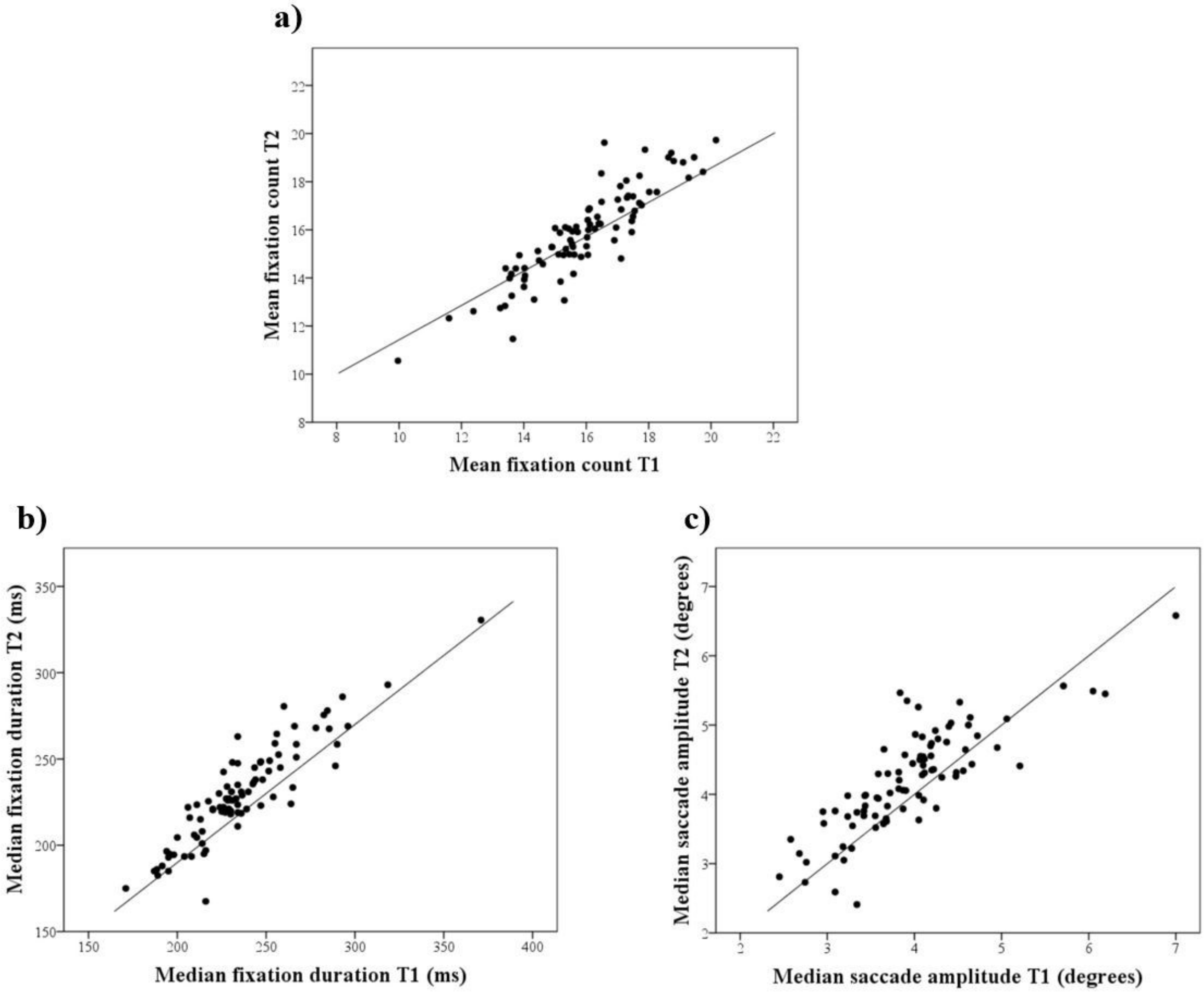

*Figure 9.* The scatterplots show the correlations of the average a) fixation counts, b) fixation durations, and c) saccade amplitudes per participant in the two phases T1 and T2.

**Reaction times.** We investigated whether groups differed in their reaction times (RT) during encoding. In line with the depth of processing hypothesis we assumed that the aesthetic group might have overall longer RTs compared to the complexity group, as aesthetic evaluations are reportedly more elaborate than merely cognitive ones. However, the *t*-test resulted to be non-significant $t(84) = -0.27$, $p = .788$, $d = .06$, with means of $M = 2424$ ms in the aesthetic and $M = 2462$ ms in the complexity condition (see also Figure 10). This is another indicator that liking ratings did not elicit deeper processing, as both groups took equally long for their evaluations.

