DECODING THE EPISTEMIC ARC: DISTINCT PHYSIOLOGICAL SIGNATURES FOR CURIOSITY, INSIGHT, UNDERSTANDING AND LIKING DURING INTERACTIONS WITH VISUAL ART

Dominik Welke
d.welke@leeds.ac.uk

School of Psychology
University of Leeds
Leeds, UK

Department of Neuroscience
Max Planck Institute for Empirical Aesthetics
Frankfurt (Main), Germany

Edward A. Vessel
evessel@ccny.cuny.edu

Department of Psychology
City College of New York (CCNY)
New York City, USA

Department of Neuroscience
Max Planck Institute for Empirical Aesthetics
Frankfurt (Main), Germany

ABSTRACT

Recent work on insight and information processing suggested feelings of *aha* as a metacognitive feedback signal. A series of processes, from curiosity (drive state) to insight (uncertainty reduction) to pleasure (reward and reinforcement), might represent a fundamental epistemic arc for motivated learning. Here, we present a paradigm that combines measurement of curiosity, insight (*aha*), understanding and liking to outline the contours of this epistemic arc and study their neuronal mechanisms. In two preregistered experiments we employ the so called "title effect" – the fact that additional semantic information accompanying an artwork (such as titles) can change how an observer understands and enjoys a piece of visual art. Participants viewed paintings (5s trials) and rated their curiosity for seeing the title. This was followed either by the original title or a dummy title lacking additional information (untitled), and a second presentation of the painting, after which participants rated strength of *aha*, aesthetic liking, and feeling of understanding the artwork.

In a behavioral online study (N=55 participants) we replicate a set of previous findings and established that visual art together with their titles can prompt strong feelings of *aha*. The association between the collected ratings is complex, partly nonlinear, and shaped by the type and amount of information provided by the stimulus material. In an EEG, ECG and eye tracking study with the same paradigm (N=49 participants) we aimed to characterize the temporal sequence of processing. Using a time-resolved decoding approach we were able to recover the categorical experimental conditions and the graded ratings: curiosity for the title could be decoded late in pretitle trials and early in the posttitle trials; *aha* could be decoded in posttitle trials; felt understanding had the most robust signature among the ratings and could be decoded from multiple signals in both pre and posttitle trials. Lastly, ratings of aesthetic liking were hard to decode from the data with only a short significant timewindow early during posttitle trials.

Keywords MVPA decoding · cognition · information processing · insight · visual aesthetics · EEG · multi-modal data

1 Intro

We are constantly confronted with novel, potentially ambiguous information, and making sense of these stimuli (i.e. forming a coherent, meaningful representation) is a cornerstone of human cognition. Often, sense-making involves insight – a sudden jump in understanding, such as when the solution to a previously unsolved problem becomes clear, or when a new piece of information allows for better understanding (Sternberg & Davidson, 1995). Salient examples can be found in the visual domain, where our perception of an ambiguous stimulus can suddenly change and allow us to perceive structure (e.g. with Mooney images; Mooney & Ferguson, 1951).

Insight could be characterized as a rapid decrease in uncertainty, or gain in information. Recent work has linked insight to the preceding state of curiosity (Van de Cruys et al., 2021): the more curiosity, the greater the *aha* – a measure of the phenomenal feeling that accompanies insight. This finding places insight within a motivational framework: curiosity is a drive state to reduce uncertainty, improved understanding is the desired outcome, and insight (or *aha*) marks moments of particularly effective uncertainty reduction. Based on their findings Van de Cruys et al. (2021) hypothesize that *aha* might act as a meta-cognitive feedback signal that can help us optimize learning efficiency and allocate limited cognitive resources to solvable tasks (see also Laukkonen et al., 2023). If insight triggers a motivational process, it should engage reward processes and be associated with pleasure – the pleasure of understanding. Indeed, some studies provide empirical evidence that *aha* can feel pleasurable (Skaar & Reber, 2020; Shen et al., 2016) and is linked to the reward system (Cristofori et al., 2018; Oh et al., 2020).

This series of processes, from curiosity to insight to pleasure, represents a fundamental epistemic arc for motivated learning. Here, we present a paradigm that combines measurement of curiosity, insight (*aha*), understanding and liking to outline the contours of this epistemic arc and study their neuronal mechanisms.

Van de Cruys et al. (2021) investigated insight and information processing using a relatively low-level task of perceptual understanding in the visual domain (Mooney images – black-and-white two tone images derived from photographs that are initially hard to interpret, but seeing the original photography can lead to a perceptual shift; Mooney & Ferguson, 1951). Yet, similar mechanisms can act higher up the cognitive hierarchy. For example, insight might be caused by combining information from different sensory modalities or by contextualization of a stimulus with prior knowledge or personal memories. In this study we use visual artworks and employ a phenomenon sometimes called the "title effect" – the fact that text based, semantic information supplementing an artwork (such as its title) can affect how an observer understands, interprets or judges a piece of visual art. In fact, insight might be a key factor in the feelings of delight or being moved many people can get from savoring art. For Mooney images it has been shown that the strength of *aha* with a particular image predicts subsequent ratings of aesthetic liking (Muth & Carbon, 2013).

1.1 The "Title effect" – a case of aesthetic *aha*?

Several psychological studies have investigated the effects of titles on different ratings of visual art. While all studies we found have reported significant effects of the applied title manipulation on one or several of their dependent measures, the findings are not entirely conclusive.

The clearest picture emerges regarding ratings of felt understanding or meaningfulness of the artwork: a total of 10 experiments in 6 studies found a significant effect of title manipulation (Russell & Milne, 1997; Russell, 2003; Millis, 2001; Leder et al., 2006; Mullennix et al., 2018; Szubielska et al., 2021), while only 2 experiments in 2 studies did not. Yet also in the studies reporting overall non-significant results, differences between title and no title condition were present (see Russell, 2003, exp. 1 and Bubi et al., 2017). Taken together, these studies provide convincing evidence that additional semantic information given in text form changes observer's subjective understanding of paintings.

When looking at aesthetic preference ratings (e.g. liking, being moved, etc.) the findings are less clear: several studies reported significantly higher ratings for presenting art with a title than without a title (Millis, 2001; Leder et al., 2006; Belke et al., 2010; Hristova et al., 2011; Gerger & Leder, 2015; Mastandrea & Umiltà, 2016; Bubi et al., 2017).

However, in many cases these effects only occurred for specific types of titles: for elaborative titles, but not descriptive or random titles (Millis, 2001; Leder et al., 2006), for semantically related, but not unrelated titles (Belke et al., 2010; Gerger & Leder, 2015), or for fabricated titles emphasizing a specific topic of the work, but not the original titles or titles deemphasizing this concept (Mastandrea & Umiltà, 2016). Furthermore, the category of artwork might matter: one study explicitly observed a difference between matching and non-matching titles only for abstract art, but not for semi-abstract or figurative art (Gerger & Leder, 2015) while most other studies were restricted to very homogeneous stimulus sets and could hence not clarify this issue. Several other studies report null results under seemingly very similar conditions: comparing titles to no titles (Russell & Milne, 1997; Millis, 2001; Russell, 2003), descriptive to more elaborate titles (Millis, 2001; Leder et al., 2006; Mullennix et al., 2018), matching to non-matching titles (Szubielska et al., 2021), or titles alone to longer text descriptions (Russell, 2003).

While none of these studies collected ratings of *aha* or insight – an effect to be established in the current study – previous work has suggested a link between insight and aesthetic liking for lower level perceptual reveals: Muth and Carbon (2013) observed that the strength of *aha* with Mooney images positively predicted ratings of aesthetic preference for the stimuli and termed this the "aesthetic aha effect".

1.2 Physiological correlates of insight and aesthetic preference

Several studies have tried to identify neurophysiological correlates of insight or aesthetic preferences, respectively.

Previous research on the neuroscience of insight has often focused on hemispheric differences; a hypothesis compatible with a set of neuroanatomical and cytoarchitectonic asymmetries in potentially relevant brain regions (especially in the context of language processing; for a review see Kounios & Beeman, 2014). However, there is a small EEG literature on insight (mainly from two research groups) demonstrating relatively consistent findings - namely bursts of gamma-band activity preceding self-reported insight moments in various text-based tasks such as remote associate tasks (Jung-Beeman et al., 2004; Sandkühler & Bhattacharya, 2008), verbal puzzles (Sheth et al., 2009), or anagram solving tasks (Oh et al., 2020). In addition to gamma-band activity, these studies less consistently reported increased alpha-power (Jung-Beeman et al., 2004; Sandkühler & Bhattacharya, 2008, both using remote associate tasks) and increased (Oh et al., 2020) or decreased (Sheth et al., 2009) beta-activity prior to the solution.

These activity patterns coarsely resemble those reported in the small frequency-domain M/EEG literature on aesthetic preferences. Studies reported significant effects in almost all of the canonical frequency-bands, but most consistently in the gamma (Lopez-Persem et al., 2020; Cela-Conde et al., 2004; Munar, Nadal, Rosselló, et al., 2012; Munar, Nadal, Castellanos, et al., 2012; Strijbosch et al., 2021) and alpha range (Munar, Nadal, Castellanos, et al., 2012; Kang et al., 2015; Strijbosch et al., 2021). Furthermore, hemispheric differences in alpha power (frontal alpha asymmetry) have been a particular target of neuroaesthetic investigation (Babiloni et al., 2013; Moon et al., 2013; Cheung et al., 2019), and some studies even decided to use an EEG asymmetry index as proxy for aesthetic preference without further verification through behavioral ratings (Babiloni et al., 2015; Lee et al., 2017).

Event related potentials (ERPs) have also been investigated in the context of aesthetic preferences. Effects on a variety of well established ERP components have been reported including posterior P2 (Righi et al., 2017; Noguchi & Murota, 2013), anterior P2 (Righi et al., 2017; Ma et al., 2015; Jiang & Cai, 2013; Wang et al., 2012), N2 (Bölte et al., 2017; Righi et al., 2017; Ma et al., 2015; de Tommaso et al., 2008; OliverRodríguez et al., 1999; Carbon et al., 2018; Augustin et al., 2011), P3 (Bölte et al., 2017; Righi et al., 2017; Ma et al., 2015; de Tommaso et al., 2008; OliverRodríguez et al., 1999), and less consistently on P1 (Righi et al., 2017; Bölte et al., 2017) and LPP (Marzi & Viggiano, 2010; Werheid et al., 2007; Schacht et al., 2008). In a previous study we further saw differences in very late potentials (3-5 s after stimulus onset; Strijbosch et al., 2021, though we could not replicate any of the previously listed ERP findings). While the presented EEG findings are not entirely conclusive (both tasks and stimuli differed across studies, regions of interest did not necessarily overlap, and some of the reported effects even go in different directions), they nevertheless suggest that preference related information can be recovered from the noninvasive scalp EEG.

Further physiological data modalities, such as eye tracking and electrocardiography (ECG) are interesting candidates as well: heart rate has been linked to aesthetic valuation (Tschacher et al., 2012), pupil size has been investigated as a potential indicator of *aha* experiences (Salvi et al., 2020) and surprise during music listening (Liao et al., 2018), and gaze patterns have been investigated as a behavior of interest in empirical aesthetics (with potential title-effects on gaze patterns/exploration behavior; Bubi et al., 2017; Kapoula et al., 2009; Temme, 1992; Hristova et al., 2011). All these physiological data types are amendable to the multivariate decoding approach applied in this study.

1.3 The current Study

We designed an extended version of the paradigm by Van de Cruys et al. (2021) and used it in two preregistered experiments: in a first purely behavioral task participants were presented with a piece of visual art and asked to rate how curious they were for the title. Then followed either the artwork's original title or a dummy stating that the piece was "untitled" (real or fake title condition randomly assigned and fully balanced across participants). They then saw the artwork again and were asked to rate their feeling of *aha*, as well as aesthetic liking, and how much they felt they understood the piece. This allows us to see if previous findings on *aha* and meaning making replicate in a new stimulus category (art), but importantly also to functionally link the different evaluation processes that to date have usually been studied separately. We compiled a diverse set of artwork stimuli with matching titles: the stimuli varied by grade of visual abstraction of the artwork (art style: figurative, semi-abstract, or abstract) and information content of the titles (title style: purely descriptive titles that did not add new information, explanatory titles that helped understand the depicted content, or evocative titles that create a link to new, remote concepts). Ideally, this should allow for a finer grained investigation of the effects and interactions of these factors on meaning making and preference. In addition to the artwork stimuli we also included several Mooney images to compare the strength of *aha* elicited by visual art and their titles to the feeling of *aha* caused by perceptual shifts. In particular, we preregistered the following hypotheses: stronger *aha* experience under additional information condition (art with real title or Mooney image) than in no additional information condition ("untitled" art); curiosity ratings predict strength of *aha* experience; *aha* ratings predict aesthetic preference ratings; art with their real title receives higher aesthetic ratings than untitled art stimuli; art with their real title receives higher understanding ratings than untitled art stimuli; an interaction with the degree of abstraction, at least for the understanding rating.

In a second experiment we aimed to investigate the physiological correlates and temporal sequence of reward processing and therefore recorded the participants' EEG, pupillometry, and heart rate in addition to the behavioral ratings. For both, insight and aesthetic preference, several partly overlapping physiological markers have been proposed (see below), but the findings are not conclusive and have to date not been investigated together. For the present study, the participants' EEG (both in time and frequency-domain), pupil traces, and instantaneous heart rate were analyzed using a time-resolved decoding approach designed to show whether and when rating related information was represented in the respective recordings (following Gwilliams & King, 2020, see methods). This part of the experiment was more exploratory and we only preregistered a few coarse hypotheses based on previous work: insight related activity in the gamma band; positive linear relation between aesthetic preference and gamma band power; a nonlinear relation between alpha power and aesthetic preference.

2 Methods

In two experiments we presented our participants with visual artworks (paintings) followed either by their original titles (as additional semantic cues that may aid interpretation of the visual information) or dummy titles without new information ("no title"). The paradigm closely followed the study by Van de Cruys et al. (2021) with Mooney images. While our two experiments differed with respect to the collected data (Exp. 1 purely behavioral, Exp. 2 behavioral and physiological), the stimulus set (Exp. 1 larger stimulus set, Exp. 2 confined subset based on ratings in Exp. 1), and the setting (Exp. 1 online, Exp. 2 lab based) the applied experimental paradigm was largely identical.

The experimental design and our hypotheses were preregistered (see https://doi.org/10.17605/OSF.IO/PAGVH for experiment 1 and https://doi.org/10.17605/OSF.IO/6KW9N for experiment 2). Any detail of the final methods that differed from the preregistered plan is clearly described and discussed as such. Information regarding participants and recording devices are reported following the standards proposed for M/EEG studies by the Committee on Best Practice in Data Analysis and Sharing (COBIDAS Pernet et al., 2018; Pernet et al., 2020) of the Organization for Human Brain Mapping.

2.1 Experiment 1 – Behavioral online study

Participants

Sample size estimation – Target sample size was estimated via *a priori* power analysis using G*Power software (Faul et al., 2007, version 3.1). As power analysis for multilevel linear models (see Statistical Analysis) is not straight forward, we powered our experiment for a similar repeated measures ANOVA. Calculation was done using "ANOVA: Repeated measures within factors" setting with alpha = .05, power = .8, Effect size $f = 0,1758631$ (equivalent to $\eta_p^2 = 0.03$, a small effect size), 1 group, 26 measurements (3x4x2 + 2 design, see below), and conservative expectation of 0 correlation among repeated measures), resulting in a target sample size of 30. As the study was conducted online and we could not control the environment of the data collection we aimed for a substantially larger sample of N = 50 participants. Note though, that we expected the data quality to be higher than usually assumed for online studies. In particular, we did not expect high data loss due to non-committal participants as the participant pool we recruited from consisted of motivated, often regular, participants in studies at our institute.

Recruitment – Participants were convenience-sampled from the general public in the Rhine-Main metropolitan region in Germany, via direct mailings to members of an institute hosted participant database (> 2000 members at time of recruitment, open to everybody to subscribe). Slots were assigned on a first come, first serve basis. There were no specific inclusion criteria for participation in this study.

Final sample – N = 56 participants enrolled in the study and N = 54 finished data collection. The 2 incomplete datasets (13/56 trials, and 53/56 trials) were still included in the analysis. From these 56, 2 participants reported past or current neurological problems (Stroke, MS) under active medical treatment. 5 other participants reported current or past episodes of mental or psychological disorders. 2 of these 5 indicated current medical treatment. The participants were between 20–82 years old (mean = 40.0 years, std = 17.0 years; with 1 participant not indicating their age). 37 of them were female, 19 male, and 0 indicated 'other'. 3 of them were left handed, 49 right handed, and 3 ambidextrous (based on self report, with 1 participant not providing these data). Participants had received between 7–26 years of education (mean = 17.4 years, std = 3.9 years; with 3 participants not providing these data). The sample showed a strong bias towards highly educated people: 54 out of 56 held the German Abitur or a university degree as their highest qualification, and 1 participant did not provide these data. See Supp. Tabs. A.1 and A.2 for full details on participant demographics.

Participants spent on average 50:12 minutes to finish the online data collection (std = 23:35, min: 31:23, 25%: 36:56, 50%: 42:24, 75%: 50:27, max: 2:22:35). For 5 participants, this calculation wasn't possible due to missing timestamps. All participants received monetary compensation of 10€ for finishing the experiment and gave their informed consent prior to participation. The study adhered to the ethical standards of the Declaration of Helsinki and was approved by the local ethics committee (Ethics Council of the Max Planck Society) under the Research Ethics Framework Application for Standard Laboratory Studies. Data were collected from December 2021 through January 2022.

Stimuli

Each participant saw a subset of 56 art stimuli and 8 Mooney images from a stimulus set of 120 art stimuli and 20 Mooney images.

Art stimuli – all art stimuli were taken from the WikiArt.org database containing thousands of images in the public domain. The set of 120 images used in this study was based on WikiArt metadata and the WikiArt Emotions dataset (Mohammad & Kiritchenko, 2018) by the following procedure: 1) we first selected all paintings from 8 western art styles of the last 2 centuries (i.e. romanticism, realism, impressionism, expressionism, cubism, surrealism, abstract expressionism, lyrical abstraction - according to WikiArt metadata), 2) we then excluded all that contained faces (according to WikiArt Emotions ratings), 3) the remaining artworks were categorized into "abstract", "semi-abstract" or "figurative" 4) and by their title into "purely descriptive" title, "explanatory" title and "evocative" title (both by the experimenter) 5) finally we selected 10 stimuli for each of the 9 factor combinations aiming for a) including stimuli from many possible art styles and b) high "art: surprise" rating from the WikiArt emotions dataset. In the 3 "explanatory title" subgroups we extracted 20 stimuli instead of 10 per bin, technically leading to 3 additional groups with explanatory titles (with slightly lower prerated "art: surprise"). This procedure yielded a total of 120 art stimuli, each accompanied by its original title (as stored in the WikiArt database, translated to German).

Mooney images – all Mooney images presented in our study were taken from the original Van de Cruys et al. 2021 study (stimuli publicly available online). Stimuli used in our study were selected by two criteria: 1) not depicting humans, and 2) the average *aha* rating in Van de Cruys et al. 2021. We took the top 10 *aha* rated stimuli and the lowest 10 *aha* rated stimuli from this study. Each Mooney image is accompanied by a matching gray-scale picture from which it was derived (presented as a "solution" cue analogous to the real titles for the art stimuli).

Experimental design

The study was laid out in a within participant factorial design with repeated measures and 5 different participant groups based on stimulus subsets (see below).

The design can be formalized as a 3x4x2 + 2 design: participants saw visual art stimuli in 3 degrees of abstraction (figurative, semi-abstract, abstract), and 4 title types (descriptive title, evocative title, 2x explanatory title), in either of 2 title reveal conditions (present the real title, or present a "no title" dummy title). These factors were fully crossed, with 2 trials for each of the 24 factor combinations (48 art stimuli). Additionally 2 groups of Mooney trials were presented (high *aha* rating, or low *aha* rating in Van de Cruys et al., 2021), with 4 trials per participant each (8 Mooney stimuli). This leads to a total 56 trials per participant (see below for more details on the stimuli).

Task procedure – Participants received an email invitation with a personalized link to an online experiment implemented in Labvanced (Scicovery GmbH, Germany). After following this link and consenting to participate, they first answered a set of ancillary questionnaires (see below). Instructions for the main experiment were displayed, followed by 5 practice trials. The practice trials presented 2 Mooney images and 3 art stimuli together with their real titles (these stimuli were not presented again in the main experiment). Then they performed the experiment, and a short debriefing questionnaire. After finishing participants were redirected to a separate questionnaire on a GDPR compliant Limesurvey server (LimeSurvey GmbH, Germany) to provide payment details for the monetary compensation.

For the main experiment we presented the participants with visual artworks (paintings) followed either by their original titles or dummy titles with no additional information ("no title"). The paradigm closely followed the experiment presented by Van de Cruys et al. (2021) with Mooney images.

For each trial participants went through the following sequence: 1) a visual stimulus was presented (5s presentation time; art or Mooney image, see below), 2) they rated how curious they were for the title or the Mooney solution (self paced response, using a slider on a continuous scale), 3) an additional semantic cue followed (3s presentation time); in case of artwork stimuli this was either the original title or a "no title" dummy (condition randomly selected, fully balanced across participants), in case of Mooney images this was the grayscale photograph the image was derived from, and 4) the initial image was presented again (5s presentation time). They then rated 5) to what extent they experienced *aha*, 6) the aesthetic appeal of the image, and 7) their perceived understanding of the image (all self paced

responses, using a slider on a continuous scale) The sequence is visualized in Fig. 1a. In this experiment the last two rating sliders (aesthetic preference and felt understanding) were presented above each other on the same page.

Randomization – The assignment of each artwork stimulus to the title or "no title" dummy condition was randomized and fully balanced across participants. The stimulus order for each participant was randomized with fully intermixed trials (not more than 1 stimulus of the same factor combination in a row). Participants were randomly assigned to either of 5 groups - the stimulus subsets in these 5 groups differed but fully overlapped with each other (i.e. each individual stimulus was presented to 2 of the 5 groups).

Behavioural measures and questionnaires

First, participants answered a set of questionnaires: the big-five personality inventory (BFI-2-XS: Rammstedt et al., 2020; Soto & John, 2017), the short boredom proneness scale (sBPS: Struk et al., 2017), the short form positive negative affect schedule (PANAS-SF: Thompson, 2007), and the Aesthetic Responsiveness Assessment (AReA: Schlotz et al., 2020) in their German versions, as well as a background questionnaire on basic demographic information (including age, gender, education, and information on mood or neurological disorders).

Data analysis

Statistical analysis and data visualization was done in Python (version 3.7.11) and R (version 4.1.3) in virtual environments on a Macintosh PC (Apple Inc). Jupyter notebooks reproducing figures and results are publicly available online at the OSF.

Linear mixed models – The main hypotheses were tested using linear mixed effects regressions with the *lmer* function from R's *lme4* package (Bates et al., 2015, version 1.1-30). Significance values for individual effect coefficients were estimated using Satterthwaite's degrees-of-freedom method implemented in R's *lmertest* package (Kuznetsova et al., 2017, version 3.1-3). To assess the overall significance of the reported linear models, Likelihood Ratio Tests were performed: LMM-coefficients were compared to reduced intercept-only models (no fixed effect, but random effects of participant and stimuli) using R's *anova* function.

To assess the title-effect the following LMM formulas were applied:

$$understanding \sim titlecondition * artstyle + (titlecondition|participant) + (titlecondition|picture) \quad (1)$$

$$liking \sim titlecondition * artstyle + (titlecondition|participant) + (titlecondition|picture) \quad (2)$$

$$aha \sim titlecondition * artstyle + (titlecondition|participant) + (titlecondition|picture) \quad (3)$$

with their respective intercept-only models

$$rating \sim 1 + (1|participant) + (1|picture) \quad (4)$$

To benchmark the strength of *aha* elicited by visual art and their titles to the feeling of *aha* caused by the Mooney images we fitted the following LMM:

$$aha \sim artstyle : titlestyle + (artstyle + titlestyle|participant) + (1|picture) \quad (5)$$

Here only data from artwork shown with their real titles were used (omitting trials labeled as "untitled"). We applied *simple coding* to compare the average *aha* rating in all interaction cells to the grand average.

To test the linear relation between curiosity and *aha* within the different title styles we applied the following LMM formula for main and intercept only model:

$$aha \sim curiosity : titlestyle + (curiosity|participant) + (1|picture) \tag{6}$$

$$aha \sim 1 + (1|participant) + (1|picture) \tag{7}$$

Local regression – Local regression plots for the figures were generated using locally weighted scatterplot smoothing (LOWESS) as implemented in the *seaborn* module for Python (Waskom, 2021, version 0.11.2).

Data visualization

Data visualizations were generated with custom code using the *seaborn* Python module (Waskom, 2021, version 0.11.2) with custom post-processing using Inkscape (https://inkscape.org).

2.2 Experiment 2 – EEG study

Participants

Sample size estimation – Given the lack of previous EEG research on the title effect or insight with art, we decided to power the study based on the behavioral effect we aimed to observe: a main effect of the stimulus category on the *aha* ratings.

We estimated the target sample size using G*Power (Faul et al., 2007, version 3.1). Again, we powered our experiment for a repeated measures ANOVA design similar to the LMMs used during actual analysis. *A priori* power analysis for "ANOVA: Repeated measures within factors" with $alpha = .05$, $power = .8$, Effect size $f = 0,1758631$ (equivalent to $\eta_p^2 = 0.03$, a small effect size), 1 group, 7 measurements (3x2 + 1 design, see below), and correlation among repeated measures $r = .42$ resulted in a target sample size of N = 38. As we expected some dropout, we aimed to collect data from at least N = 45 participants.

The correlation between repeated measures was estimated based on the online study data, as the average correlation coefficient of the participant-average insight-ratings across the 7 stimulus factor combinations used in the present study.

Recruitment – Participants were convenience-sampled from the general public in the Rhine-Main metropolitan region in Germany. Recruitment was performed via direct mailings to members of an institute hosted participant database (> 2000 members at time of recruitment, open to everybody to subscribe). Slots were assigned on a first come, first serve basis. Inclusion criteria for recruitment were age between 18–55 years, good command of the German language, good eyesight and no need to wear glasses during the study (as this decreases the quality of the eye tracking), no known neurological disorders, and no participation in the previous online study.

Final sample – 50 participants enrolled and finished the data collection. Despite the exclusion criteria made clear in the recruitment, 1 participant reported ongoing neurological problems (essential tremor), and was thus excluded from analysis, leaving 49 datasets for analysis. A few participants had periods of missing data (EEG, ECG, eye tracking, and/or ratings) due to minor bugs or poor quality recordings, but were not generally excluded from analysis.

The participants were between 18–52 years old (mean = 28.0 years, std = 7.2 years). 28 of them were female, 19 male, and 2 indicated 'other'. 8 of the participants were left handed, 41 right handed, and 0 ambidextrous (Flinders Handedness survey; Nicholls et al., 2013). Eye dominance was assessed by the experimenter (see below): 17 participants showed left eye, 28 right eye, and 4 no eye dominance. Participants indicated having received between 8–27 years of education (mean = 17.3 years, std = 3.9 years; with 1 participant not providing these data). The sample showed an

extreme bias towards highly educated people: all 49 participants indicated the German Abitur or a university degree as their highest qualification. 11 in 49 participants reported current or past episodes of mental or psychological disorders with 7 of these 11 indicating current medication. Although it was not a criterion for exclusion, we also assessed the participants caffeine intake on the day of the experiment: it ranged between 0–3.61 mg/kg body weight (mean = 1.06 mg/kg, std = 0.923 mg/kg). 8 out of 49 participants indicated zero caffeine intake before participation. See Supp. Tabs. A.1 and A.2 for full details on participant demographics.

All participants received monetary compensation of 7€ per 30 minutes and gave their informed written consent prior to participation. The study adhered to the ethical standards of the Declaration of Helsinki and was approved by the local ethics committee (Ethics Council of the Max Planck Society) under the Research Ethics Framework Application for Standard Laboratory Studies. Data were collected from March 2022 through July 2022.

Stimuli

Each participant saw 48 art stimuli and 16 Mooney images on a calibrated LCD monitor. These stimuli were preselected from the larger stimulus set of 120 art stimuli and 20 Mooney images from Exp. 1 using the rating data from Exp. 1.

Image selection – From the 140 stimuli described above for Exp. 1 we selected 64 based on a linear mixed model with the formula:

$$aha \sim titleshown + (titleshown|participant) + (titleshown|picture) \tag{8}$$

i.e. a model that predicts the insight rating from the title condition (real title or no title), with random slopes and intercepts for each participant and each stimulus (meaning that the strength of the title effect can vary for each participant and each stimulus). The slope coefficients for all stimuli were extracted from the fitted model and ranked. For this study we chose the 16 stimuli with the highest slope within each picture category (abstract art, semi-abstract art, figurative art, Mooney images) – these were the images with the strongest title-effect on *aha* ratings. The third fully crossed condition from Exp. 1 (type of title: descriptive, explanatory, evocative) was omitted here, as this would have drastically decreased the freedom in the selection process. The ranked slope coefficients and the selection are depicted in Supp. Fig. A.1. A full list of stimuli used in this study can be found in Supp. A.8.

Experimental design

The study was laid out in a within participant factorial design with repeated measures. The design can be formalized as a 3x2 + 1 design: participants saw visual art stimuli in 3 degrees of abstraction (figurative, semi-abstract, abstract), in either of 2 title reveal conditions (present the real title, or present a "no title" dummy title). These factors were fully crossed, with 8 trials for each of the 6 factor combinations (48 art stimuli). Additionally, 16 trials with Mooney images were presented per participant. This led to a total 64 trials per participant. As opposed to Exp. 1, all participants saw the same stimuli though with randomly drawn title assignments.

Task procedure – First the participant answered a set of ancillary questionnaires (see below) digitally implemented in Labvanced (Scicovery GmbH, Germany), while the experimenter prepared the EEG cap. After that the participant entered the cabin, the EEG cap was mounted, and peripheral electrodes (ECG) were attached. EEG electrodes were filled with electrolyte gel and electrodes were adjusted until impedance were below or close to 10 kOhm. After finishing preparations, the effects of chewing, blinking, and contracting neck muscles on the EEG signal were demonstrated to the participant to raise awareness for these sources of artifacts, along with a demonstration of individual alpha oscillations during a short eyes closed period. The experimenter left the cabin, and instructions for the main experiment were displayed, followed by 3 practice trials. The practice trials presented 1 Mooney image and 2 art stimuli together

with their real titles. These stimuli were not presented again in the main experiment. After the end of the training the experimenter entered the cabin, adjusted the participants sitting position and the head rest, and set up the eye tracker. From then on, the participant was asked to keep her position in the head rest and restrain from chewing, swallowing or moving the head during trials. The main experiment then started with a calibration of the eye tracker (horizontal and vertical calibration at 13 positions) followed by the first trials. Trials were presented in blocks of 10 stimuli, and after each block participants could choose to take a short self paced break to relax. After finishing the experiment electrode positions were digitized, and participants filled a short debriefing questionnaire digitally implemented in Labvanced (Scicovery GmbH, Germany).

The applied experimental paradigm was largely identical to the one described above for Exp. 1 (see also Fig. 1a). Yet unlike the online study, all pictures and titles were preceded by a fixation cross, presented for $2-2.5s$ (jittered), the inter stimulus interval was $1-1.5s$ (jittered), and all rating sliders (including aesthetic preference and felt understanding) were presented on separate pages. The routine was implemented in custom code using PsychoPy (Peirce, 2007).

Randomization – The assignment of each art stimulus to the title or "no title" dummy condition was fully randomized across participants. However, the randomization was controlled such that within each art style (abstract, semi-abstract, figurative) the final number of title and no-title trials was balanced. To keep the assignment distribution balanced for the whole data collection, a random assignment was done only for every other participant (1,3,5, and so on), while the subsequent participants (2,4,6, and so on) got the assignments of the previous participant flipped around. The stimulus order for each participant was randomized, with fully intermixed trials (not more than 2 stimuli of the same type in a row).

Behavioural measures and questionnaires

In addition to the screening tools used in Exp. 1, caffeine intake before the visit (Bühler et al., 2014), and handedness (FLANDERS; Nicholls et al., 2013) were assessed via systematic questionnaires. Ocular dominance was determined by the experimenter using a variation of the Porta test (as described in Roth, 2002).

All trial-wise ratings (curiosity, *aha*, aesthetic liking, understanding) were collected using a continuous scale with a slider controlled by moving the mouse. A response was logged by clicking the mouse button, but participants could not log a response without first moving the mouse in order to prevent non-responding. Both the ratings and response times were logged. The ends of the scales were marked with "+" and "–" and each scale was accompanied by the corresponding questions: "Wie neugierig sind Sie auf den Titel / die Auflösung?" ("How curious are you for the title / the solution?") for the curiosity rating, "Hatten Sie ein Aha! Erlebnis?" ("Did you have an Aha! experience?") for the *aha* rating, "Wie ästhetisch ansprechend fanden Sie das Bild?" ("How aesthetically appealing did you find the picture?") for the aesthetic rating, or "Denken Sie, dass Sie das Bild verstanden haben?" ("Do you think you understood the picture?") for the understanding rating respectively. The order of the ratings was fixed. In the more detailed task description preceding the study, all ratings were introduced – the full texts can be found in the Appendix.

Data acquisition and devices

Study environment – EEG preparation and main experimental routine took place in an acoustically shielded cabin (model: IAC 120a, IAC GmbH, Germany; internal dimension: $2.74 * 2.54 * 2.3m$). Participants were seated in a chair and placed their chin on a chin rest with forehead support (SR Research Head Support, SR Research Ltd., Canada). The distance between the chin rest and the screen was 72 cm. The height of the desk was adjusted such that participants sat in a comfortable upright position that they were able to sustain for the time of the study. All data was collected by the same experimenter, with assistance during EEG gel preparation. The participant could contact the experimenter any time via a room microphone installed in the cabin. The experiment was run on a PC running 64-bit Microsoft Windows 7.1.7601 service pack 1 (Microsoft Corporation, USA), using PsychoPy3 standalone software (Peirce, 2007, version 3.0.7). Visual stimuli were presented on a 24 inch BenQ XL2420Z screen (BenQ Corporation, Taiwan) with

nominal frame rate of 144 Hz, resolution 1920 x 1080 px (mirrored for the experimenter). Luminance and color calibration was performed using the screen's settings and validated using a chromameter (Konica Minolta CS-150) with a staircase routine from the PsychoPy monitor center; the measured gamma curve can be found online in the OSF repository.

EEG and peripheral physiology data acquisition – Bio-signals were amplified using a BrainVision actiCHamp 128 Amplifier with BIP2Aux Adapters (Brain Products GmbH, Germany) with 24-bit digitization and an analogue band-pass filter from DC - 280 Hz. The amplifier was DC battery powered by 2 ActiPower PowerPacks (Brain Products GmbH, Germany). Data was recorded continuously with 1 kHz sampling frequency using BrainVision Recorder software (Brain Products, version 1.21.0303) on an independent recording PC running Microsoft Windows 7. Triggers from the experimenter PC to the recording system were sent via the parallel port. EEG data were collected using a 64 channel actiCAP system with active Ag/AgCl electrodes without active shielding (Brain Products GmbH, Germany), placed according to extended international 10-20 localization system (Commitee, 1958; Oostenveld & Praamstra, 2001) with FCz online recording reference and AFz ground. Before placing the cap, forehead and skin behind the ears were wiped with alcohol to enhance the impedance. The cap was positioned by centering the Cz electrode on the axes Nasion to Inion and left ear to right ear. After the end of the data collection, electrode positions were digitized using a Polhemus FastTrak system (Polhemus, USA) and Curry software (Compumedics, version 8). ECG data was recorded via a bipolar auxiliary channel of the EEG amplifier: electrodes were placed on the right mid clavicle and lower left rib cage, in correspondence with the II Einthovens derivation, using disposable electrodes (Covidien, USA). A photo diode (Brain Products GmbH, Germany) connected to another auxiliary channel of the amplifier pointed at a dummy stimulus in the lower right corner of the screen to verify parallel-port triggers sent from the presentation computer. In order to enhance EEG data quality the participants were asked in the email invitation to not apply hair styling products before the study.

Eye tracking data acquisition – Binocular eye position (gaze coordinates on the screen) and pupil size were recorded using a desktop mount EyeLink 1000 Plus eye tracking system (SR Research Ltd.) and EyeLink 1000 Plus Host software running on an independent recording PC. Connection between the experimenter PC and the EyeLink recording system was established via ethernet, and the tracker system was controlled using the PyLink Python module (SR Research Ltd., version 1.11.0.0). Epoched recordings were made for each block, with a sampling frequency of 500 Hz. Trigger messages were sent from the presentation PC to the eye tracker (identical messages as for the EEG). The eye tracker was re-calibrated preceding every block using a horizontal and vertical calibration routine at 13 positions. In order to enhance data quality the participants were asked in the email invitation to not apply make-up before the study, especially eyeliner and mascara.

Data analysis

Data preprocessing, processing, and statistical analysis was done on the institute's high performance computer (HPC) using Python (version 3.10.4), R (version 3.6.1) in conda virtual environments, and Matlab (The MathWorks Inc., version R2019b).

EEG data early preprocessing

EEG Data were transferred into BIDS format (Pernet et al., 2019, version 1.6.0) using the MNE-BIDS Python module (Appelhoff et al., 2019, version 0.10). Data were preprocessed using the PREP pipeline (version 0.56.0, using the default parameter set; Bigdely-Shamlo et al., 2015) implemented as an EEGLAB plugin in MATLAB. For details on further processing, see the sections below. Diverging from the preregistered analysis plan, we did not apply a fully-automatic multi-step preprocessing and cleaning pipeline including ICA (e.g., AUTOMAGIC; Pedroni et al., 2019). The less aggressive fully automatic early-stage preprocessing pipeline we used (i.e. PREP; Bigdely-Shamlo et al., 2015) consisted of only line noise removal, reference to (robust) average, and detection and interpolation of noisy channels. This decision was based on 1) recent work suggesting that heavy preprocessing and/or data rejection

can sometimes be less helpful than intended (Delorme, 2023) as any gain in SNR has to outweigh the decrease in statistical power due to data rejection, and 2) work that showed that ICA based data cleaning can leave residuals in the data (Dimigen, 2020; Robbins et al., 2020) and that those can still be problematic if subsequent analyses are based on sensitive machine learning algorithms (Quax et al., 2019; Thielen et al., 2019), as planned in this study.

Time resolved decoding from EEG, pupil size, and instantaneous heart rate

In the main analysis we wanted to predict different targets (stimulus categories and ratings) from the multivariate physiological data (EEG, pupil size, heart rate). The temporal decoding was done in a sliding window approach: independent decoders were fit at each time sample across the 64 EEG sensors (or 2 pupil traces, or single heart rate channel). This analysis yielded a decoding time course indicating if and when a rating or condition can be linearly decoded from the physiological signals. The decoding time courses were then passed on to cluster-corrected second-level statistics across participants (see below).

Decoding pipelines – We built 3 different categorical classification pipelines for repeated presentation (1st/2nd presentation of the image), the stimulus condition (Mooney/art) and the title condition (real or uninformative dummy title) and 4 different continuous prediction pipelines for the different ratings (curiosity, *aha*, liking, understanding; all transformed to participant wise z-scores). The decoder design closely follows Gwilliams and King (2020), implemented in MNE Python: We applied logistic regressions for categorical classification (scikit-learn; Pedregosa et al., 2011, default parameters: C = 1) and ridge regression for continuous prediction (default parameters: alpha = 1). Decoding performance was assessed in a stratified 10-fold cross-validation. Decoding accuracy was scored using an AUC metric for logistic regression, and spearman R for ridge regression. All decoders were provided with data normalized by the mean and the standard deviation in the training set.

Data features – These decoding pipelines were applied to different data: first, to single-trial ERPs on all sixty four channels – continuous PREP preprocessed EEG data were band-pass filtered (0.1-40 Hz, FIR, MNE default parameter) and cut into epochs (-0.5-5 s around stim onset); baseline correction was applied by subtracting the average of the baseline period (-0.5-0 s) from the entire epoch (for each trial and channel individually) yielding single-trial ERPs. Data were re-referenced to infinite reference (REST; Yao, 2001, implemented in MNE). Lastly, a fully automatic trial-rejection procedure was applied (based on the min-max amplitude of the trial data within each participant): a gamma distribution was fit to the empirical distribution of min-max amplitudes of all trials used in the respective decoding subset, and epochs with a value above the .9 percentile of the fitted distribution were rejected as outliers. While this step reduced our trial number, we deemed it necessary, as the scaling step (see below) makes the applied decoding pipeline vulnerable to outliers.

The second dataset for decoding was 64 channel oscillatory activity in the canonical EEG frequency-bands. PREP preprocessed continuous EEG data were band-pass filtered to the respective frequency-band (delta: .1-4 Hz; theta: 4-8 Hz; alpha: 8-23 Hz; beta: 13-30 Hz; gamma: 30-80 Hz – FIR, MNE default parameter) and cut into initially longer than needed epochs (-2–6.5 s for delta and -1–5.5 s for the other bands, to avoid edge artifacts in the analyzed data); no baseline correction was applied at this stage. Analogous to ERPs, the data were re-referenced to infinite reference. To reveal oscillatory activity the evoked response was then subtracted from every single trial and the envelope of the Hilbert transformed EEG was calculated as the time-frequency response (TFR). Lastly, epochs were cropped to the time window intended for decoding (-0.5–5 s) and baseline correction was applied subtracting the average of the baseline period (-0.5–0 s) from the entire epoch (for each trial and channel individually). The same fully automatic trial-rejection as for ERP data was performed.

The third dataset was the binocular pupil trace. Raw continuous data were low-pass filtered (cutoff 10 Hz, FIR, MNE default parameter), missing data due to eye blinks were linearly interpolated (using MNE's *interpolate_blinks* function), and trials with a min to max amplitude two standard deviations larger than the average were rejected as outliers.

The fourth dataset was instantaneous heart rate. ECG data recorded alongside EEG was loaded using MNE-python and preprocessed using the neurokit2 python module (version 0.2.7; Makowski et al., 2021): recordings were checked for inverted signal using neurokit2's *ecg_invert* function and corrected if necessary, cut into initially longer than needed epochs (-3.5–8 s; no baseline correction) followed by data cleaning (neurokit's *ecg_clean* function), R-peak detection (neurokit's *ecg_peak* function), and conversion to instantaneous heart rate (neurokit's *ecg_rate* function, all with default parameters). Lastly, epochs were cropped to the time window intended for decoding (-0.5–5 s). Recording quality in each trial was assessed using neurokit's *ecg_quality* function (using the scoring method following Zhao & Zhang, 2018) and all trials not scored as "excellent" were excluded from analysis.

ERP, TFR, pupil, and heart rate data were then down-sampled to 50 Hz prior to decoding (using MNE's *resample* function with default parameters).

Trial rejection – Below we report trial rejection across all the different decoding pipelines within a given data modality. For decoding of repeated exposure (1st or 2nd presentation) and stimulus category (Mooney or visual art) trials with all 64 stimuli were used, all other analyses were performed only on visual art trials (48 trials per participant). The criteria for rejecting outlier trials in the different data modalities is described above. In addition, a participant was skipped entirely within a given decoding pipeline if 1) there were fewer trials left than required for cross validation, or 2) if the variance in a participant's ratings was to low for the decoding model to converge.

The cleaning and trial rejection procedures described above led to a moderate trial rejection in the low percentage range. Note that the maximum trial number varied by pipeline. ERP – data from 46–49 participants (mean=48.2 out of 49, or 98.45%) and within those 91.3–95.6% (mean=93.51%) of the maximum possible trials were used. EEG delta – data from 47–49 participants (mean=48.3 out of 49, or 98.64%) and within those 91.0–95.3% (mean=92.85%) of the maximum possible trials were used. EEG theta – data from 41–49 participants (mean=47.9 out of 49, or 97.71%) and within those 87.8–93.3% (mean=91.06%) of the maximum possible trials were used. EEG alpha – data from 45–49 participants (mean=48.3 out of 49, or 98.64%) and within those 88.7–94.5% (mean=91.46%) of the maximum possible trials were used. EEG beta – data from 47–49 participants (mean=48.5 out of 49, or 98.89%) and within those 93.3–96.6% (mean=95.01%) of the maximum possible trials were used. EEG gamma – data from 47–49 participants (mean=48.4 out of 49, or 98.70%) and within those 91.8–96.6% (mean=94.29%) of the maximum possible trials were used. Pupil dilation – data from 47–49 participants (mean=48.3 out of 49, or 98.64%) and within those 98.9–99.4% (mean=99.07%) of the maximum possible trials were used. Heart rate – data from 45–49 participants (mean=47.0 out of 49, or 95.98%) and within those 95.8–98.3% (mean=97.12%) of the maximum possible trials were used.

Second-level statistical analysis of decoding accuracy – We evaluated the significance of the uni-variate and multivariate decoding results using a one-sample permutation cluster test (Maris & Oostenveld, 2007) implemented in MNE's *permutation_cluster_1samp_test* function.

To account for noise in the individual decoding results, decoding timecourses for each participant were mildly smoothed prior to the permutation test with a Gaussian filter (using scipy's *gaussian_filter* function, with sigma=1.0).

We centered the data around the theoretical chance level of the respective metric (0.5 for AUC, 0 for Spearman R). Then the permutation cluster test was performed: first, one-sided one-sample t-tests were done at each location in time of the original data. Spatio-temporally adjacent data-points were clustered based on a cluster-forming threshold equivalent to $p < 0.05$ and the test statistic for each cluster was found as the sum of t-values across time. Then a permutation null distribution was generated (in 10000 permutation steps): randomized data were created from the original data with random sign flips, and a new set of clusters and test statistics were formed. The original test statistics were then compared to the permutation null distribution to assess statistical significance.

Data visualization

Data visualizations for the time-resolved decoding analysis were generated with custom code using matplotlib Python module (Hunter, 2007, version 3.8.0), exported in .svg format and subject to custom post-processing using Inkscape (https://inkscape.org).

3 Results

3.1 Experiment 1 – Behavioral ratings of curiosity, insight, aesthetic preference, and understanding

To test for the presence of a "title effect" on the various behavioral ratings we fit a set of linear mixed models (LMMs) to our data from the behavioral online study (Eqs. 1, 2,3). For this analysis, only the ratings of art stimuli were used, ignoring the Mooney images. For comparison however, the figures include ratings for the Mooney images as well.

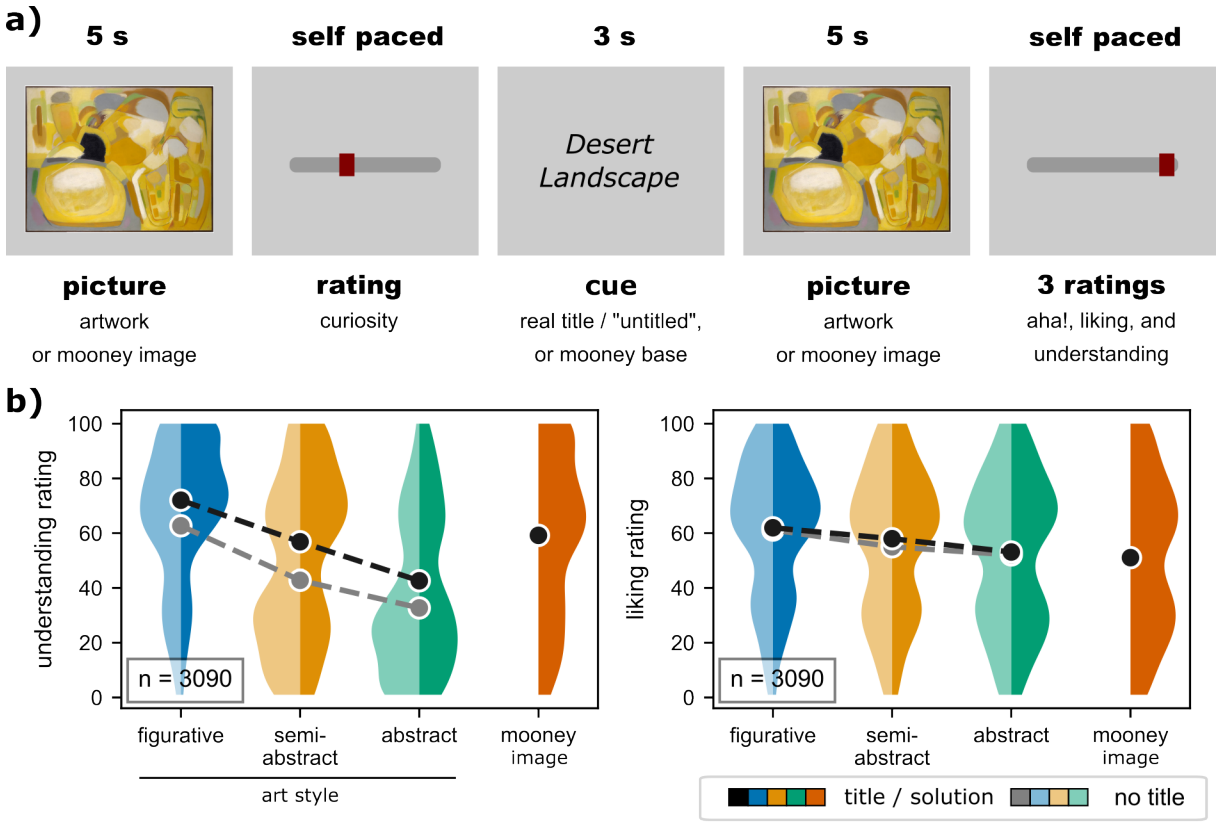


Figure 1: **Artworks with their genuine titles evoke higher feeling of understanding than "untitled" art (replicating previous findings). a) Experimental sequence:** Participants viewed images of visual art and were asked to rate their curiosity for its title. Then followed artwork title – either the original title or an uninformative dummy ("untitled"). Then they saw the original image again, and were asked to rate their feeling of *aha*, aesthetic liking, and if they felt they understood the piece. Besides visual artworks, a series of Mooney images (Mooney & Ferguson, 1951) was presented (randomly intermixed). Instead of a real or fake title, the base photograph used to create the Mooney image was shown. **b) Replicating a title-effect on understanding, but not on aesthetic liking:** violin plots show the distribution of all ratings (n=3090 across 56 participants) separately for the stimulus categories (visual art in 3 levels of figurativeness plus Mooney images), split by title condition (real or no title; all Mooney images were accompanied by their real base photography and here categorized as titled). We reproduce a strong effect of the title manipulation on felt understanding, present in all 3 categories of artwork, but no significant differences for aesthetic liking.

Informative titles increase Aha and felt understanding, but not aesthetic appeal

We were able to replicate previous findings of a title effect for perceived understanding, but not for aesthetic ratings. A visualization of the data (splitting the artwork stimuli by grade of abstraction) is shown in Fig. 1b. The title effect on felt understanding was substantial ($slope = 11.1$, $SE = 1.51$, $p = 6.77 * 10^{-10}$ using Satterthwaite's degrees-of-freedom method) and the likelihood ratio test was significant ($\chi^2[5] = 173.88$, $p < 2.2 * 10^{-16}$). The title effect on aesthetic rating, however, was only marginal ($slope = 1.7$, $SE = 0.94$, $p = .079$ using Satterthwaite's degrees-of-freedom method) and the likelihood ratio test was not significant ($\chi^2[5] = 6.53$, $p = .258$). We did not see any hint that this would be different in any of the art or title styles.

Besides these previously investigated effects on aesthetic preference and felt understanding we could confirm our hypothesis that the title manipulation had also an effect on the *aha* ratings (see Fig. 2a). The title effect on *aha* was even stronger than on understanding ratings ($slope = 17.9$, $SE = 2.08$, $p = 1.53 * 10^{-12}$ using Satterthwaite's degrees-of-freedom method) and the likelihood ratio test was significant ($\chi^2[5] = 564.95$, $p < 2.2 * 10^{-16}$). Importantly, the effect of the title manipulation on *aha* ratings varied across stimuli, with a few of them even receiving lower aha ratings with the actual title than with the dummy titles (see Supp. A.2 and Supp. Fig. A.1).

To benchmark the strength of *aha* elicited by visual art and their titles against the feeling of *aha* caused by perceptual reveals of the Mooney images we fit an additional LMM (Eq. 5) on *aha* ratings from all trials with titled art and Mooney images (i.e. omitting trials with art stimuli but in the no title condition). This model design compared the average *aha* rating in these interaction cells to the grand average. Not surprisingly we observed the highest insight ratings for Mooney images ($slope = 19.1$, $SE = 4.60$, $p = 6.82 * 10^{-5}$) but semi-abstract art with explanatory titles also caused significantly stronger *aha* than the average ($slope = 11.3$, $SE = 4.77$, $p = .01978$). On the other hand, all art styles in combination with descriptive titles resulted in significantly lower average *aha* (figurative art with descriptive titles: $slope = -13.4$, $SE = 4.73$, $p = .00542$, semi-abstract art with descriptive titles: $slope = -10.4$, $SE = 4.95$, $p = .03869$, and abstract art with descriptive titles: $slope = -11.1$, $SE = 4.86$, $p = .02474$). The other groups did not significantly differ from the grand mean.

Type and amount of gained information shapes the interrelation of ratings

Using local regression we found that *aha* ratings were predictive of both aesthetic preference and understanding ratings (see Fig. 2b and c). This relation was relatively linear for aesthetic preference, and did not depend much on the picture type (see LOWESS fits in Fig. 2c). We thereby replicate previous observations with Mooney images (Muth & Carbon, 2013) in visual art. For understanding ratings however, we saw evidence for a nonlinear interaction with the stimulus category: while the association was strong and linear for abstract art, it seemed more quadratic (U-shaped) for semi-abstract and figurative art. Mooney images also exhibited a U-shaped association (see LOWESS fits in Fig. 2b). This pattern would be expected if these low aha–high understanding trials were due to a high initial understanding of the images that leaves little uncertainty to reduce by the additional information of title or Mooney solution (see also Supp. Fig. A.3).

By means of another LMM (Eq. 6; likelihood ratio test: $\chi^2[5] = 773.51$, $p < 2.2 * 10^{-16}$) we also found a very robust linear relation between previously rated curiosity and *aha* ratings in the Mooney images ($slope = 0.635$, $SE = 0.0348$, $p < 2 * 10^{-16}$ – meaning that incrementing the curiosity rating by 1 point resulted in .635 point higher *aha* rating). This matches previously reported values for this relation of $r = 0.68$ (on the image level, see Van de Cruys et al., 2021). The LMM also found a significant effect with smaller slope for titled art ($slope = 0.370$, $SE = 0.0314$, $p = 4.91 * 10^{-15}$). These findings might suggest that curiosity generally predicted *aha* – yet, for untitled art the slope was much lower ($slope = 0.111$, $SE = 0.0313$, $p = .001$). Again, local regression (LOWESS, see Fig. 2d) revealed a non-linear interaction with the type of title cue: while Mooney images and art with explanatory titles (i.e., titles leaving little uncertainty what is depicted in the artwork) exhibited a relatively linear relationship between curiosity and *aha*, this association was weaker and markedly nonlinear for the titles that reduced uncertainty to a lower degree or

not at all (e.g., descriptive titles that solely described in words what was depicted on the image and "untitled" dummy titles that added no new information at all). Here, for low to medium-high curiosity trials the LOWESS fit of *aha* ratings monotonically increased, but then flattened and even fell off again for high curiosity trials. This is consistent with a disappointment effect for trials with high curiosity but no subsequent relief (i.e. without actual information gain through the title). Lastly evocative titles, which introduce and link to a new concept not readily depicted in the visual

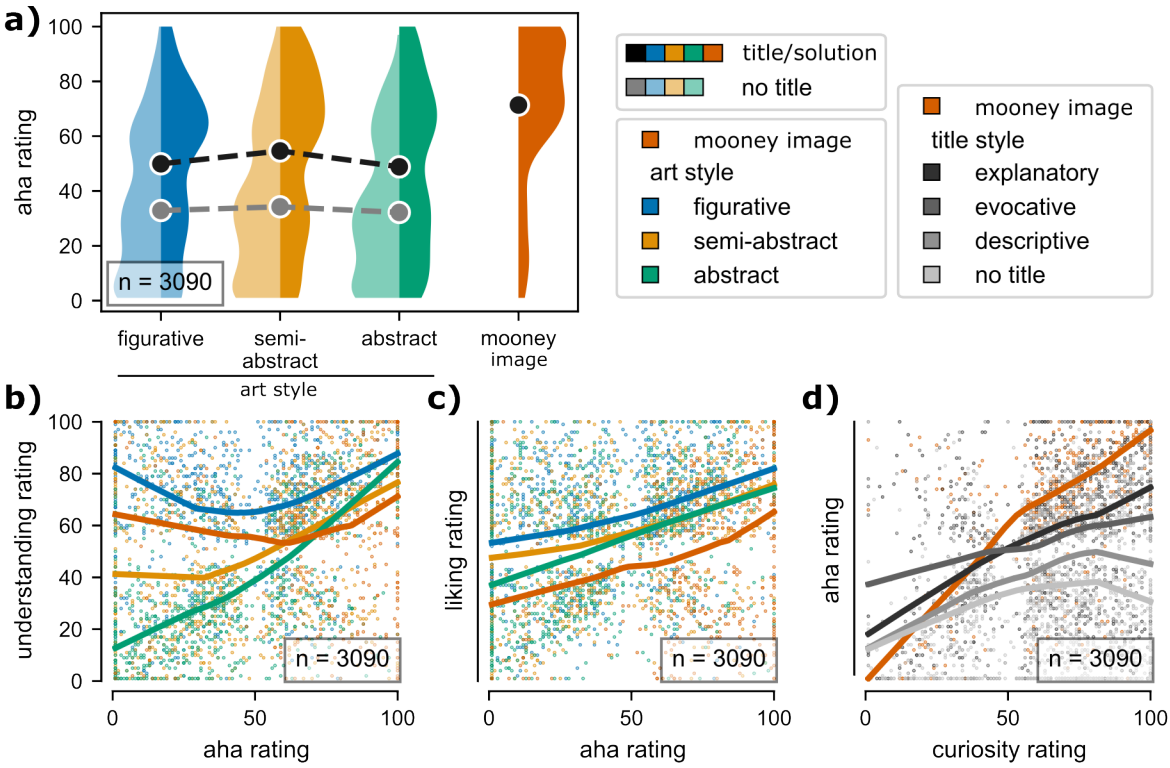


Figure 2: **Using the title effect to evoke insight with visual art. a) Establishing a title effect on felt *aha*:** the title manipulation strongly affected the participants' feeling of *aha* (present in all 3 categories of artwork); the effect is even larger than on felt understanding. **b) Relation between *aha* and understanding interacts with stimulus category:** for abstract art (green LOWESS fit) there is strong linear relation between feelings of *aha* and understanding. On the other hand, figurative art (blue LOWESS fit) exhibits a clear non-linear U-shaped relation with high understanding both for very low and very high *aha* trials (consistent with the hypothesis, that low *aha* could result from trials in which felt understanding was high from the beginning, and no uncertainty was to be reduced). Semi-abstract art (yellow LOWESS fit) exhibits an intermediate pattern. The Mooney images (orange LOWESS fit) also show a U-shaped relation, but on average lower understanding ratings than figurative art. **c) Positive relation between *aha* and aesthetic liking:** the association is present in all stimulus categories and appears linear. As reported before, average aesthetic liking is highest for figurative art (blue LOWESS fit). Mooney images (orange LOWESS fit) cause the lowest average aesthetic liking. **d) Relation between curiosity and insight interacts with information content of solution cues:** we reproduce a strong linear relation between curiosity and *aha* for Mooney images (orange LOWESS fit; previously reported by Van de Cruys et al., 2021) and see a similar, slightly weaker relation for artwork with explanatory titles (i.e., titles leaving no uncertainty about what is depicted in the artwork; darkest LOWESS fit). However, for the titles that reduce uncertainty to a lesser degree or not at all (e.g., descriptive titles that solely describe in words what is depicted on the image and "untitled" dummy titles that add no semantic information at all; light gray LOWESS fits) the association becomes weaker and markedly nonlinear. For low to medium-high curiosity trials the LOWESS fit of *aha* ratings is monotonically increasing, but then flattens and even falls off again for high curiosity trials. Evocative titles, which introduce and link to a new concept not readily depicted in the visual image (dark-gray LOWESS fit), exhibit a weaker, but linear relationship, with higher *aha* ratings in low curiosity trials, compared to the other title conditions.

s

image, exhibited a weaker but linear relationship, with higher *aha* ratings in low curiosity trials compared to the other title conditions.

3.2 Experiment 2 – Neuronal correlates and the temporal sequence of information processing

To study the temporal profile and neuronal mechanisms of the epistemic meaning making process we repeated the experiment in the lab, recording observers' EEG, heart rate, and pupil size. These physiological signals were analyzed using a time-resolved decoding approach designed to show whether and when relevant information was represented in the physiological data (see methods). We aimed to decode both the experimental conditions (as a proof of concept) as well as the various behavioral ratings collected during the trials. Importantly, we ran these decoding pipelines on data from both the pretitle trials and the posttitle trials separately, to reveal potential effects of the interim presented title/solution (except in one pipeline that specifically decoded whether trials belonged to the pretitle or posttitle condition). We further ran the pipelines on a difference signal (posttitle – pretitle trials); these results are reported in Supp. A.6. Findings are described below, and relevant decoding results are shown in Figs. 3 and 4. Decoding plots for all data modalities and all conditions are compiled in Supp. A.7.

Robust decoding of presentation, image category and title condition

The categorical experimental manipulations we applied could be decoded quite robustly from the recorded data. A more verbose description of the results can be found in Supp. A.5. Decoding worked very well for the contrast between the 1st and 2nd presentation of any given image stimulus (see Fig. 3a). Decoding was reliably possible from almost all data modalities. This strongly suggests that the stimuli are not processed identically during repeated presentation; rather, cognitive processing differs in a reliably detectable way. Whether this is caused by mere repetition suppression (cf. Kim, 2017) or other factors specific to the study such as the interim contextualization with additional information (i.e. the real or fake title, or the Mooney solution respectively) or the differing implicit or explicit tasks (e.g., generating ratings) can not be clearly determined at this point.

The presented image category (Mooney image or painting) could also be reliably decoded right from the beginning from both EEG and pupil data (see Fig. 3b). Notably, except in the pupil data (see discussion) classification worked more reliably during the second image presentation than the first (i.e. higher decoding performances, and more sustained significance); this was the case for most of the investigated decoding tasks in this study.

Turning to the questions more central to this study, the title condition in artworks (real title or a dummy title) could also be significantly decoded from the time-domain signals (see Fig. 4a). However, none of the EEG frequency-bands could be used to reliably decode the title condition. When decoding from ERPs during second viewing (posttitle), accuracy was the highest in a sustained but relatively early window after image onset (significant from 0.3 s onward with 2 short intermissions). Decoding from pupil size was significant in a time-window between 1.47–2.37 s in the posttitle condition. Heart rate during second viewing predicted the condition significantly in two time-windows from -0.50–0.64 s and 3.31–5.00 s, respectively. As to be expected, the title condition could not be significantly decoded from any of the recorded data modalities during first viewing (when the title or dummy had not yet been presented).

Physiological signatures of subjective processing experience

We now turn to the question of whether the subjective aspects of the perceiver's processing experience can be decoded from the physiological signals. Participants ratings of curiosity, liking and understanding could be significantly decoded from ERP data at some point during the 5 s image presentation, while insight could be robustly decoded from pupil size, and understanding from heart rate (see Fig. 4b). Felt understanding was the one among our candidates that could be best decoded. The time-domain EEG allowed good decoding in the early pretitle trial (significant from 1.21–1.83 s after image onset). In the posttitle trial, decoding rate stays above chance throughout the whole trial, early on with values similar to the pretitle trials (though not significant) but then building up toward the end of the image

presentation and becoming significant in 2 late time-windows (2.39–3.11 s and 3.41–3.80 s after image onset). Delta band power in the late posttitle trial also significantly predicted the ratings (4.20-4.60 s after image onset). Heart rate in posttitle trials showed marked significant decoding rate for understanding in 2 time-windows (-0.50–0.50 s and 2.71–3.88 s after image onset) that are very similar to those for decoding the title condition (though a little earlier; see above). Pupil dilation also exhibited above chance decoding, with pre- and posttitle timecourses very similar to the ERP data, but did not become significant.

Ratings of curiosity for the title could be significantly decoded from the late pretitle pupil signal (4.06-4.66 s after image onset) and the early posttitle time domain EEG (0.93-1.19 s and 1.23-1.59 s after image onset).

Insight (*aha*) ratings could be significantly decoded from pupil size early in posttitle trials (0.62-1.57 s after image onset).

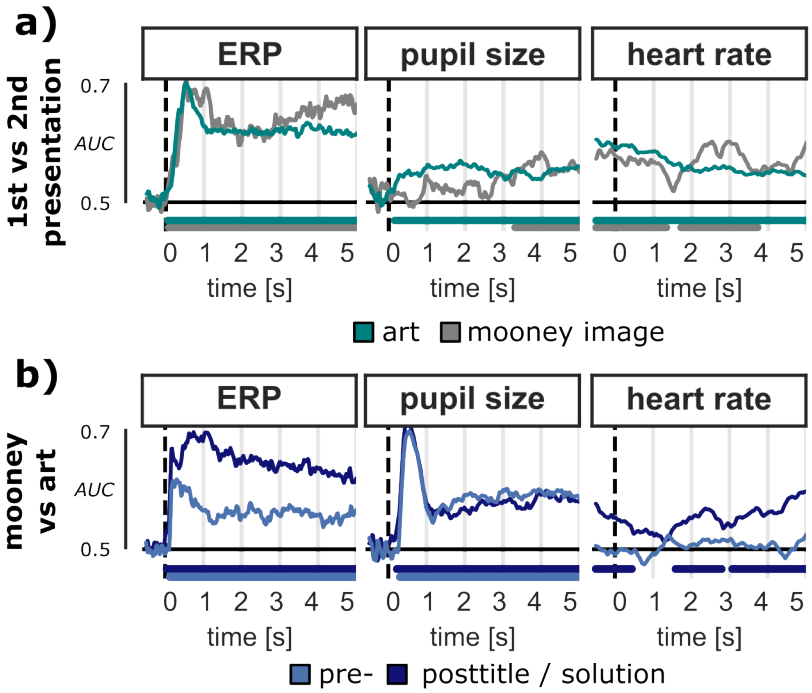


Figure 3: **Decoding repeated presentation and image category (artwork and Mooney trials) – a) 1st or 2nd presentation of the image:** decoding is reliably possible from almost all data modalities but best from the time-domain EEG (significantly above chance throughout the full trial, with a marked peak around 0.5–1 s after image onset). In pupil trace decoding works better with art than with Mooney images, where decoding becomes significant only at the end of the trial. This might be partly due to smaller trial counts in this category (48 art and 16 Mooney trials per participant). Instantaneous heart rate seems to predict the category slightly better early in the trials. