Virtual Reality (2026) 30:123
https://doi.org/10.1007/s10055-026-01372-9

ORIGINAL ARTICLE

Aesthetic experience in the physical and immersive virtual space

Giulia Tomasetig[1,2] · Angelo Maravita[1,2] · Cosmo Laera[3] · Stefano Pizzi[3] · Eraldo Paulesu[1,2,4] · Lucia Maria Sacheli[1,2,4]

Received: 24 November 2025 / Accepted: 3 April 2026

Abstract
Art contemplation typically occurs in museums or art galleries. However, aesthetic experience has been predominantly investigated in laboratory settings, limiting the generalizability of the results. Virtual reality (VR) can create immersive and realistic simulations, potentially bridging this gap. However, the extent to which aesthetic engagement in virtual environments parallels that of physical exhibitions remains to be tested. To address this question, twenty-four art naïve participants visited both a physical photographic exhibition and its virtual counterpart designed for Oculus Quest 2. We assessed both explicit aesthetic experiences (liking, curiosity, and emotional involvement) and implicit behaviors towards art (viewing times) in both settings. By employing Bayesian linear mixed-effects analyses, we found evidence of no difference between the two environments in terms of liking and emotional involvement. However, the participants who first visited the exhibition in the physical context reported lower curiosity towards the photographs during the second session (i.e., in VR). We also found strong evidence for the lack of a modulatory role of context over phenomena generally found in real-world exhibition, i.e., the *museum fatigue* and the relationship between viewing times and aesthetic experience. Moreover, one year later, participants similarly rated the vividness of their memory and their willingness to revisit the virtual and physical exhibition. Altogether, these findings suggest that aesthetic experiences in a controlled physical context can be comparable to those experienced in its immersive digital twin, with meaningful scientific and cultural implications.

Keywords Aesthetics · Virtual reality · Digital twin

1 Introduction

Art is deeply rooted in our culture and communities. Throughout time and civilizations, human beings have consistently turned to art to give form to emotions and thoughts that often elude verbal expression. In its myriad forms, art is a powerful vehicle for transmitting ideas, values, and history, functioning as a social and cultural anchor (Zaidel 2018). Yet, beyond its societal role, the encounter with art remains essentially a subjective, hedonic experience, capable of resonating with the most intimate aspects of the self (Vessel et al. 2012, 2013). We know that art appreciation is a complex phenomenon arising from the profound interplay between observer, artwork, and environment (Leder et al. 2004; Leder & Nadal 2014; Leder & Pelowski 2021; Pearce et al. 2016; Pelowski et al. 2017a, b), in which emotion–valuation, sensory–motor, and meaning–knowledge processes are involved (Chatterjee & Vartanian 2014, 2016). Although art contemplation usually occurs in museums or art galleries, most empirical studies on aesthetic experience have been carried out in laboratory settings. These allow examining the effects of specific manipulated factors while controlling confounding variables, thus providing obvious methodological advantages. Yet, several studies show that the laboratory setting is not a good proxy of the real one, failing to capture the complexity and depth of the aesthetic experience (for a review, see Pelowski et al. 2017a, b), producing findings that are not fully generalizable to real-world circumstances (Grüner et al. 2019; Pelowski et al. 2017a, b; Specker et al. 2017; Specker et al. 2023a, b; Szubielska et al.

✉ Giulia Tomasetig
g.tomasetig@campus.unimib.it

✉ Lucia Maria Sacheli
lucia.sacheli@unimib.it

1 Psychology Department and NeuroMi—Milan Center for Neuroscience, University of Milano-Bicocca, Milan, Italy

2 Mind and Behavior Technological Center, University of Milano-Bicocca, Milan, Italy

3 Academy of Fine Arts of Brera, Milan, Italy

4 IRCCS Orthopedic Institute Galeazzi, Milan, Italy

Published online: 24 April 2026

2021; Szubielska & Imbir 2021). Indeed, on average, people like artworks more, find them more interesting, experience more positive affective responses, and observe paintings for longer in the museums/art galleries than in the laboratory (Brieber et al. 2014; Brieber et al. 2015a, b; Grüner et al. 2019; Locher et al. 1999, 2001; Locher & Dolese 2004; Specker et al. 2017; Szubielska et al. 2021; Szubielska & Imbir 2021, but see Darda et al. 2025). Artworks are also remembered better and longer when viewed in a museum (Brieber et al. 2015a, b; Specker et al. 2017). Current experimental approaches, dominated by laboratory studies, may thus have led to an understanding of the aesthetic experience at least in part detached from real-life art-observer interactions, calling for more ecologically valid studies (Pelowski et al. 2017a, b). Several aspects differentiate the real context from the laboratory one. In the latter, artworks are typically displayed on a computer screen, with dimensions that obviously do not match the original sizes. Texture is another factor that distinguishes an original artwork from its reproduction, especially for impasto paintings and sculptures, which possess three-dimensional qualities. These are flattened when reproduced in two dimensions, resulting in the loss of both haptic and visual information. Moreover, in the laboratory, artworks are typically displayed one after another on the screen, making viewers passive receivers with little or no control over what they observe (Pelowski et al. 2017a, b). In contrast, in the museum context, visitors intentionally approach artworks and freely move within the exhibition. Furthermore, laboratory settings frequently present sequences of artworks without coherent thematic logic – often relying on the randomization of stimuli order. To maximize generalizability of the results, these studies present participants with unrelated artworks, differing perhaps in style, historical period, or artist. In museums, by contrast, beholders visit thematic exhibitions, and not random images: curators carefully select where and how to position artworks, creating structured exhibition pathways that contribute to conveying the overarching meaning of the exhibition. According to Pelowski and colleagues (Pelowski et al. 2017a, b), all the factors mentioned above can influence the interaction with art and exert an impact on aesthetic experience that can be interpreted in the light of the influential theoretical model proposed by Leder and colleagues (2004, 2014). Indeed, the model proposes that the aesthetic experience begins even before perception, with the context that contributes to shaping expectations. Then, the aesthetic experience is described as a sequence of processing stages that lead to two outputs: cognitive-based aesthetic judgments, and affective-based aesthetic emotions. Aesthetic judgments refer to explicit evaluative appraisals, which, operationally, are typically captured by collecting ratings of liking, beauty, interestingness, or understanding. Aesthetic emotions, in turn, denote the affective experience elicited during the encounter with the artwork and, operationally, can be assessed via self-reported emotional responses (e.g., emotional involvement, arousal, valence) and complemented by physiological indices (Leder et al. 2004; Leder & Nadal 2014). Thus, the context in which art is experienced (e.g., a museum vs. a laboratory setting) can influence both cognitive- and affective-based explicit judgments by shaping expectations as well as the perceptual experience. In this regard, curiosity would act as a "bridge" between cognitive-based aesthetic judgments and affective-based aesthetic emotions. Within the predictive processing framework, which is increasingly adopted in the last few years as a framework for understanding aesthetic phenomena (Frascaroli et al. 2024; Van de Cruys et al. 2024), curiosity is considered as an "epistemic emotion" or "metacognitive feeling" (Vogl et al. 2021) that signals an "information need" and acts as a drive to information-seeking to grasp the meaning of an artwork (Frascaroli et al. 2024; Van de Cruys et al. 2024). Curiosity is often theorized as arising from the perception of a knowledge gap, motivating information-seeking behavior aimed at reducing that gap (Kenett et al. 2023; Loewenstein 1994). Importantly, it arises when a stimulus is partly uncertain yet offers an epistemic promise: the implicit expectation that individuals, by getting additional information, can make progress in resolving that uncertainty and achieve deeper understanding (Van de Cruys et al. 2024).

While these outputs reflect the explicit counterpart of aesthetic experience, theoretical interest is also given to implicit behaviour towards art (such as viewing times and viewing distance). Theoretical debate has long discussed whether and how artistic experience and viewing time influence one another (Brieber et al. 2020). Leder and Nadal (2014), in their theoretical model (Leder & Nadal 2014), argued that a key component of (deep) aesthetic experience may be "its long extension in time", thus highlighting that viewing times may represent a critical element associated with the aesthetic experience. On the one hand, this is one of the reasons why laboratory studies are criticized for showing works of art for too limited time intervals (a few seconds): restrictions on viewing time can constrain how deeply viewers engage in visual and semantic processing, thereby shaping the beholder's overall experience (for a study on this topic see, for instance, Brieber et al. 2020). The considerations reported above make it necessary to analyse implicit behaviour towards art (such as viewing times) when aiming to investigate the impact of context on the aesthetic experience.

Given these premises, it becomes necessary to identify new approaches for investigating the processes and phenomena that occur during the encounter with art in a more ecologically valid manner (Specker 2025). One solution has been conducting field studies directly within the physical spaces where exhibitions are hosted. An increasing number of studies are adopting this approach, employing mobile measures such as eye-tracking and portable brain imaging technologies like electroencephalography (EEG) (for example, see Castellotti et al. 2023; Pelowski et al. 2018; Shemyakina et al. 2024). Nevertheless, such studies also have several practical challenges and methodological disadvantages regarding experimental control. For instance, the visibility of artworks or responses may be influenced by the presence of other visitors (see, for instance, Gulhan et al. 2021; Szubielska & Imbir 2021), or restoration work may temporarily restrict access to certain areas of the museum. Moreover, it is often difficult, if not impossible, to manipulate certain variables (e.g., wall color) or to modify the building's layout (e.g., to assess the impact of architectural elements on the aesthetic experience). Similarly, the arrangement of artworks cannot be changed if it has been determined by the curatorial design of the exhibition.

A compelling alternative could be virtual reality (VR), immersing participants in a setting closely resembling the real-world environment. VR is a technology that has increasingly been employed in recent years to simulate immersive and realistic experiences. It offers the significant advantage of both providing a high degree of realism and enabling high control of experimental conditions, potentially combining the strengths of laboratory settings with those of real-world environments, while overcoming the limitations typically associated with each of them (Rodriguez-Boerwinkle et al. 2022). An immersive VR art gallery could place individuals within a meaningful context, providing realistic experiences and faithfully translating the key features of a naturalistic setting into a simulated one (Kuliga et al. 2015), allowing the investigation of the aesthetic experience from a more ecological yet controlled perspective. However, to the best of our knowledge, only a couple of studies to date have addressed whether immersive VR constitutes a valid proxy in this regard and to what extent people's aesthetic experiences are comparable between real and immersive virtual contexts. One of these studies compared museum visitors' gaze patterns between physical and VR settings (Gulhan et al. 2021). The comparison revealed no significant effect of the environment, indicating a similarity in visual exploration patterns between the two contexts. Nonetheless, participants observed only a single installation, and the analyses were limited to eye movement patterns, without assessing or comparing the broader explicit cognitive and affective components of the aesthetic experience between the two settings. In another study, Marín-Morales and colleagues (2019) compared affective (i.e., valence and arousal) and physiological responses elicited by visiting an exhibition in a real-world setting versus experiencing the same exhibition in an immersive virtual environment. The authors found no significant difference in self-reported arousal and valence between the two contexts for the vast majority of stimuli. However, subjective evaluations only focused on the emotional impact of each stimulus, neglecting the more cognitive dimension of the aesthetic experience; moreover, the ratings were collected only at the end of the entire visit, potentially introducing biases such as recency effects, primacy effects, and memory distortions. To date, no study has systematically investigated whether both subjective explicit (in both its cognitive and affective components) aesthetic experience and implicit behavior towards art (e.g., viewing time) gathered in a physical setting can be considered comparable to the one collected in its digital twin (i.e., highly detailed virtual replica), by collecting evaluations in real time, thus capturing the unfolding of the aesthetic experience. Moreover, the few studies conducted thus far have employed only a limited number of artistic stimuli, without examining whether phenomena typically observed in museum contexts occur similarly in virtual environments. Such phenomena include, for example, (i) the relationship between viewing times and aesthetic judgments – namely, the tendency for individuals to spend more time observing artworks they find more appealing (Brieber et al. 2014; Goller et al. 2019), and (ii) *museum fatigue*, an effect documented both in museums (Robinson 1928) and lab-based settings (Mikuni et al. 2022) referring to the reduction in attention and interest – as indexed by decreasing viewing times – as visitors/participants are exposed to an increasing number of artworks (Bitgood 2009). Previous research has shown that context (museum vs laboratory) modulates both the relation between art experience and viewing time (Brieber et al. 2014) and the *museum fatigue* phenomenon (Robinson 1928), but evidence regarding the immersive virtual environment is lacking.

1.1 Aims

In the present manuscript we aimed to investigate the extent to which users' aesthetic experiences are analogous between a physical context and its immersive digital twin. In two different sessions, participants were invited to (physically and virtually) visit the photograph exhibition entitled *Domani*, which brings together photographs by young artists created in the post-lockdown climate, as a collective reflection on expectations, fears, and desires for the post-pandemic era. The exhibition was held within the University of Milano-Bicocca (Milano, Italy) in collaboration with the Brera Academy of Fine Arts (Milan, Italy). Our methodological choices were based on Leder and colleagues' influential theoretical model and on the predictive processing framework. For these reasons, we collected explicit self-report judgments on the cognitive-based (i.e., liking) and affective-based (i.e., emotional involvement) components of the aesthetic experience in both contexts, as well as curiosity. Moreover, given the critical role played by aesthetic experience "extension in time" (Leder & Nadal 2014), we also measured viewing time as an index of implicit behavior towards art in the real and virtual environment. Overall, this experimental design enabled us to achieve our primary aim of comparing the real and virtual contexts (Aim1) on cognitive-based and affective-based explicit judgments, curiosity, as well as implicit measures. Then, we exploratorily investigated whether phenomena usually reported during physical exhibitions also occur within virtual immersive environments (Aim2). More specifically, we focused on: i) the relationship between viewing times and aesthetic judgments, and ii) the *museum fatigue*. For the former, we investigated whether viewing times predicted art experiences (i.e., liking, curiosity, emotional involvement) and whether context (i.e., physical/VR) moderates such an effect. For the latter, we assessed whether viewing times decrease as participants see more photographs, and whether context moderates this effect. Finally, we assessed whether, one year after data collection, participants rated the vividness of their memory, overall exhibition appreciation, and willingness to revisit the exhibition similarly between contexts (Aim3).

Given that we wanted to evaluate the validity of a null hypothesis (i.e., the absence of an effect of context), we adopted a Bayesian perspective, which allowed us to quantify how much the data provide evidence in favor of the null hypothesis compared to the alternative one, overcoming the limits of frequentist statistics in discussing null effects (Wagenmakers et al. 2018).

2 Material and methods

2.1 Participants

Twenty-six art naïve subjects participated in the within-subjects experiment. However, one of them did not join the second session, while another was excluded due to inappropriate behaviors during the real tour (i.e., he got too close to the works, touching them and leaving the "observation station"), thus leaving the final sample of twenty-four subjects (sixteen female, mean age 23.44 ± 2.55). All the participants had normal or corrected to normal vision, and none reported a history of neurological or psychiatric disorder. The experimental protocol was approved by the Ethics Committee of the University of Milano-Bicocca, and the study was conducted in accordance with the Declaration of Helsinki. Written informed consent for study participation was obtained from all participants. One participant also provided explicit informed consent for the publication of identifiable images in an open-access online format. Accordingly, two photographs depicting him with the experimental setup were taken during this participant's sessions.

2.2 Power analysis

To determine the adequacy of the sample size to address Aim1 (i.e., comparing explicit aesthetic judgments between the physical and virtual contexts), we conducted a simulation-based Bayesian power analysis. We simulated 100 datasets, each representing a hypothetical experiment with 24 participants. Since in our experiment each participant observed 37 photographs/set of photographs—each made by a different artist—twice (once in each context: physical and virtual), this results in a total of $24 \times 37 \times 2 = 1776$ observations for each dataset. As recommended by Vasishth and colleagues (2023) and Wang & Gelfand (2002), we generated prior predictive data using informative *sampling* priors (or *design* priors; Schönbrodt & Wagenmakers 2018). *Sampling* priors should be based on existing prior information and beliefs to tune the design and simulate the datasets; thus, we defined them relying on parameter estimates reported in a recent study by Specker and colleagues (2023a), who applied linear mixed-effects models to investigate the role of artwork genuineness on different components of art experience (i.e., arousal, valence, liking, interest, and understanding). To obtain some initial ballpark estimates for the different variance components and fixed effects, we extracted standard deviations associated with random effects and fixed intercepts across four of the five aesthetic outcome variables reported: arousal, valence, liking and interest (see Specker et al.' Table S1 , Specker et al. 2023a). We did not consider the parameters related to understanding, as our study did not assess this dimension. We averaged the standard deviations for viewer-level random effects (mean SD = 0.3825) and artist-level random effects (mean SD = 0.675) and rescaled them to a 0–100 metric by multiplying by 14.29, based on the difference between the original response scale used by Specker et al. (0–7) and the one used in the current study (0–100). This yielded SDs of approximately 5.47 for viewer-level variability and 9.65 for artist-level variability. The mean of the intercepts across these models was 4.165, which—rescaled to the 0–100 range—led to a grand mean of 59.52, used as the intercept in the data-generating process. A standard deviation of 20 was added to reflect residual (intra-individual) variability. So, in each simulated dataset, the dependent variable (art appreciation) was generated under the null hypothesis of no effect of context by combining i) a global mean of 59.52; ii) a random variation across participants (sampled from a normal distribution with SD = 5.47); iii) a random variation across photographs made by different artists (sampled from a normal distribution with SD = 9.65); and iv) a residual error term. All values were truncated to fall within the possible response range of 0–100. For each dataset, we estimated two Bayesian linear mixed-effects models using the brms package (Bürkner 2017a, 2017b) in R: i) the alternative model, which included the fixed effect of context and by-subjects and by-photographs random intercepts and slopes; ii) the null model, which excluded the fixed effect of context but preserved the same random structure. As suggested by Wang & Gelfand (2002) and Schönbrodt & Wagenmakers (2018), we fitted the models and computed the Bayes Factor (BF_{01}) using less informative priors (i.e., *fitting* or *analysis* priors) to allow the data driving the inference and reduce the risk of biasing the results in favor of a particular hypothesis. The same *fitting* priors were also applied in the data analyses. The prior distribution on the fixed effect of context was set to *normal(0, 10)*. For the intercept, we used *normal(50, 20)*, and for random effects (subject and photograph), we used *normal(0, 10)*. Both models were fitted with 4 chains, 4000 iterations each (the warmup was of 2000 iterations), and *adapt_delta = 0.95*. The full script is available at the following link: https://osf.io/8fpvk/. We found that, if no effect of context exists (i.e., the null hypothesis is true, as simulated in our datasets), a sample size of $N = 24$ allowed us to obtain moderate evidence (i.e., $BF_{01} > 3$) in favor of the null hypothesis in 95% of cases.

2.3 Stimuli

The stimuli were taken from the photograph exhibition "*DOMANI. Aspettative, previsioni e desideri per l'epoca post-pandemica*", exposed within the University of Milano-Bicocca spaces. The underlying theme of the exhibition was the tension between isolation and the return to social life, between intimacy and collective space, depicted with a lost yet hopeful gaze at the future. The exhibition was curated by Professor Cosmo Laera of the Brera Academy of Fine Arts, and it consisted of 38 artistic photographs (or sets of related photographs), each made by a different young artist affiliated with the Brera Academy of Fine Arts (for the complete list of stimuli, see https://osf.io/8fpvk/). Each artwork comprised 1 to 9 photographs of various dimensions (range: from 16×20.5 to 74.5×100 cm^2), depicting landscapes, objects, and human figures expressing the sense of bewilderment but also of rebirth that followed the lockdown due to the COVID-19 outbreak and resolution. Each artwork was complemented by a label with the author's name, the title of the artwork, and a QR-code which links to the *BreraBicocca* website (http://www.brerabicocca.it/) where more information is available. However, participants were asked not to scan the QR-code during the experiment. Employing photographs as stimuli—and not paintings or sculpture, for example—allowed us to explore the effect of context, minimizing the possible confound related to the "genuineness", that is, being in front of the original work of art vs its reproduction (Specker et al. 2023a, b) and the effect of texture. Indeed, the same high-resolution digital images that were printed on physical supports for the real exhibition were used as textures of the three-dimensional (3D) mounts within the virtual environment. The 3D VR reconstruction of the hallway where the photograph exhibition was displayed was developed using the Unity game engine (www.unity3d.com) and was designed to replicate physical space faithfully. We indeed measured both the corridor and the photographs' dimensions to ensure the virtual environment, including the artworks, was recreated with the same size and proportions as the physical one. Illustrative photographs of the physical environment and screenshots of the virtual one are shown in Fig. 1. The web version of the 3D virtual environment is available at the following link: https://play.unity.com/en/games/fef5bee8-8258-49fe-ba27-ee0ad4db7279/unityartvrweb. The virtual scenario was customized for Oculus Quest 2 (https://www.meta.com/), and the entire system was connected to the research computer through Meta Quest Link. In order to achieve pseudo-naturalistic navigation and mitigate motion sickness, visitors could move within the virtual environment thanks to the teleport navigation (Clifton & Palmisano 2020). We configured it by using the Unity XR Interaction Toolkit package.

2.4 Apparatus

Within the physical context, the entire tour was recorded using a GoPro camera, and the data were stored with a sample rate of 60 Hz. For each photograph or group of photographs, through a tablet, participants provided their aesthetic evaluations (i.e., liking, curiosity, and emotional involvement) on a horizontal slider ranging from 0 to 100 (see Fig. 2a). Within the virtual context, the video of the visit was recorded with the Unity Recorder package with a sample rate of 60 Hz. Through user interfaces, using the controllers, participants gave their evaluations at the three questions on the same horizontal slider ranging from 0 to 100 (see Fig. 2b). In both contexts, the three questions were presented in sequential order, and participants had to confirm each rating by pressing the button "send" to pass to the next question.

2.5 User experience (UX) measures

At the end of the virtual session, participants completed two questionnaires on their user and virtual experience. The first

Fig. 1 The figure shows a comparison between the physical context where the exhibition was set up (left) and its virtual reproduction (right). Regarding the physical setting, although the labels indicating the title and author of the artwork, as well as "observation stations", were not yet in place when the photos were taken, they were present during data collection. The same applies to the "observation stations" in the virtual environment

one was the System Usability Scale (SUS) (Brooke 1996), a widely used ten-statement scale on the perceived usability of the computer system they employed for the virtual tour. For each statement, participants indicated their agreement on a five-point Likert scale ranging from 1 ("Strongly Disagree") to 5 ("Strongly Agree"). The computed total score ranges from 0 to 100 with 2.5-point increments. The second instrument was the Independent Television Commission Sense of Presence Inventory (ITC-SOPI) (Lessiter et al. 2001), a forty-four-item scale addressing the individual's thoughts and feelings during and after the VR experience on a five-points Likert scale ranging from 1 ("Strongly Disagree") to 5 ("Strongly Agree"). The questionnaire results in four-factor scores, each ranging from 1 to 5: Sense of Physical Space (19 items), which refers to the feeling of being physically positioned in the simulated environment; Engagement (13 items), which provides a measure of a user's involvement, interest and enjoyment with the content of the experience; Ecological Validity (5 items), which relates to the credibility and realism of the content and the naturalness of the simulated environment; and Negative Effects (6 items), which describes the adverse psychological reactions of exposure to the virtual experience (i.e., symptoms of cybersickness such as dizziness, eyestrain and nausea). Given that our environment did not contain characters, as suggested in the scoring instructions, we excluded item B23 in computing the Sense of Physical Space factor score.

2.6 Procedure

Each participant visited the exhibition both in the physical and virtual context (two different sessions). To avoid carry-over effects, the order of the conditions (physical and VR) was counterbalanced, and participants were pseudo-randomly assigned to either a physical-first or a VR-first context. The second session occurred 4–8 days after the first

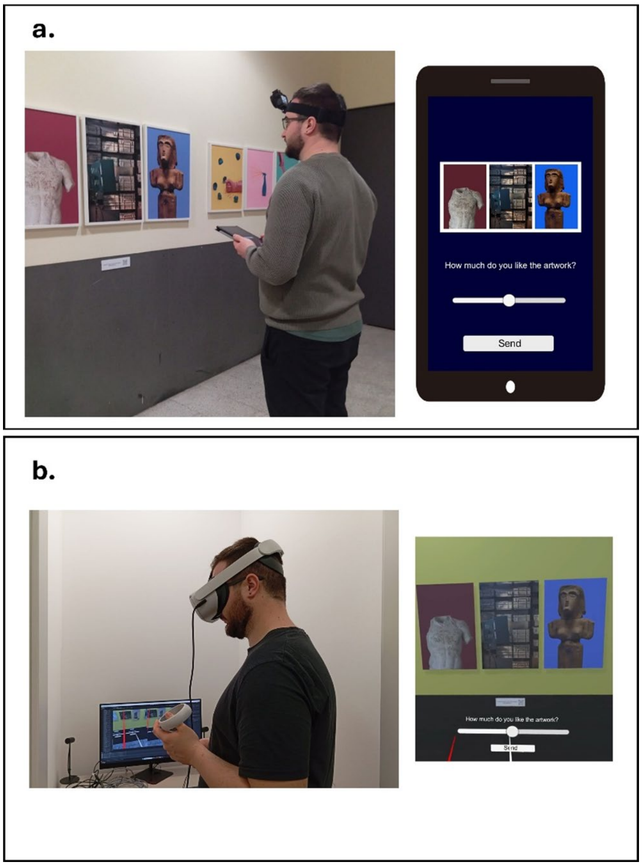

Fig. 2 The figure shows the experimental setup and user interface for providing ratings in the physical **a**. and virtual **b**. contexts. Explicit informed consent was obtained from the participant for taking and publishing these identifying images

one, and mainly within the same time slot. In both conditions, participants could explore the photographic exhibition following a pre-established path. This choice, often adopted in previous studies comparing museum vs laboratory aesthetic experiences (Brieber et al. 2014; Brieber et al. 2015a, b; Szubielska et al. 2021; Szubielska & Imbir 2021) ensured that all participants viewed the artworks in the same fixed order in both environments. In the physical context, the pursuable path was outlined by placing adhesive tape on the floor: in correspondence with the center of each photograph/set of photographs, we delineated an “observation station” with tape, which was placed 1.30 m away from the wall where the artworks were hung. We chose that distance because it allowed participants to observe the details of both large and small-sized artworks comfortably. Participants had to visit the exhibition lingering within the “observation stations”, while they could observe the artworks without time restrictions. We recognize that these methodological choices may limit the overall ecological validity of the experience, as participants were not free to move within the exhibition; however, we would also like to highlight that guided navigation, following a pre-specified pathway, is something that may occur (with looser constraints) in the museum context as well, where curators carefully select where and how to position artworks, creating structured routes that contribute to conveying the exhibition's overarching meaning. In any event, we decided to keep the viewing distance and the order of artworks presentation fixed for all participants in both contexts to prevent any impact of these factors on our results. In light of empirical evidence for the relationship between viewing distance and aesthetic experience (Carbon 2017, 2020; Pelowski et al. 2017a, b), we wanted to ensure that this element was controlled by design. The time spent viewing each photograph/set of photographs was calculated post-hoc from the records obtained through a GoPro camera, positioned on the participants' heads with a suitable harness (see Fig. 2a for an illustration of the experimental set-up). The use of the GoPro also allowed us to control that participants actually observed artworks before providing their aesthetic evaluations. They were indeed instructed to observe the works of art for as long as they wished and, when they felt satisfied, to provide on a slider their ratings on liking, curiosity, and emotional involvement (i.e., *How much do you like the artwork? How curious are you about the artwork? How much does the artwork emotionally involve you?)* by using a tablet. The judgments were given for each artwork (that could be composed by a single photograph or by a set of related photographs made by the same artist) before moving to the next observation station. Supplementary Video S1 provides a clearer example of how the visit and the collection of aesthetic evaluations took place within the physical context.

On the day of the virtual session, for each participant, before starting with the virtual tour of the exhibition, the vertical distance from the floor to the participant's eyes was measured by using a measuring tape. This value was subsequently input into Unity to configure the height of the main camera within the XR Rig relative to the ground. This adjustment ensured that the participants’ virtual height and point of view within the teleportation anchors accurately corresponded to their own in the physical context. In this way, the visual angle with which the participants observed the photographs was the same between the two contexts. Within the virtual environment, participants observed the artworks in the same order as in the physical context and could move through teleportation. Teleportation anchors were placed at the equivalent of 1.3 m from the wall and in correspondence with the center of each photographic work, delineating “observation stations” (as in the physical context). Also in this case, participants could observe each work within the “observation stations” without any time limitations. When they felt satisfied, they pressed a button on the controller: by doing so, the questions and sliders appeared on the scene, and they could provide their explicit judgments. See Supplementary Video S2 for an example.

At the end of the virtual tour, participants indicated how often they visit art exhibitions, and they completed two questionnaires (i.e., the SUS and ITC-SOPI) on their user experience in the virtual environment.

One year after participating in the experiment, we contacted the participants again and asked them to answer the following questions on a scale from 1 to 10: i) *How vivid is your memory of visiting the exhibition in the physical/virtual context?* ii) *How much did you like the exhibition overall in the physical/virtual context?* iii) *How much would you like to visit the exhibition again in the physical/virtual context?* Twenty-one out of twenty-four individuals responded.

2.7 Data analysis

Our primary aim was to evaluate whether explicit aesthetic responses (i.e., liking, curiosity, and emotional involvement) and viewing time were comparable between the physical and virtual context (Aim1, analysis 1). Then, we exploratorily investigated whether phenomena that are usually reported within the museum context (i.e., the *museum fatigue* and the relationship between fixation times and aesthetic appreciation) also occurred in our study in both contexts (Aim2, analysis 2). Finally, we assessed whether, one year after the experience, participants rated the vividness of their memory, their overall appreciation of the exhibition, and their willingness to revisit it similarly between contexts (Aim 3, Analysis 3). While for the third analysis we performed Bayesian paired-sample t-tests, we performed linear mixed effects model analyses to address the first two aims and support scientific considerations on the need to employ statistical models that appropriately take into account the variance due to both viewers and artworks (Specker 2025; Specker et al. 2023a); this statistical approach follows what suggested by theoretical models, which assume that art experience emerges as an interaction between artwork's features and the beholder (Leder et al. 2004; Leder & Nadal 2014; Pelowski et al. 2017a, b). Yet, as we wanted to evaluate the strength of evidence in favor of the null hypothesis (i.e., the absence of difference between physical and virtual contexts) on art experience, we employed a Bayesian approach. The rationale is to consider the Bayes Factor (BF_{01}) a statistical metric that quantifies the strength of evidence that the data provides in favor of the null hypothesis relative to the alternative one: a BF_{01} higher than 3 indicates substantial evidence supporting the null hypothesis, whereas a BF_{01} lower than 0.3 indicates substantial evidence in favor of the alternative hypothesis (Jarosz & Wiley 2014). All Bayesian linear mixed-model analyses were performed using the R package *brms* (Bürkner 2017b, a), which use the probabilistic programming language Stan. Moreover, by adopting this method, it becomes easier to avoid convergence issues that often arise in the estimation of mixed-effect models when using frequentist methods. By relying on Hamiltonian Monte Carlo (Duane et al. 1987; Neal 2011) and its extension, the No-U-Turn Sampler (NUTS) (Hoffman & Gelman 2014), these algorithms converge much more quickly than frequentist mixed models estimated through maximum likelihood optimization (Hoffman & Gelman 2014). In our analyses, even when specifying the maximal random-effects structure supported by the data, we did not encounter convergence problems (i.e., all Rhat values were lower than 1.01). The scripts and the corresponding data are available at the following link: https://osf.io/8fpvk/. Subjects and photographs were included as random-effect variables, adopting the most complex random-effect structure and estimating both the random intercept and the random slopes in all models' parameters. More details of the specific models tested in each analysis are provided below.

Although participants observed 38 artworks, all the analyses were performed on a total of 37 photographs (or sets of photographs), each realized by a different artist. The first photograph was indeed excluded from the analyses since it was employed as training for participants, to show them how to provide the evaluations (especially for the virtual context), and to correctly position the GoPro in a way it was possible to clearly distinguish – analyzing post hoc the recordings– when the participants were looking at the pictures and when at the tablet.

In both contexts, viewing times were operationalized for each photograph as the time effectively spent by the individual in observing the artwork before providing an evaluation. To compute them, we used the VSDC Free Video Editor software. We imported the videos recorded from the GoPro (physical context) and the Unity Recorder (virtual context) and analyzed them frame by frame, calculating the effective fixation duration from the moment participants entered the observation stations and started watching the images to the moment they looked away from the artworks to provide their evaluations on the tablet or with the user interface.

We removed from the analyses trials (1.24%) in which it was unclear whether participants were looking at the stimuli (e.g., because they were adjusting their headset).

2.7.1 Analysis 1

First of all, we wanted to evaluate whether art appreciation, as measured by explicit ratings, and implicit behavior towards art were comparable between physical and virtual contexts. So, for each dependent variable of interest (i.e., liking, curiosity, emotional involvement, and viewing time), the BF was computed by comparing the alternative model

(M_1) with the null model (M_0), which was identical to M_1 except for the fixed effect of context, that was excluded to estimate the evidence in favor of the absence of an effect of context on the different dimensions of art appreciation and implicit behavior. As an additional robustness check, to assess whether the order of sessions (ViewingOrder: Physical-first vs VR-first) biased the main results, we compared the best-supported model from the primary analyses with another model (M_2) in which we included the main effects of ViewingOrder, Context, and their interaction as additional fixed effects.

M_1: DV ~ 1 + Context + (1 + Context | Subject) + (1 + Context | Photograph)
M_0: DV ~ 1 + (1 + Context | Subject) + (1 + Context | Photograph)
M_2: DV ~ 1 + Context * ViewingOrder + (1 + Context | Subject) + (1 + Context | Photograph)

Where DV were, in separate analyses, judgments on liking, curiosity, emotional involvement, and viewing time.

Concerning the analyses on explicit measures, we implied student distribution. A weakly informative prior was set for the fixed effect of context (i.e., normal(0,10)) and the intercept prior (i.e., normal(50,20)). Weakly informative priors were also set for the random effects (i.e., normal(0,10)). Concerning viewing times, we implied a log-normal model (see, for example, Specker et al. 2023a), thus setting the parameters on the log scale. A weakly informative prior was set for the fixed effect of context (i.e., normal(0,0.5)) and the intercept prior (i.e., normal(2.55, 0.8). Weakly informative priors were also set for the random effects (i.e., normal(0,1)). The choices for the intercept and random effects priors were informed by existing literature, which reports median fixation times between 11 and 21 s, a mode of 10 s and a considerable variability across subjects (Smith & Smith 2001; Smith et al. 2017; Specker et al. 2023a; Tröndle & Tschacher 2012).

2.7.2 Analysis 2

To exploratorily examine whether viewing times predict art appreciation differently (or not) between physical and virtual contexts, we conducted three different Bayesian linear mixed effects models analyses with art experience ratings (i.e., liking, curiosity and emotional involvement) as dependent variables and viewing times, context, and their interaction as fixed effects. Weakly informative priors were set for the fixed effects regression coefficients (i.e., normal(0,10)) and the intercept prior (i.e., normal(50,20)). Weakly informative priors were also set for the random effects (i.e., normal(0,10)). In these analyses, we implied a student distribution, and the BF was computed by comparing the alternative model (M_1) with a null model (M_0), which was equal to M_1 but without the effect of interest – the interaction between viewing times and context – thus with just the main effects. Also for analysis 2 we performed an additional robustness check, assessing whether the order of sessions (ViewingOrder) biased the results. So, we compared the best-supported model with another model (M_2) in which we included the main effect of ViewingOrder and its interactions as additional fixed effects.

M_1: DV ~ 1 + Context + ViewingTime + Context:ViewingTime + (1 + ViewingTime + Context + Context:ViewingTime | Subject) + (1 + ViewingTime + Context + Context:ViewingTime | Photograph)
M_0: DV ~ 1 + Context + ViewingTime + (1 + ViewingTime + Context + Context:ViewingTime | Subject) + (1 + ViewingTime + Context + Context:ViewingTime | Photograph)
M_2: DV ~ 1 + Context * ViewingTime * ViewingOrder + (1 + ViewingTime + Context + Context:ViewingTime | Subject) + (1 + ViewingTime + Context + Context:ViewingTime | Photograph)

Where DV were, in separate analyses, judgments on liking, curiosity, and emotional involvement.

Then, to evaluate whether there are differences in the *museum fatigue* phenomena between contexts, we performed an analysis with viewing times as the dependent variable, photograph number, context and their interaction as fixed effects. We implied a log-normal model, setting weakly informative priors for the fixed effect regression coefficients (i.e., normal(0,0.5)) and the intercept prior (i.e., normal(2.55, 0.8). Weakly informative priors were also set for the random effects (i.e., normal(0,1)). The BF was computed by comparing the alternative model (M_1) with a null model (M_0), which was equal to M_1 but without the effect of interest, that is, the interaction between photograph number and context. We then compared the best-supported model with another model (M_2) in which we included the main effect of ViewingOrder and its interactions as additional fixed effects.

M_1: ViewingTime ~ 1 + Context + PhotographNumber + Context:PhotographNumber + (1 + Context + PhotographNumber + Context:PhotographNumber | Subject) + (1 + Context | Photograph)
M_0: ViewingTime ~ 1 + Context + PhotographNumber + (1 + Context + PhotographNumber + Context:PhotographNumber | Subject) + (1 + Context | Photograph).

M_2: ViewingTime ~ 1 + Context * PhotographNumber * ViewingOrder + (1 + Context + PhotographNumber + Context:PhotographNumber | Subject) + (1 + Context +| Photograph)

2.7.3 Analysis 3

To exploratorily investigate whether participants rated similarly between physical and virtual contexts: i) the vividness of their memory, ii) their overall appreciation of the exhibition, and iii) their willingness to revisit it, we performed three separate Bayesian paired-sample t-tests using the default Cauchy(0, 0.707) prior distribution.

3 Results

3.1 Analysis 1

Liking. Concerning M_1, the 95% credible interval around the estimate of the effect of context included the 0 (95%CI from -2.44 to 2.23), indicating that liking evaluations did not differ between contexts (see Fig. 3a). The BF_{01} was equal to 8.18, suggesting that the observed data were more than 8 times more likely to occur under M0, thus providing moderate evidence in favor of the null hypothesis (i.e., no effect of context on liking ratings). Moreover, the data were more than 23 times more likely to occur under the M_0 compared with the M_2 (BF_{02}=23.11), providing strong evidence against the effect of context, order of sessions (ViewingOrder) and their interaction on liking ratings.

Curiosity. Concerning M_1, the results showed an effect of context, suggesting we can be 95% confident of lower ratings within the virtual context compared to the physical one (95%CI from −5.13 to −0.11). The BF_{01} was 0.96, indicating that evidence in favor of M_1 remains anecdotal, requiring caution in interpreting these results. However, the Bayes factor indicated strong evidence in favor of M_2 over M_1 (BF_{12}=0.09, that is, the data are about 11 times more likely under M_2 than under M_1), suggesting that a model that also includes the order of viewing and its interaction with context can better explain the data. Indeed, the results of M_2 showed that, additionally to a main effect of context (95%CI from −8.50 to −2.45), we can be 95% confident of a context by order of viewing interaction (95%CI from 1.98 to 10.79, see Fig. 3b). Post-hoc comparisons were extracted using the *emmeans* package, and results indicate that, after having first observed the artworks in the physical context, during the second session (i.e., in VR) participants show less curiosity (95% highest posterior density intervals from 2.43 to 8.48). Such a decline in curiosity ratings was not present in participants who visited the exhibition first in VR (95% highest posterior density intervals from −4.28 to 2.40). Importantly, this was not associated with a general between-group difference, because the 95% credible interval around the estimate of the main effect of order of viewing included the 0 (95%CI from -10.85 to 9.51).

Emotional involvement. Concerning M_1, the 95% credible interval around the estimate of the effect of context

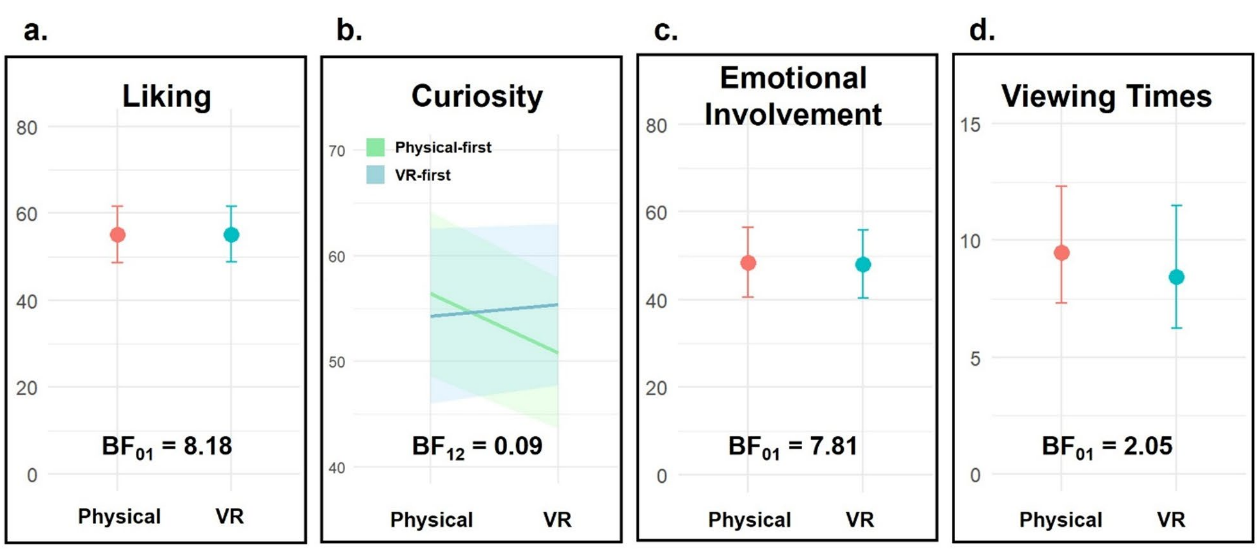

Fig. 3 The figure shows the results of Bayesian linear mixed-model analyses (Aim1) including Liking **a**., Curiosity **b**., Emotional Involvement **c**. and Viewing Times **d**. as dependent variables. Judgments were provided on a slider (see Fig. 2) ranging from 0 to 100. Error bars represent the 95% credible intervals

included the 0 (95%CI from -2.62 to 1.92), highlighting that evaluations did not differ between contexts (see Fig. 3c). The BF_{01} was 7.81, indicating moderate evidence in favor of M0, that is, that data were more likely to occur under the null hypothesis of no effect of context compared to the alternative one. Moreover, the data were more than 27 times more likely to occur under the M_0 compared with the M_2 (BF_{02} = 27.07), providing strong evidence against the effect of context, order of sessions and their interaction on emotional involvement ratings.

Viewing times. Concerning M_1, the 95% credible interval around the estimate of the effect of context included the 0 (95%CI from −0.31 to 0.07), suggesting no differences between physical and virtual exhibition (see Fig. 3d). However, the strength of evidence in favor of the null hypothesis remained at an anecdotal level (BF_{01} = 2.05), requiring caution in interpreting these results. Even if data were less likely under the M_0 compared with the M_2 (BF_{02} = 0.73), the strength of evidence in favor of M_2 remained anecdotal and the 95% credible interval around the estimate of the effect of context (95%CI from −0.45 to 0.00), order of viewing (95%CI from −0.02 to 0.81) and their interaction (95%CI from −0.09 to 0.58) included the 0.

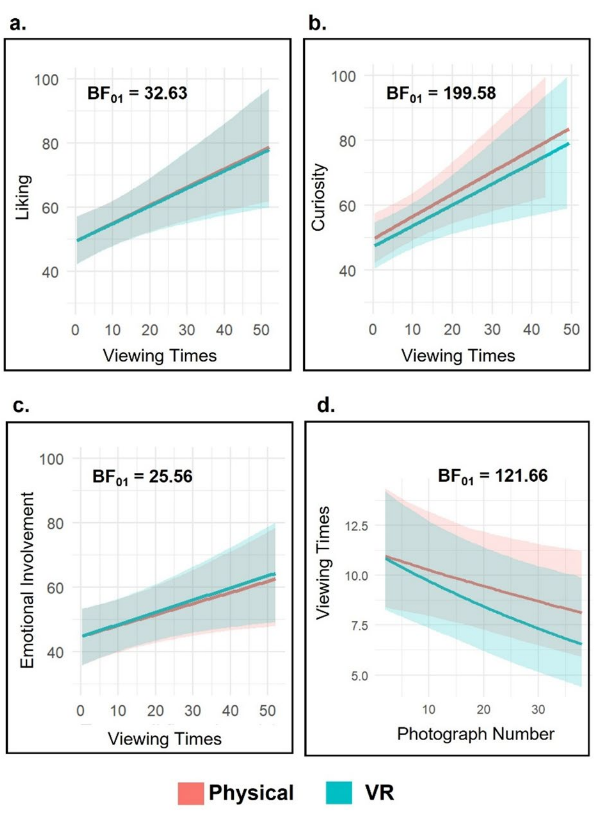

Fig. 4 The figure shows the results of the Bayesian linear mixed-model analyses (Aim2) examining the interaction effect between Viewing Times and Context on Liking (a.), Curiosity (b.), and Emotional Involvement (c.), indicating evidence in favor of the absence of a significant role of the factor Context in moderating the results. The figure also shows the results indicating the absence of an interaction effect between Photograph Number and Context on Viewing Times (d.), suggesting that the so-called *museum fatigue* effect can be replicated in virtual reality. Error bars represent the 95% credible intervals

3.2 Analysis 2

We found that viewing times predicted liking and curiosity. We can be 95% confident that, for each unit (second)-increase in viewing times, the liking judgments increase of a value between 0.20 and 1.00, and the curiosity judgments of a value between 0.31 and 1.21. Instead, concerning the emotional involvement, the 95% credible interval around the estimate of the effect of viewing times included the 0 (95%CI from −0.05 to 0.61), suggesting viewing times had no effect on judgments on the emotional component of aesthetic experience.

Concerning the viewing times by context interaction, no effect emerged in any of the three analyses, indicating that the context does not seem to moderate this effect (Fig. 4a, 4b, 4c). Indeed, the observed data were strongly more likely to occur under M_0, that is, the model that did not include the interaction, than the model that included it, for liking (BF_{01} = 32.63), curiosity (BF_{01} = 199.58), and emotional involvement (BF_{01} = 25.56). Data were also extremely more likely under the M_0 compared with the M_2 for liking (BF_{02} = 236,174.5), curiosity (BF_{01} = 294,267.5), and emotional involvement (BF_{01} = 208,308.3).

Furthermore, people tended to observe photographs for shorter times as they progressed through the exhibition and saw more artworks, confirming the *museum fatigue* phenomenon. We can be 95% confident that viewing times, on a log scale, decrease by a value between −0.0162 and −0.0009 (that is, a reduction between 1.6% and 0.1%) for each new photograph seen. Such a decline in viewing times was not likely to be different between physical and virtual contexts (see Fig. 4d). Indeed, the results showed extreme evidence (BF_{01} = 121.66) for the model M0 that did not include the photograph number by context interaction than the model that included it, as well as for the M_0 compared with the M_2 (BF_{02} = 77,285.54).

3.3 Analysis 3

Vividness of memory. The Bayes factor (BF_{01} = 4.25) showed moderate evidence in favor of the null hypothesis (95%CI from −0.45 to 0.34), suggesting that participants' evaluations on vividness of memory about the visit did not differ between physical and virtual contexts (see Fig. 5a).

Overall appreciation. The 95% credible interval around the estimate of the effect of context included the 0 (95%CI from −0.18 to 0.64), suggesting no differences between

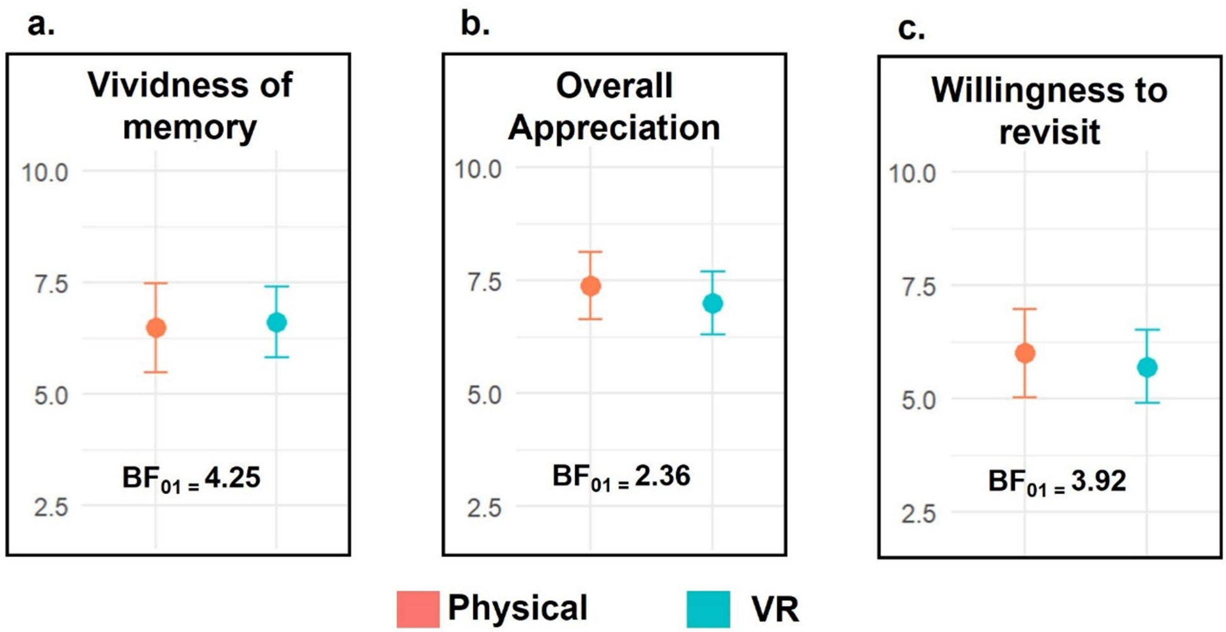

Fig. 5 The figure shows the results of the Bayesian paired sample t-test (Aim3) on Vividness of memory (**a**), Overall Appreciation (**b**), and Willingness to revisit the exhibition (**c**). Judgments were provided on a scale ranging from 0 to 10. Error bars represent the 95% credible intervals

Table 1 The table shows the descriptive statistics of the scores given by the participants in each of the four scales of the Independent Television Commission Sense of Presence Inventory (ITC-SOPI) and the System Usability Scale (SUS)

	Min	Max	Median	Mean	SD
Spatial presence (ITC-SOPI)	2.44	4.28	3.14	3.21	0.44
Engagement (ITC-SOPI)	2.92	4.38	3.73	3.71	0.36
Ecological validity (ITC-SOPI)	2.6	4.8	4.1	3.98	0.60
Negative effects (ITC-SOPI)	1	3.83	1.75	1.93	0.76
SUS	50	95	80	80.10	10.97

physical and virtual exhibition on the overall appreciation of the exhibition (see Fig. 5b). However, the strength of evidence in favor of the null hypothesis remained at an anecdotal level ($BF_{01} = 2.36$), requiring caution in interpreting these results.

Willingness to revisit the exhibition. The Bayes factor ($BF_{01} = 3.92$) showed moderate evidence in favor of the null hypothesis (95%CI from −0.30 to 0.50), suggesting that participants' evaluations on willingness to visit again the exhibition did not differ between physical and virtual contexts (see Fig. 5c).

3.4 Questionnaires

Descriptive statistics of the four scales of the ITC-SOPI and the SUS are shown in Table 1. Concerning ITC-SOPI questionnaire, on average, participants gave ratings that fall within the typical range reported in studies revealing comparable experiences between VR and real-world contexts (see, for example, Green et al. 2025). Moreover, on average, participants gave good usability (i.e., SUS) ratings (median = 80 out of 100; mean = 80.1; SD = 10.97, range from 50 to 95), since both the median and mean values align with the standard benchmark score of 80 used to mark above average system usability (Lewis 2018). Regarding the frequencies with which participants visit art exhibitions, 13 participants (54.16% of our sample) declared to visit them almost never (less than twice a year), 8 participants (33.33%) sometimes (three/four times a year), 2 participants (8.33%) often (more than once a month), and one participant (4.1%) on average once a month.

4 Discussion

The extent to which the aesthetic experience fostered within immersive virtual environments truly mirrors that evoked by the direct encounter with physical exhibitions was so far an open issue. We aimed to fill this gap by evaluating whether explicit aesthetic responses (i.e., liking, curiosity, and emotional involvement) and viewing times are comparable between visiting a physical exhibition and its virtual counterpart in an immersive environment. We also explored whether phenomena such as *museum fatigue* and the relationship between viewing time and aesthetic responses occur both during virtual and physical visits, and we evaluated whether, one year after the experience, participants' ratings on the vividness of their memory, their overall exhibition appreciation, and their willingness to visit again the exhibition were similar between physical and virtual contexts. To address these questions, we employed a Bayesian approach in data analysis, which allows for quantifying support for one model over an alternative one and thus possibly providing evidence in favor of the null hypothesis, that is, in the context of the present experiment, of similar results between aesthetic experience taking place in the virtual and physical space.

Regarding explicit judgments, our results provided moderate evidence supporting the absence of a difference between physical and virtual contexts in terms of liking and emotional involvement. Concerning curiosity, instead, the physical context seemed to lead to greater curiosity toward artistic photographs compared to the virtual one. This effect, however, was influenced by the order of viewing (i.e., whether participants started from the physical or the virtual exhibition): after having first encountered the artworks in the physical context, during the second session (i.e., in VR) participants showed less curiosity towards the photographs. Such a decline in curiosity was not present in participants who visited the exhibition first in VR. These findings are consistent with a previous study, in which participants, having encountered the artworks in the museum a week before, found the laboratory experience less interesting (Brieber et al. 2015a, b). In this regard, it is important to acknowledge that, in our experiment, unlike the physical context, the labels displaying the title of the artwork and the artist's name in the virtual context were not clearly legible. Previous literature suggests that, for non-experts in contemporary art, the presence or absence of minimally informative labels (which only provide the title of the artwork) may not affect levels of liking or emotional responses to contemporary art (Leder et al. 2006; Szubielska et al. 2021; Szubielska & Imbir 2021), nor influence interest ratings (Gerger & Leder 2015), but rather impact understanding (Leder et al. 2006). So, these findings underscore that even minimal information conveyed by labels can serve as a tool for achieving a greater understanding and reducing uncertainty. Furthermore, some participants reported that, for smaller artworks, certain details in the photographs within the virtual context were not always clearly distinguishable. These aspects may explain part of the differences in curiosity judgments. According to the predictive processing framework, curiosity increases when subjects implicitly expect that, by gaining more information, artworks can become more understandable (Frascaroli et al. 2024; Van de Cruys et al. 2024). Therefore, it could be possible that individuals who visited the VR environment first and then encountered the artworks a second time in the physical context may have maintained a high level of curiosity due to the possibility of gathering additional perceptual/visual details (e.g., small details difficult to discern in VR) and textual information (like labels with the artwork's title that may not be clearly legible in VR) to achieve a greater understanding. On the other hand, those who have already gathered this information during the first visit in the physical space might have less to discover (and to understand) on their second visit (in VR), losing curiosity in the images. Even if this is a speculative interpretation, further studies will be needed in the future to clarify this result. Future studies should also take into account understanding ratings.

Finally, viewing times did not appear to differ between the two contexts, although the evidence supporting the absence of a difference remained anecdotal and requires further studies before being conclusively established.

Regarding our second aim, our results provided strong evidence against the moderation effect of context on certain phenomena related to the art experience. Specifically, we found that in both the physical and virtual contexts i) visitors tended to observe artworks for progressively shorter durations as they moved through the exhibition, replicating the effect defined as *museum fatigue* by previous studies (Bitgood 2009), and ii) viewing times predicted some explicit judgments (i.e., liking and curiosity) related to the aesthetic experience. It is worth noting that previous studies exploring the link between art appreciation and fixation times focused on the more cognitive dimensions, such as liking, interest, ambiguity, and understanding, overlooking the emotional ones (Brieber et al. 2014). Our results replicate these findings on the link between viewing times and the cognitive components of the aesthetic experience (i.e., liking), while the emotional one seems to be less influenced by the participants' looking behavior, an effect that may be worth further exploring in future research. At the same time, we acknowledge that longer viewing times may reflect attempts to resolve interpretive uncertainty (coherently, we found a positive relationship between curiosity and viewing times). Future studies should therefore include understanding measures to better disentangle the respective contributions of appreciation-related responses and meaning-making difficulty to viewing behavior.

Finally, we found evidence of no difference between the two environments in terms of vividness of memory of the visit one year after the experience and willingness to revisit the exhibition.

In an era increasingly defined by the transformative possibilities of emerging technologies (Anwar et al. 2025), these findings, even if preliminary, could have profound implications on both scientific and cultural levels.

From a scientific standpoint, VR could represent a powerful tool that could overcome the inherent limitations of traditional laboratory settings (Kuliga et al. 2015), offering the possibility of studying the psychological processes involved in aesthetic experiences from a more ecologically valid perspective and within a controlled environment. Moreover, VR opens up methodological possibilities, enabling investigating the effect of variables that are otherwise unfeasible to manipulate in the real world, while overcoming practical challenges that often limit research in physical settings.

On a cultural level, the implications of these results could be equally significant. In recent years, there has been a growing push to leverage digital technologies to preserve and promote cultural heritage (Bekele et al. 2018). Events such as the fire at Notre-Dame Cathedral in Paris, coupled with the COVID-19 pandemic and war conflicts, have catalyzed reflection on the urgent need to digitally safeguard cultural landmarks and ensure their virtual accessibility to a global audience, breaking down physical, geographical, and generational barriers while promoting its implementation in educational, tourism, and creative sectors (see https://digital-strategy.ec.europa.eu/en/news/commission-proposes-common-european-data-space-cultural-heritage). The digitization of cultural heritage is on the European agenda, with precise European Commission's recommendations to the Members State and several Horizon Europe funding calls in 2023, 2024 and 2025 (European Commission: Directorate-General for Research and Innovation 2022, 2024). More and more cultural institutions are embracing this digital shift by offering virtual exhibitions and online tours to make their collections more accessible. The digitization of art and cultural sites has gained momentum in recent years and is set to become an increasingly prevalent way of experiencing and engaging art in the future, complementing traditional methods and broadening the scope of access for global audiences. For all these reasons, we think investigating the extent to which the aesthetic experience fostered by immersive virtual environments can truly replicate the encounter with physical art exhibitions is of great interest. Addressing this question is crucial for realizing the full potential of VR in the cultural sector: it is essential to evaluate whether shifting the medium through which we experience art and culture alters the quality or certain aspects of that experience. By understanding these potential differences, we can help curators and institutions design virtual exhibitions that not only mirror physical ones but are also tailored to the unique attributes of virtual environments, ensuring that these digital platforms offer experiences as rich as their physical counterparts.

5 Limitations and future directions

While further studies with larger samples are needed, our findings provide initial evidence supporting the comparability of the aesthetic experience between a physical environment and its immersive virtual counterpart. Nevertheless, we recognize that our study presents some constraints and limitations, which limit the generalizability of the findings. First, it is worth keeping in mind that our physical context was not a museum or an art gallery. The physical space of the museum itself embodies culturally relevant factors, creating an optimal context for experiencing art that fosters an "aesthetic attitude" (Cupchik & László, 1992). This effect is known as the *white cube effect,* referring to the fact that a museum setting can transform any ordinary object into an art object that warrants aesthetic consideration (Leder et al. 2004). Previous research has also discussed the potential of an *inverse white cube effect*, particularly in relation to artistic photographs (Brieber et al. 2015a, b): artworks displayed in an unconventional physical context can, in turn, elevate that space into a museum (Brieber et al. 2015a, b). This effect may have made our participants feel like they were in a gallery. In any case, future studies should compare the aesthetic experience of visiting an exhibition in a physical museum/gallery with its virtual counterpart.

Moreover, to isolate the effect of context (physical vs virtual) while avoiding possible confounds related to the texture and materiality information conveyed by the pieces of art, our study employed photographic artworks as stimuli. Photographs, being two-dimensional, are easier to render in VR compared to paintings (especially those with impasto) and sculptures. Future studies are thus necessary to understand how transferable the results are to different artistic stimuli, where the material elements (materials used, supports, brushstrokes) could play a fundamental role.

Furthermore, in the attempt to increase control over the experimental design, we acknowledge that some of our choices may have reduced the ecological aspect of the experience. To be considered a fully ecological experience, participants should be able to freely explore the exhibition in time and space (Carbon 2017, 2019, 2020). This means allowing participants to observe artworks for as long as they wish, enabling them to move within the exhibition space without constraints. It implies, among other things, allowing them to return to the same artwork multiple times, as well as to move closer or further away from the works themselves. Instead, in our experiment, we constrained participants to follow a pre-established path during both the physical and virtual tours, thereby limiting the effective freedom of the experience. Future studies could allow participants to experience and visit the virtual and physical exhibition in a truly free way, investigating whether explorative behaviors are comparable between contexts. Rodriguez-Boerwinkle and colleagues recently developed an open-source non-immersive virtual gallery, finding that participants navigated the online gallery and interacted with the artworks in ways consistent with typical visitor behaviours in real-world art museums (Rodriguez-Boerwinkle et al. 2022). However, on this point, the literature lacks a direct comparison between physical galleries and their immersive digital twins. Additionally, our sample predominantly consisted of young participants not particularly interested in art, declaring themselves to attend art exhibitions rarely. Therefore, it would be valuable to investigate whether different results emerge among museum enthusiasts, frequent exhibition-goers, and art experts. Previous research has indeed shown that, while art-trained individuals rated genuine artworks differently from reproductions, this was not the case for laypeople (Locher et al. 2001), suggesting that genuineness may matter more for those with a deeper knowledge of art. If art lovers should experience art differently within immersive virtual environments compared to traditional museum settings, such findings could offer invaluable insights for virtual exhibition curators. Certain exhibitions, by their very nature, tend to attract a broader audience of less art-savvy individuals, while others may appeal to a more niche audience of professionals. Tailoring specific strategies based on the exhibition type and the target audience could significantly enhance the overall experience, ensuring that it remains as satisfying and engaging as possible. Finally, in our experiment, participants visited the exhibition individually in both physical and virtual contexts. However, we recognize that the social aspect—specifically, the experience of visiting an exhibition with others—plays a significant role in shaping the overall experience (see, for instance, Tröndle et al. 2012). Looking ahead, the potential of the Metaverse will allow for the simulation of such social interactions in immersive virtual environments, enabling people to attend virtual exhibitions together. This advancement will create an opportunity to explore whether the social dimension of the experience holds different significance in physical versus virtual settings.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10055-026-01372-9.

Acknowledgments We are grateful to the BreraBicocca project (http://www.brerabicocca.it) for making this work possible.

Author contributions Author contributions. G.T.: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing—original draft. A.M.: Conceptualization, Funding acquisition. C.L.: Conceptualization. S.P.: Conceptualization. E.P.: Conceptualization, Methodology, Writing—original draft. L.M.S.: Conceptualization, Methodology, Formal analysis, Writing—original draft, Supervision. All authors reviewed the manuscript.

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement. The project was funded by the Italian Ministry of Research (MUR), Dipartimento di Eccellenza 2023–2027 (Project Number 2023-NAZ-0206).

Data availability All data are available in the Open Science Framework at the following address: https://osf.io/8fpvk/.

Declarations

Conflict of interests Authors declare that they have no competing interests.

References

Anwar MS, Yang J, Frnda J, Choi A, Baghaei N, Ali M (2025) Metaverse and XR for cultural heritage education: applications, standards, architecture, and technological insights for enhanced immersive experience. Virtual Real. https://doi.org/10.1007/s10055-025-01126-z

Bekele MK, Pierdicca R, Frontoni E, Malinverni ES, Gain J (2018) A survey of augmented, virtual, and mixed reality for cultural heritage. J Comput Cult Herit 11(2):1–36. https://doi.org/10.1145/3145534

Bitgood S (2009) Museum fatigue: a critical review. Visit Stud 12(2):93–111. https://doi.org/10.1080/10645570903203406

Brieber D, Forster M, Leder H (2020) On the mutual relation between art experience and viewing time. Psychol Aesthet Creat Arts 14(2):197–208. https://doi.org/10.1037/aca0000210

Brieber D, Leder H, Nadal M (2015a) The experience of art in museums. Empir Stud Arts 33(1):95–105. https://doi.org/10.1177/0276237415570000

Brieber D, Nadal M, Leder H (2015b) In the white cube: Museum context enhances the valuation and memory of art. Acta Psychol Amst 154:36–42. https://doi.org/10.1016/j.actpsy.2014.11.004

Brieber D, Nadal M, Leder H, Rosenberg R (2014) Art in time and space: context modulates the relation between art experience and viewing time. PLoS ONE 9(6):e99019. https://doi.org/10.1371/journal.pone.0099019

Brooke J (1996). SUS–A quick and dirty usability scale . In Taylor & Francis (Ed.), *Usability evaluation in industry* (pp. 189–194).

Bürkner PC. (2017a). Advanced bayesian multilevel modeling with the R package brms.

Bürkner PC (2017b) Brms: an R package for bayesian multilevel models using Stan. J Stat Softw. https://doi.org/10.18637/jss.v080.i01

Carbon C-C (2017) Art perception in the museum: how we spend time and space in art exhibitions. I-Perception. https://doi.org/10.1177/2041669517694184

Carbon C-C (2019) Empirical approaches to studying art experience. J Percept Imaging 2(1):010501-1-010501–010507. https://doi.org/10.2352/J.Percept.Imaging.2019.2.1.010501

Carbon C-C (2020) Ecological art experience: how we can gain experimental control while preserving ecologically valid settings and contexts. Front Psychol. https://doi.org/10.3389/fpsyg.2020.00800

Castellotti S, D'Agostino O, Mencarini A, Fabozzi M, Varano R, Mastandrea S, Baldriga I, Del Viva MM (2023) Psychophysiological and behavioral responses to descriptive labels in modern art museums. PLoS ONE 18(5):e0284149. https://doi.org/10.1371/journal.pone.0284149

Chatterjee A, Vartanian O (2014) Neuroaesthetics. Trends Cogn Sci 18(7):370–375. https://doi.org/10.1016/j.tics.2014.03.003

Chatterjee A, Vartanian O (2016) Neuroscience of aesthetics. Ann N Y Acad Sci 1369(1):172–194. https://doi.org/10.1111/nyas.13035

Clifton J, Palmisano S (2020) Effects of steering locomotion and teleporting on cybersickness and presence in HMD-based virtual reality. Virtual Real 24(3):453–468. https://doi.org/10.1007/s10055-019-00407-8

Cupchik GC, László J (1992) Emerging visions of the aesthetic process: In psychology, semiology, and philosophy. Cambridge University Press

Darda KM, Estrada Gonzalez V, Christensen AP, Bobrow I, Krimm A, Nasim Z, Cardillo ER, Perthes W, Chatterjee A (2025) A comparison of art engagement in museums and through digital media. Sci Rep 15(1):8972. https://doi.org/10.1038/s41598-025-93630-0

Duane S, Kennedy AD, Pendleton BJ, Roweth D (1987) Hybrid Monte Carlo. Phys Lett B 195(2):216–222. https://doi.org/10.1016/0370-2693(87)91197-X

European Commission: Directorate-General for Research and Innovation. (2022). European collaborative cloud for cultural heritage. https://data.europa.eu/doi/https://doi.org/10.2777/83783

European Commission: Directorate-General for Research and Innovation. (2024). Horizon Europe strategic plan 2025–2027. https://data.europa.eu/doi/https://doi.org/10.2777/092911

European Commission: Directorate-General for Research and Innovation. (2024). Shaping the future of cultural heritage : exploring the nexus between the cultural heritage cloud, the common European data space for cultural heritage, and the European open science cloud. https://data.europa.eu/doi/https://doi.org/10.2777/12321

Frascaroli J, Leder H, Brattico E, Van de Cruys S (2024) Aesthetics and predictive processing: grounds and prospects of a fruitful encounter. Philos Trans R Soc B Biol Sci. https://doi.org/10.1098/rstb.2022.0410

Gerger G, Leder H (2015) Titles change the esthetic appreciations of paintings. Front Human Neurosci. https://doi.org/10.3389/fnhum.2015.00464

Goller J, Mitrovic A, Leder H (2019) Effects of liking on visual attention in faces and paintings. Acta Psychol Amst 197:115–123. https://doi.org/10.1016/j.actpsy.2019.05.008

Green AD, Clark A, Rechdan J, Guppy A (2025) Virtual reality provides an eyewitness experience that is similar to real life. Appl Cognit Psychol. https://doi.org/10.1002/acp.70099

Grüner S, Specker E, Leder H (2019) Effects of context and genuineness in the experience of art. Empir Stud Arts 37(2):138–152. https://doi.org/10.1177/0276237418822896

Gulhan D, Durant S, Zanker JM (2021) Similarity of gaze patterns across physical and virtual versions of an installation artwork. Sci Rep 11(1):18913. https://doi.org/10.1038/s41598-021-91904-x

Hoffman MD, Gelman A (2014) The No-U-turn sampler: adaptively setting path lengths in Hamiltonian monte carlo. J Mach Learn Res 15(1):1593–1623

Jarosz AF, Wiley J (2014) What are the odds? a practical guide to computing and reporting bayes factors. J Probl Solv. https://doi.org/10.7771/1932-6246.1167

Kenett YN, Humphries S, Chatterjee A (2023) A thirst for knowledge: grounding curiosity, creativity, and aesthetics in memory and reward neural systems. Creat Res J 35(3):412–426. https://doi.org/10.1080/10400419.2023.2165748

Kuliga SF, Thrash T, Dalton RC, Hölscher C (2015) Virtual reality as an empirical research tool—exploring user experience in a real building and a corresponding virtual model. Comput Environ Urban Syst 54:363–375. https://doi.org/10.1016/j.compenvurbsys.2015.09.006

Leder H, Belke B, Oeberst A, Augustin D (2004) A model of aesthetic appreciation and aesthetic judgments. Br J Psychol 95(4):489–508. https://doi.org/10.1348/0007126042369811

Leder H, Carbon C-C, Ripsas A-L (2006) Entitling art: influence of title information on understanding and appreciation of paintings. Acta Psychol 121(2):176–198. https://doi.org/10.1016/j.actpsy.2005.08.005

Leder H, Nadal M (2014) Ten years of a model of aesthetic appreciation and aesthetic judgments: the aesthetic episode–developments and challenges in empirical aesthetics. Br J Psychol 105(4):443–464. https://doi.org/10.1111/bjop.12084

Leder H and Pelowski M (2021). Empirical aesthetics. In: The oxford handbook of empirical aesthetics (pp. 921–942). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198824350.013.43

Lessiter J, Freeman J, Keogh E, Davidoff J (2001) A cross-media presence questionnaire: the ITC-sense of presence inventory. Presence Teleoperators Virtual Environ 10(3):282–297. https://doi.org/10.1162/105474601300343612

Lewis JR (2018) The system usability scale: past, present, and future. Int J Hum Comput Interact 34(7):577–590. https://doi.org/10.1080/10447318.2018.1455307

Locher P, Dolese M (2004) A comparison of the perceived pictorial and aesthetic qualities of original paintings and their postcard images. Empir Stud Arts 22(2):129–142. https://doi.org/10.2190/EQTC-09LF-JRHA-XKJT

Locher P, Smith J, Smith L (2001) The influence of presentation format and viewer training in the visual arts on the perception of pictorial and aesthetic qualities of paintings. Perception 30(4):449–465. https://doi.org/10.1068/p3008

Locher P, Smith L, Smith J (1999) Original paintings versus slide and computer reproductions: a comparison of viewer responses. Empir Stud Arts 17(2):121–129. https://doi.org/10.2190/R1WN-TAF2-376D-EFUH

Loewenstein G (1994) The psychology of curiosity: a review and reinterpretation. Psychol Bull 116(1):75–98. https://doi.org/10.1037/0033-2909.116.1.75

Marín-Morales J, Higuera-Trujillo JL, Greco A, Guixeres J, Llinares C, Gentili C, Scilingo EP, Alcañiz M, Valenza G (2019) Real vs. immersive-virtual emotional experience: analysis of psycho-physiological patterns in a free exploration of an art museum. PLoS ONE 14(10):e0223881. https://doi.org/10.1371/journal.pone.0223881

Mikuni J, Specker E, Pelowski M, Leder H, Kawabata H (2022) Is there a general "art fatigue" effect? A cross-paradigm, cross-cultural study of repeated art viewing in the laboratory. Psychol Aesthet Creat Arts 16(2):343–360. https://doi.org/10.1037/aca0000396

Neal RM (2011). MCMC Using Hamiltonian dynamics. In: handbook of Markov chain Monte Carlo (pp. 113–162). Chapman and Hall/CRC. https://doi.org/10.1201/b10905-6

Pearce MT, Zaidel DW, Vartanian O, Skov M, Leder H, Chatterjee A, Nadal M (2016) Neuroaesthetics. Perspect Psychol Sci 11(2):265–279. https://doi.org/10.1177/1745691615621274

Pelowski M, Forster M, Tinio PPL, Scholl M, Leder H (2017a) Beyond the lab: an examination of key factors influencing interaction

with ‘real’ and museum-based art. Psychol Aesthet Creat Arts 11(3):245–264. https://doi.org/10.1037/aca0000141
Pelowski M, Leder H, Mitschke V, Specker E, Gerger G, Tinio PPL, Vaporova E, Bieg T, Husslein-Arco A (2018) Capturing aesthetic experiences with installation art: an empirical assessment of emotion, evaluations, and mobile eye tracking in Olafur Eliasson’s “Baroque, Baroque!” Front Psychol. https://doi.org/10.3389/fpsyg.2018.01255
Pelowski M, Markey PS, Forster M, Gerger G, Leder H (2017b) Move me, astonish me… delight my eyes and brain: the Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. Phys Life Rev 21:80–125. https://doi.org/10.1016/j.plrev.2017.02.003
Robinson ES (1928) The behavior of the museum visitor. American Association of Museums, Washington, DC
Rodriguez-Boerwinkle RM, Boerwinkle MJ, Silvia PJ (2022) The open gallery for arts research (OGAR): an open-source tool for studying the psychology of virtual art museum visits. Behav Res Methods 55(2):824–842. https://doi.org/10.3758/s13428-022-01857-w
Schönbrodt FD, Wagenmakers E-J (2018) Bayes factor design analysis: Planning for compelling evidence. Psychon Bull Rev 25(1):128–142. https://doi.org/10.3758/s13423-017-1230-y
Shemyakina NV, Nagornova ZhV, Grokhotova AV, Galkin VA, Vasenkina VA, Biryukova SV, Potapov YuG (2024) A study of EEG characteristics of aesthetic perception and assessment of painting works during museum visits: neuroaesthetic study. Hum Physiol 50(4):345–357. https://doi.org/10.1134/S0362119724700798
Smith JK, Smith LF (2001) Spending time on art. Empir Stud Arts 19(2):229–236. https://doi.org/10.2190/5MQM-59JH-X21R-JN5J
Smith LF, Smith JK, Tinio PPL (2017) Time spent viewing art and reading labels. Psychol Aesthet Creat Arts 11(1):77–85. https://doi.org/10.1037/aca0000049
Specker E (2025) A personal perspective on psychology of aesthetics and the arts: ecologically valid, interdisciplinary, and diverse methodologies. Creat Res J 37(2):293–300. https://doi.org/10.1080/10400419.2023.2269339
Specker E, Arató J, Leder H (2023a) How are real artworks and reproductions judged? The role of anchoring in empirical investigations of the genuineness effect. J Exp Soc Psychol 108:104494. https://doi.org/10.1016/j.jesp.2023.104494
Specker E, Fekete A, Trupp MD, Leder H (2023b) Is a “real” artwork better than a reproduction? A meta-analysis of the genuineness effect. Psychol Aesthet Creat Arts 17(3):294–306. https://doi.org/10.1037/aca0000399
Specker E, Tinio PPL, van Elk M (2017) Do you see what I see? An investigation of the aesthetic experience in the laboratory and museum. Psychol Aesthet Creat Arts 11(3):265–275. https://doi.org/10.1037/aca0000107
Szubielska M, Imbir K (2021) The aesthetic experience of critical art: the effects of the context of an art gallery and the way of providing curatorial information. PLoS ONE 16(5):e0250924. https://doi.org/10.1371/journal.pone.0250924
Szubielska M, Imbir K, Szymańska A (2021) The influence of the physical context and knowledge of artworks on the aesthetic experience of interactive installations. Curr Psychol 40(8):3702–3715. https://doi.org/10.1007/s12144-019-00322-w
Tröndle M, Tschacher W (2012) The physiology of phenomenology: the effects of artworks. Empir Stud Arts 30(1):75–113. https://doi.org/10.2190/EM.30.1.g
Tröndle M, Wintzerith S, Wäspe R, Tschacher W (2012) A museum for the twenty-first century: the influence of ‘sociality’ on art reception in museum space. Mus Manag Curat 27(5):461–486. https://doi.org/10.1080/09647775.2012.737615
Van de Cruys S, Frascaroli J, Friston K (2024) Order and change in art: towards an active inference account of aesthetic experience. Philos Trans R Soc Lond B Biol Sci. https://doi.org/10.1098/rstb.2022.0411
Vasishth S, Yadav H, Schad DJ, Nicenboim B (2023) Sample size determination for Bayesian hierarchical models commonly used in psycholinguistics. Comput Brain Behav 6(1):102–126. https://doi.org/10.1007/s42113-021-00125-y
Vessel EA, Starr GG, Rubin N (2012) The brain on art: intense aesthetic experience activates the default mode network. Front Human Neurosci. https://doi.org/10.3389/fnhum.2012.00066
Vessel EA, Starr GG, Rubin N (2013) Art reaches within: aesthetic experience, the self and the default mode network. Front Neurosci. https://doi.org/10.3389/fnins.2013.00258
Vogl E, Pekrun R, and Loderer K (2021). Epistemic emotions and metacognitive feelings. In: trends and prospects in metacognition research across the life span (pp. 41–58). Springer International Publishing. https://doi.org/10.1007/978-3-030-51673-4_3
Wagenmakers EJ, Marsman M, Jamil T, Ly A, Verhagen J, Love J, Selker R, Gronau QF, Šmíra M, Epskamp S, Matzke D, Rouder JN, Morey RD (2018) Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. Psychon Bull Rev 25(1):35–57. https://doi.org/10.3758/s13423-017-1343-3
Wang F, Gelfand AE (2002) A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. Stat Sci 17:193–208
Zaidel DW (2018). Culture and art: importance of art practice, not aesthetics, to early human culture (pp. 25–40). https://doi.org/10.1016/bs.pbr.2018.03.001