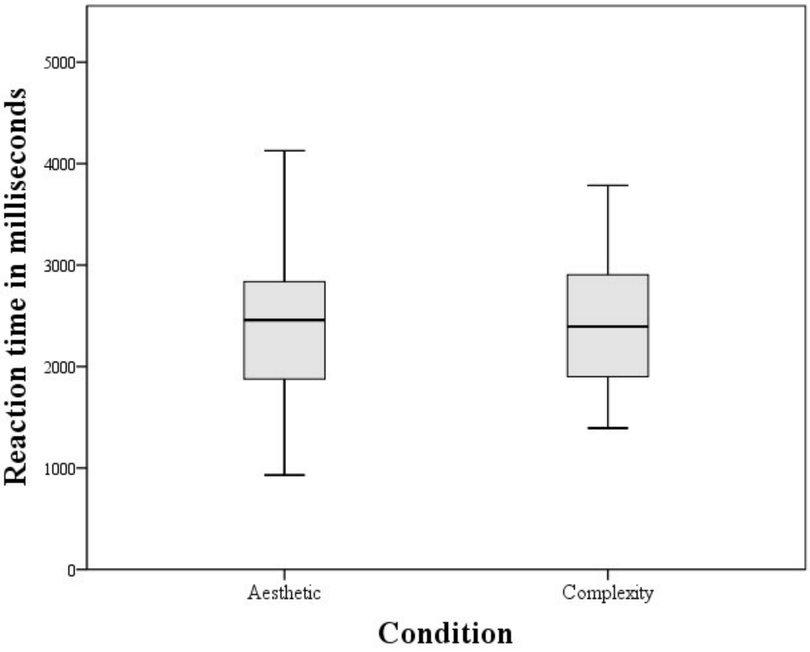

*Figure 10.* This boxplot shows mean reaction times for each condition respectively. Participants in the aesthetic group did not show longer reaction times than participants in the complexity group.

**Liking ratings.** In the aesthetic condition participants were asked to rate how much they liked each image. We now investigated whether this rating was related to recognition accuracy. One possibility would be that subjects remembered pictures better if they liked them more. Another option would be that images that were very liked or disliked stayed in memory better than images that participants were rather indifferent to. To find out more about this relation we calculated the hit rate for each of the seven rating categories. In this case we only used hits because we had to be certain that participants viewed the same image twice, once in T1 and then again in T2. Including correct rejections would bias the outcome as participants rated a different image in T1 than was shown in T2. The rating given in T1 is therefore irrelevant if the image in T2 is not the same as recognition is logically not possible. Therefore, this analysis only includes the 30 stimuli that were equal in T1 and T2, excluding the 30 "new" distractor stimuli in the recognition phase. Due to their heteroscedastic and non-normal distribution we then compared the hit rates of each rating category using the non-parametric Kruskal-Wallis test, which resulted to be significant, $H(6) = 23.88$, $p = .001$, $\varepsilon^2 = .09$. Mean ranks ranged from min = 105 in rating category 3 to max = 170 in rating category 7. Figure 11 illustrates the distribution of hit rates per category using the mean ranks of the Kruskal-Wallis test.

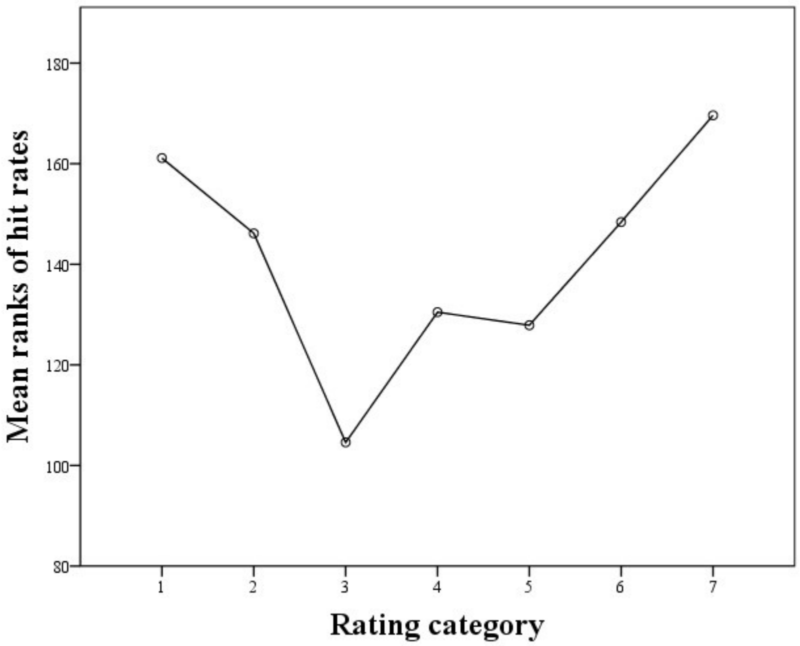

*Figure 11.* Depicted are mean ranks of hit rates for each like rating category. The trend indicates that more extreme ratings in both directions lead to higher hit rates. This graph serves descriptive purposes only.

The trend suggests that images which were rated either very high (category 6 and 7) or very low (category 1 and 2) on the liking scale were subsequently better recognized. To test this assumption, we conducted 21 pairwise comparisons between all rating groups using the Mann-Whitney test. After applying the Bonferroni-Holm method to control the familywise error most of the previously significant results turned out to be non-significant (see Table 3), except for the comparison between rating category 3 and 7; $U = 372$, $z = -3.74$, $p < .001$, $r = -.43$, with $\alpha_{\text{corrected}} = .002$.

Table 3

*Selected Pairwise Comparisons Between Rating Categories*

| Comparison Rating Category | | $\alpha_{\text{corrected}}$ | $p$ | $r$ |
|---|---|---|---|---|
| 3 | 7 | .002 | .000185* | -.43 |
| 5 | 7 | .003 | .0025 | -.35 |
| 1 | 3 | .003 | .0033 | -.35 |
| 3 | 6 | .003 | .005 | -.3 |
| 4 | 7 | .003 | .008 | -.3 |
| 2 | 3 | .003 | .02 | -.25 |
| 1 | 5 | .003 | .027 | -.26 |
| 3 | 5 | .004 | .047 | -.21 |

*Note.* Depicted are pairwise comparisons that yielded significant results after the Mann-Whitney test $p < .05$. The other 13 comparisons were not significant. After correcting $\alpha$ using the Bonferroni-Holm method only one comparison resulted to be significant.

*$p < \alpha_{\text{corrected}}$

Finally, we took a more general look at the liking ratings. We wanted to know how images were generally perceived by our sample and whether they were liked or disliked on average. The median liking rating of a total of $N = 2769$ ratings was $Mdn = 4$, which corresponds to the answer "neutral". The interquartile range ($IQR = 3\text{-}5$) reveals that 50% of the ratings lie between rating category 3 and 5, indicating that overall participants rated images neither very positive nor very negative.

**Security rating.** During the recognition test in phase 2 participants were asked whether they remembered the images from phase 1. Additionally, they had to rate how sure they were about their answer on a scale from 1 to 7. We now wanted to know how this security rating is linked to recognition accuracy. Intuitively it seems logical that being certain about having seen the image before or not should result in high hit and correct rejection rates. And indeed, the Spearman correlation between the security rating and the rate of hits and correct rejections per category yielded a significant moderate correlation, $r = .42$, $p < .001$, indicating that higher security ratings coincide with better subsequent recognition.

Finally, we examined how sure participants were in general about their answers in the recognition test. The median of a total of $N = 5160$ security ratings was $Mdn = 6$ ($IQR = 5\text{-}7$), which reveals that overall participants felt quite secure in their performance on the recognition test. We also examined whether there were differences between groups in their subjective security. As a result of deeper processing it would be reasonable to hypothesize that participants in the aesthetic condition might have felt surer about their answers in the recognition test. However, the Mann-Whitney test comparing the aesthetic to the complexity group was not significant, $U = 3262974.5$, $z = -.969$, $p = .333$, $r = .04$, indicating that both groups were equally sure about their performance.

## Discussion

What makes aesthetic processing special? This was the main question of interest for this study in which we tried to identify specific characteristics of aesthetic experiences. Previous research indicates that certain eye movement patterns during an aesthetic mode of processing result in memory enhancing effects (Choe et al., 2017; Kardan et al., 2015; Mills et al., 2011). Specifically, a global processing strategy with high fixation counts, low fixation durations, and large saccade amplitudes is characteristic for aesthetic judgement tasks and has been shown to positively affect recognition rates due to the detailed representations it creates.

We therefore expected this specific viewing pattern from participants in the aesthetic judgement condition and compared it to another group that rated the same images on their visual complexity. To clarify whether these eye movement parameters indeed induce stronger representations we additionally compared the recognition rates of both groups and expected participants in the aesthetic condition to achieve higher scores. However, all our hypotheses had to be rejected, as neither recognition performance nor eye movement patterns differed significantly between the two groups. These results replicate the findings of Wallraven et al. (2009), who found eye movement patterns do be highly similar in participants rating aesthetic appeal and participants rating visual complexity of artworks. Instead of aesthetic appeal ratings we used liking ratings as they may be more suitable in representing and activating the subjective aesthetic experience rather than triggering objective judgements on aesthetic appeal. However, the lack of difference in the results of both studies indicates that both tasks are essentially equal in their outcome.

The fact that viewing patterns were very similar in the aesthetic and the complexity group alone does not necessarily contradict the depth of processing hypothesis. There are numerous other eye movement parameters that can be analyzed complementing the ones we investigated in this study. Scanpath comparison, for example, often provides a very extensive picture of how exactly the image was observed. It allows the examination of the exact fixation location and the temporal fixation sequence. A close-up exploration of scanpaths in both groups may therefore reveal differences in viewing patterns that are not observable with the eye movement parameters that we have focused on here. However, the features we measured do tell us something about the representations they have formed. That fixation count, fixation duration and saccade amplitude were not significantly different in both conditions suggests that the representations that were created during encoding were also highly similar, if not equal. And indeed, the recognition test revealed that there were no significant differences in recognition accuracy between the aesthetic and the complexity group. This shows that higher-level liking evaluations on artworks did not lead to memory advantages via enhanced representations compared to lower-level complexity judgements which contradicts our expectations and refutes the depth of processing hypothesis.

Not only the measured eye movement parameters but also the lack of difference in reaction times during encoding in both groups further undermines the depth of processing hypothesis. Unlike Wang and colleagues (2016) who found longer reaction times in the aesthetic condition compared to a visual search task, our data yielded contradictory results. From a depth of processing perspective liking judgements should be more exhaustive and encompass both cognitive and affective components leading to longer elaboration times, as was shown by Wang et al. (2016). That our data do not support this finding may have two reasons. The first reason could be task similarity. Instead of visual search we used visual complexity judgements as another incidental encoding task. By choosing two rating tasks the processing time for an evaluation may be highly similar or even equal, which would explain why Wang et al. (2016) found a difference in reaction times and we did not. However, even if rating tasks tend to require similar processing times, there should still be variance among this category according to the depth of processing theory. If aesthetic processing is as deep as postulated, then the time it takes to arrive at an aesthetic judgement should be significantly different from lower level cognitive judgements. This brings us to the second explanation that highlights an important technicality. In contrast to Wang et al. (2016) our study participants were given a fixed time limit of five seconds. Only after these five seconds had passed, participants were able to rate the image. In contrast, participants in the study of Wang et al. (2016) were free to choose how much time they spent for each rating. This is a crucial aspect because it weakens the informative value of reaction times in our study. If participants had already made up their mind before five seconds they still had to wait until after the time had passed to rate the image. This obviously constricts the validity of reaction times as a measure, which is why no hypothesis about differences between the two groups was included in the confirmatory analysis. In sum, even though the similarity of the task may have led reaction times to be more alike, the main reason for the missing effect was probably caused by the study's design.

Whether or not the results of the reaction times in our study are credible enough to count as evidence against the depth of processing hypothesis is disputable. What is certain, however, is that neither the measured eye movement parameters nor the measure of recognition accuracy differed significantly between the two groups. One explanation for the lack of difference between the two conditions is that the mere act of making a rating may induce similar or even equal processing strategies independent of rating type. Murty et al. (2015) arrived at a similar conclusion after they found memory enhancing effects by simply making a choice. This would explain why there was no difference in recognition accuracy between the two conditions. If the act of making a rating itself leads to better memory, then making aesthetic judgements is not an adequate measure to induce and capture an aesthetic experience. Maybe the concept of liking is simply too shallow to evoke and represent the depth that aesthetic experiences entail. Whether this conclusion is valid or not can be examined by replicating one of the studies using aesthetic judgements and replacing them with any other kind of rating. If the outcome is the same, then the resulting memory enhancing effects may not be attributed to an aesthetic mode of processing but rather to the rating task itself. This highlights again the importance of comparing incidental encoding tasks of the same type. Visual search has been used extensively alongside aesthetic judgements to highlight the advantages of aesthetic processing on memory (Choe et al., 2017; Kardan et al., 2015; Mills et al., 2011; Wang et al. 2016). But if making ratings independent of rating type facilitates better memory consolidation, it is not clear what kind of mechanism lies behind the advantages of giving a rating and why. The only thing that would be certain is that aesthetic experience could thereby not be the underlying construct.

Another potential explanation for the congruent outcomes of both conditions could be that aesthetic processing is simply not as special as assumed. Wallraven and colleagues (2009) already emphasized that aesthetic and complexity judgements both evoke global search strategies which may have led to similar representations of the artworks, which is also what our data revealed. The viewing pattern similarity questions the uniqueness of aesthetic processing and implies that aesthetic judgements essentially do not differ from other rating tasks. However, in an fMRI study Jacobsen and colleagues (2006) were able to demonstrate differences in brain area activation while making aesthetic judgements compared to symmetry ratings. In a next step these tasks could be compared using eye tracking to examine whether these differences in activation are manifested in the participants' eye movement patterns as well. Conversely, it would be interesting to conduct an fMRI study comparing aesthetic and complexity judgements. If both tasks induce similar or equal activation patterns in the brain, it can be concluded that aesthetic and complexity judgements are governed by the same mechanism. If different brain areas are active, however, it would give further insight into the nature of aesthetic processing and the distinction to complexity judgements, which eye tracking failed to do. The best strategy for future studies should therefore be to focus on comparing various kinds of descriptive rating tasks and compare them to evaluative aesthetic judgement tasks using a variety of neurophysiological methods.

Other factors that may have played a crucial part in the study's outcome are inter-subject variability and the selected stimulus set. Participants were very consistent in the way they viewed the images, but there was variation between the subjects. Each participant had their own strategy; some tended to make more fixations, others less, some dwelled longer on the same spot while making short and quick jumps and others did the opposite. This individual pattern transcended from the encoding phase to the recognition phase irrespective of task type. However, our study was not designed to take these distinct viewing strategies into account. Only a within-subjects design could clarify whether there are differences in fixation patterns in the two conditions for each individual respectively. If each person performs both the aesthetic and the complexity task in the same experiment, we could find out whether an aesthetic processing mode results in substantial changes in viewing strategy within participants.

But not only the personal characteristics but also the selected stimulus set plays a major role in the guidance of participants' eye movements. We chose a rather heterogenous stimulus set with images varying in complexity and style to control for potential confounding influences. However, this brings up an issue that was already discussed by Fuchs et al. (2011). In their study they demonstrated how image salience attracts the viewer's attention and how different stimulus categories influence viewing behavior. They found that abstract artworks compared to representative artworks tend to induce longer fixations as processing efforts to master the artwork's content increase. This led to differences in average fixation counts and fixation durations for each stimulus category respectively. In our study the stimulus set also consisted of various image types, but the majority of artworks was representational, which is why it was not included as a factor in the statistical analysis. However, future studies may take that into account by comparing the aesthetic and the complexity condition using a balanced set of abstract and representational artworks and examine whether image type influences fixation patterns differently in the two conditions.

That the selected artworks in this study were overall neither liked nor disliked is also relevant, especially for studies trying to investigate aesthetic experiences. As described in the beginning, aesthetic experiences are unique states, in which the observer is excited or moved by an artwork (Chatterjee & Vartanian, 2014; Cupchik et al., 2009; Di Dio & Gallese, 2009; Massaro et al., 2012). This is hardly inducible during an eye tracking experiment in the laboratory, using images that may not satisfy everyone's taste and are only shown for a limited amount of time. But the idea was that any perceivable object or scene can be approached with an aesthetic orientation or mode of processing, which theoretically should suffice to profit from memory advantages. It would be interesting, however, to move out of this standardized environment and do more studies in museums where observers can really appreciate the artworks. Another option would be to question participants beforehand about their preferred art styles and then match the stimuli accordingly. Capturing each participant's individual taste is definitely a challenge but it might be worth the effort. Our data suggest that images were better remembered the higher the liking rating was. There was also a trend in the opposite direction with disliked images being remembered quite well, but the results were not significant. This leads to the question whether there would have been a difference in recognition accuracy between the aesthetic and the complexity group had the stimulus set consisted of more stimuli that were liked by the participants. Support for this reasoning comes from neuroimaging studies in which judged-beautiful stimuli induced different brain activation compared to neutral and ugly rated images (Di Dio & Gallese, 2009; Kawabata & Zeki, 2004). Especially activation in the medial orbitofrontal cortex (OFC), which is related to reward, as well as the insula and right amygdala, which are core emotion centers, was higher when images were liked relative to those that were disliked. This shows that liked images may be better suited in evoking aesthetic experiences and may thereby lead to memory advantages, which highlights the importance of the selected stimulus set.

Another dependent variable worth mentioning is the security rating during the recognition test which revealed that participants in the aesthetic condition were equally sure about their answers compared to the ones in the complexity group. This also speaks against the depth of processing hypothesis, because according to that view deeper processing in the aesthetic condition should lead to higher confidence in the recognition test compared to the non-aesthetic condition. However, that there was no difference between the two groups suggests that over all participants believed to remember the images equally well. Both groups rated to be generally quite sure about their performance and the positive moderate correlation between the security rating and recognition accuracy indicates that the answers were mostly correct when the rating was higher. Including a security rating in studies with a recognition test has the advantage that it reveals information about the perceived difficulty of the test itself aside from the actual measure of recognition accuracy. Wang and colleagues (2016) describe it as a metacognitive measure of the participant's ability to introspect on the accuracy of their memory. Future studies may include varying test difficulty as an independent variable and examine whether and how it affects recognition accuracy and the participants' metacognitive performance in an aesthetic versus a non-aesthetic condition.

Finally, we inspected how eye movement parameters differed in the encoding phase from the recognition phase in which participants partly viewed the same images for a second time. Eye tracking studies focusing on the development of eye movements during repeated presentation of the same stimuli have found evidence that the fixation count decreases (Sharot, Davidson, Carson, & Phelps, 2008), fixation duration increases, and saccade amplitude decreases (Kaspar & König, 2011). Contrary to these findings, the pairwise comparisons we conducted of these eye movement parameters in T1 and T2 yielded no significant difference in fixation count, an increase in saccade amplitude and a decrease in fixation duration, although this was only significant for the complexity condition. One explanation for these contradictory results could be that the task type itself induced different fixation patterns. Studies on top-down influences on eye movements already showed that people generally direct their gaze toward informative task-relevant locations (Kardan et al., 2016; Henderson et al., 2009; Rothkopf et al., 2007) and also intuitively it seems reasonable to expect different fixation patterns in a rating task and a recognition test. However, Foulsham and colleagues (2012) had a very similar study design. They also conducted a recognition test after asking participants to memorize images of natural scenes. Then they compared eye movements between correct and incorrect trials of the recognition test and found that trials with correct responses led to fixation patterns that were more similar in T1 and T2 compared to incorrect trials. Scanpath comparison further revealed that correct recognition trials were characterized by partial reconstruction of the original scanpath from the encoding phase, which was also supported by other studies (Foulsham & Underwood, 2008; Foulsham & Kingstone, 2013; Holm & Mäntylä, 2007). According to scanpath theory (Noton & Stark, 1971) reinstating a sequence of eye fixations during the recognition phase activates the previously created representations, facilitating better recognition of the image. Inconsistent study-test fixations imply sampling of new information, which means that the image was not recognized. The essential difference between the Foulsham et al. (2012) study and ours is that during memorization participants may have already tried to fixate relevant regions that facilitate subsequent recognition. During the recognition test itself participants profited from reinstating their previous fixations, thereby refixating the previously memorized features, which facilitates better recognition. In our study on the other hand, they merely rated images without expecting any kind of examination. Participants evaluating aesthetic preference or visual complexity are unlikely to choose a memorization strategy in which they focus on distinctive features that can later be retrieved more easily. Instead, they probably approach the image more holistically to come to an elaborate rating decision. During the recognition test they then rely on their representations and try to identify certain features that look familiar. This may have led to different fixation patterns and thus caused the discrepancy between our results and previous findings. Whether scanpath recapitulation was functional in scene memory and whether there were differences in scanpath similarity between the aesthetic and the complexity condition can be investigated in follow-up studies. In general, scanpath comparison looks like a promising approach because a lot of information about the exact fixation locations can be retrieved. However, computer programs specialized in scanpath comparison are still scarce and relatively underdeveloped. Most studies tried to tackle this issue by simplifying the design; focusing only on a limited number of fixations, for example, or restraining the viewing time to a few seconds only (Foulsham et al., 2012; Holm & Mäntylä, 2007; Sharot et al., 2008). But with increased viewing time comparing scanpaths becomes highly complex. Vector-based software programs like MultiMatch or ScanMatch are commonly used but are limited in their ability to capture both the spatial and temporal properties of scanpaths (Dewhurst et al., 2012). Nevertheless, they are useful tools in measuring and predicting people's eye movements and pave the way for advanced eye tracking studies.

Aside from the already mentioned suggestions for future research in this area we would like to suggest two more research questions worth investigating. The first is whether and how gender influences aesthetic processing. Through magnetoencephalography (MEG), it was already shown that different brain areas were active in men and women when judging images on their beauty (Cela-Conde et al., 2009). Now it would be interesting to examine whether these differences in brain activation have an impact on eye movement patterns and whether the formed representations differ among men and women for an aesthetic and a non-aesthetic condition respectively. The second research proposal is to replicate some of the studies about aesthetic experiences and replace lay people with experts. Art experts generally tend to use a more global image processing strategy compared to lay people (Pihko et al., 2011; Vogt, 1999; Zangemeister et al., 1995). This leads to more abstract representations of artworks as more knowledge can be integrated, resulting in greater appreciation (Augustin & Leder, 2006; Leder et al., 2012). It is therefore probably easier to induce aesthetic experiences in experts compared to lay people, because of their superior knowledge and interest in the field. The only downside about studies with art experts is that the resulting findings are not generalizable to the entire population. However, they may reveal interesting aspects about aesthetic experiences that may not be detectable with lay persons.

## Conclusion

This study aimed at investigating aesthetic experiences and the benefits of aesthetic processing. Using eye tracking as an instrument to capture visual attention we tried to identify the composition and structure of representations that were created during the exposure to artworks. Contrary to the depth of processing hypothesis we did not find differences in eye movement patterns during an aesthetic judgement task compared to a visual complexity rating task. Both conditions seem to induce global processing strategies leading to similar representations and thereby equal recognition performance. That we failed to identify distinct fixation patterns during the aesthetic judgement task can have two reasons: either aesthetic processing is not as special and unique as presumed, or the rating task itself is the key component of memory advantages in studies using aesthetic judgements. Further research using a variety of incidental encoding tasks is necessary to clarify which eye movement patterns exactly may be attributed to an aesthetic experience. By extracting the exact characteristics of aesthetic processing future studies should be able to predict recognition performance and even distinguish various kinds of mental states. That artworks which were rated higher on the aesthetic preference scale were also better remembered does indicate that aesthetic appeal comes with certain benefits in memory enhancement. We therefore believe that even though our study was not able to highlight the advantages of aesthetic processing on memory, aesthetic experiences are something special and should be explored further.

# References

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103*, 62-70. doi:10.1037/h0036799

Augustin, M. D., & Leder, H. (2006). Art expertise: A study of concepts and conceptual spaces. *Psychology Science, 48*, 135-156.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage, 76*, 412-427.

Bernstein, L. J., Beig, S., Siegenthaler, A. L., & Grady, C. L. (2002). The effect of encoding strategy on the neural correlates of memory for faces. *Neuropsychologia, 40*(1), 86-98.

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research, 15*, 719-722.

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

Castelhano, M., & Henderson, J. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12*(6), 1017-1040.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9*(3), 1-15. doi:10.1167/9.3.6

Cela-Conde, C. J., Ayala, F. J., Munar, E., Maestú, F., Nadal, M., Capó, M. A., del Río, D., López-Ibor, J. J., Ortiz, T., Mirasso, C., & Marty, G. (2009). Sex-related similarities and differences in the neural correlates of beauty. *Proceedings of the National Academy of Sciences,* 106(10), 3847-3852.

Cela-Conde, C. J., García-Prieto, J., Ramasco, J. J., Mirasso, C. R., Bajo, R., Munar, E., Flexas, A., del-Pozo, F., & Maestú, F. (2013). Dynamics of brain networks in the aesthetic appreciation. *Proceedings of the National Academy of Sciences, 110*, 10454-10461.

Chatterjee, A., & Vartanian, O. (2014). Neuroaesthetics. *Trends in Cognitive Sciences, 18*(7), 370-375.

Choe, K.W., Kardan, O., Kotabe, H. P., Henderson, J. M., & Berman, M. G. (2017). To search or to like: Mapping fixations to differentiate two forms of incidental scene memory. *Journal of Vision, 17*(12), 1-22, doi:10.1167/17.12.8

Craik, F. I. M. (2002). Levels of processing: Past, present, and future? *Memory, 10*(5-6), 305-318.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Cunningham, W.A., Johnson, M.K., Gatenby, J.C., Gore, J.C., & Banaji, M.R., 2003. Neural components of social evaluation. *Journal of Personality and Social Psychology, 85*, 639–649.

Cupchik, G. C., Vartanian, O., Crawley, A., & Mikulis, D. J. (2009). Viewing artworks: Contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain and Cognition, 70*, 84-91.

Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behav Res, 44*, 1079–1100. doi:10.3758/s13428-012-0212-2

Di Dio, C., & Gallese, V. (2009). Neuroaesthetics: a review. *Current Opinion in Neurobiology, 19*, 682-687. doi:10.1016/j.conb.2009.09.001

Draschkow, D., Wolfe, J. M., & Võ, M. L. -H. (2014). Seek and you shall remember: Scene semantics interact with visual search to build better memories. *Journal of Vision, 14*(8), 1-18, doi:10.1167/14.8.10

Faul, F., Erdfelder, E., Lang, AG., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi: https://doi.org/10.3758/BF03193146

Foulsham, T., Dewhurst, R., Nyström, M., Jarodzka, H., Johansson, R., Underwood, G., & Holmqvist, K. (2012). Comparing scanpaths during scene encoding and recognition: A multi-dimensional approach. *Journal of Eye Movement Research, 5*(4), 1-14.

Foulsham, T., & Kingstone, A. (2012). Fixation-dependent memory for natural scenes: An experimental test of scanpath theory. *Journal of Experimental Psychology: General. 142,* 41-56. doi:10.1037/a0028227

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8*(2), 1-17. doi:10.1167/8.2.6

Frith, C. D., & Nias, D. K. B. (1974). What determines aesthetic preferences? *Journal of General Psychology, 91*, 163-173.

Fuchs, I., Ansorge, U., Redies, C., & Leder, H. (2011). Salience in paintings: bottom-up influences on eye fixations. *Cognitive Computation, 3*(1), 25-36. doi:10.1007/s12559-010-9062-3

Gartus, A., & Leder, H. (2013). The small step toward asymmetry: Aesthetic judgment of broken symmetries. *i-Perception, 4*(5), 352-355. doi:dx.doi.org/10.1068/i0588sas

Glazek, K. (2012). Visual and motor processing in visual artists: Implications for cognitive and neural mechanisms. *Psychology of Aesthetics, Creativity, and the Arts, 6*(2), 155-167. doi:10.1037/a0025184

Grady, C. L., Bernstein, L. J., Beig, S., & Siegenthaler, A. L. (2002). The effects of encoding task on age related differences in the functional neuroanatomy of face memory. *Psychology and Aging, 17*(1), 7-23.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., & Cohen, J.D., 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389-400.

Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences, 21*(1), 15-23.

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour, 1*, 743-747. doi:https://doi.org/10.1038/s41562-017-0208-0.

Henderson, J. M., Malcom, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review, 16*, 850-856. doi:10.3758/PBR.16.5.850

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15*, 566-573. doi:10.3758/PBR.15.3.566

Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. Journal of Experimental Psychology: *Human Perception and Performance, 25*(1), 210-228.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 113-136.

Holm, L., & Mäntylä, T. (2007). Memory for scenes: Refixations reflect retrieval. *Memory & Cognition, 35*(7), 1664-1674.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

Ishai, A. (2011). Art compositions elicit distributed activation in the human brain. In A. P. Shimamura and S. E. Palmer (Eds.), *Aesthetic Science: Connecting Minds, Brains, and Experience* (pp. 337-355). Oxford University Press: New York.

Jacobsen, T., Schubotz, R. I., Höfel, L., & Cramon, D. Y. V. (2006). Brain correlates of aesthetic judgment of beauty. *Neuroimage, 29*(1), 276-285. doi:10.1016/j.neroimage.2005.07.010

Josephs, E. L., Draschkow, D., Wolfe, J. M., & Võ, M. L.-H. (2016). Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica, 169*, 100-108.

Kardan, O., Berman, M. G., Yourganov, G., Schmidt, J., & Henderson, J. M. (2015). Classifying mental states from eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1502-1514.

Kardan, O., Henderson, J. M., Yourganov, G., & Berman, M. G. (2016). Observers' cognitive states modulate how visual inputs relate to gaze control. *Journal of Experimental Psychology: Human Perception and Performance, 42*(9), 1429-1442.

Kaspar, K., & König, P. (2011). Overt attention and context factors: the impact of repeated presentations, image type, and individual motivation. *PloS One, 6*(7), e21719. doi:10.1371/journal.pone.0021719

Kawabata, H., & Zeki, S. (2004). Neural correlates of beauty. *Journal of Neurophysiology, 91*, 1699-1705.

Koide, N., Kubo, T., Nishida, S., Shibata, T., & Ikeda, K. (2015). Art expertise reduces influence of visual salience on fixation in viewing abstract paintings. *PloS ONE, 10*(2), 1-14. doi:10.1371/journal.pone.0117696

Kosslyn, S.M., Alpert, N.M., Thompson, W.L., Maljkovic, V., Weise, S.B., Chabris, C.F., Hamilton, S.E., Rauch, S.L., Buonanno, F.S. (1993). Visual mental imagery activates topo-graphically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience, 5*, 263-287.

Le Bihan, D., Turner, R., Zeffiro, T., Cuendo, C., Jezzard, P., Bonnerot, V. (1993). Activation of human primary visual cortex during visual recall: A magnetic resonance imaging study. *Proceedings National Academy of Sciences USA, 90*, 11802-11805.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology, 95*, 489-508. doi:10.1348/0007126042369811

Leder, H., Gerger, G., Dressler, S. G., & Schabmann, A. (2012). How art is appreciated. *Psychology of Aesthetics, Creativity, and the Arts, 6*(1), 2-10. doi:10.1037/a0026396

Locher, P., Krupinski, E., Mello-Thoms, C., & Nodine, C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision, 21*, 55-77. doi:10.1163/156856807782753868

Locher, P., & Nodine, C. (1987). Symmetry catches the eye. In J. K. O'Regan & A. Levy-Schoen (Eds), *Eye movements: From physiology to cognition* (pp. 353-361). Holland: Elsevier.

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology, 3*(4), 525-551.

Marzi, T., & Viggiano, M. P. (2010). When memory meets beauty: Insights from event-related potentials. *Biological Psychology, 84*(2), 192-205. doi:10.1016/j.biopsycho.2010.01.013

Massaro, D., Savazzi, F., Di Dio, C., Freedberg, D., Gallese, V., Gilli, G., & Marchetti, A. (2012). When Art Moves the Eyes: A Behavioral and Eye-Tracking Study. *PLoS ONE, 7,* e37285. doi:10.1371/journal.pone.0037285

Mellet, E., Tzourio, N., Crivello, F,, Joliot, M., Denis, M., Mazoyer, B. (1996). Functional anatomy of spatial mental imagery generated from verbal instructions. *Journal of Neuroscience, 16*, 6504-6512.

Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision, 11*(8), 17-17. doi:10.1167/11.8.17

Moll, J., de Oliveira-Souza, R., Bramati, I.E., & Grafman, J., 2002. Functional networks in emotional moral and nonmoral social judgments. *Neuroimage, 16*, 696-703.

Murty, V. P., DuBrow, S., & Davachi, L. (2015). The simple act of choosing influences declarative memory. *The Journal of Neuroscience, 35*(16), 6255-6264. doi:10.1523/JNEUROSCI.4181-14.2015

Nadal, M., Marty, G., & Munar, E. (2006). The search for objective measures of aesthetic judgment: The case of memory traces. *Empirical Studies of the Arts, 24*(1), 95-106.

Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts, 28*(2), 173-191.

Nodine, C. F., Locher, P. J., & Krupinski, E. A. (1993). The role of formal art training on perception and aesthetic judgment of art compositions. *Leonardo, 26*, 219-227. doi:10.2307/1575815

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research, 11*, 929–942.

O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change blindness as a result of 'mudsplashes'. *Nature, 398*, 34.

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*(4), 1-19. doi:10.16910/jemr.2.2.4

Peacock, C. E., Hayes, T. R., & Henderson, J. M. (2019). Meaning guides attention during scene viewing, even when it is irrelevant. *Attention, Perception, & Psychophysics, 81*(1), 20-34. doi:https://doi.org/10.3758/s13414-018-1607-7

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

Quian Quiroga, R., & Pedreira, C. (2011). How do we see art: an eye-tracker study. *Frontiers in Human Neuroscience, 5,* 98. doi:10.3389/fnhum.2011.00098

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7*(16), 1-20. doi:10.1167/1167.1114.1116.

Ramachandran, V. S., & Hirstein, W. (1999). The science of art. *Journal of Consciousness Studies, 6*(6–7), 15-51.

Sharot, T., Davidson, M. L., Carson, M. M., & Phelps, E. A. (2008). Eye movements predict recollective experience. *PLoS One, 3*(8), e2884. doi:10.1371/journal.pone.0002884

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261-267.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207-222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single trial learning of 2500 visual stimuli. *Psychonomic Science, 19*, 73-74.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition, 12*, 473-494. doi:10.1080/13506280444000409

Velichkovsky, B. M. (2002). Heterarchy of cognition: The depths and the highs of a framework for memory research. *Memory, 10*, 405-419. doi:10.1080/09658210244000234

Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the XXVII conference of the Cognitive Science Society* (pp. 2283-2288). Mahaw, NJ.: Lawrence Erlbaum Associates.

Vogt, S. (1999). Looking at paintings: patterns of eye movements in artistically naïve and sophisticated subjects. Leonardo, 32(4), 325-328.

Vogt, S.E., & Magnussen, S. (2007). Expertise in pictorial perception: Eye movement patterns and visual memory in artists and laymen. *Perception, 36*, 91-100. doi:10.1068/p5262

Wallraven, C., Cunningham, D., Rigau, J., Feixas, M., & Sbert, M. (2009). Aesthetic appraisal of art: from eye movements to computers. In O. Deussen, D. Fellner, & N. Dodgson (Eds.), *Computational Aesthetics 2009: Eurographics Workshop on Computational Aesthetics in Graphics, Visualization and Imaging* (pp. 137-144). Aire-La-Ville, Switzerland: Eurographics.

Wang, T., Cant, J. S., & Cupchik, G. (2016). The impact of depth of aesthetic processing and visual-feature transformations on recognition memory for artworks and constructed design patterns. *Empirical Studies of the Arts, 34*, 193-220.

Wolfe, J. M., Horowitz, T. S., & Michod, K. O. (2007). Is visual attention required for robust picture memory? *Vision Research, 47*(7), 955-964.

Yarbus, A. L. (1967) *Eye Movements and Vision*. Plenum. New York. (Originally published in Russian 1962).

Zangemeister, W. H., Sherman, K., & Stark, L. (1995). Evidence for a global scan path strategy in viewing abstract compared with realistic images. *Neuropsychologia, 33*, 1009-1025. doi:10.1016/0028-3932(95)00014-T

Zeki, S. (1980). The representation of colours in the cerebral cortex. *Nature, 284*, 412-418.

Zysset, S., Huber, O., Ferstl, E., & von Cramon, D.Y., 2002. The anterior frontomedian cortex and evaluative judgment: an fMRI study. *Neuroimage, 15*, 983– 991.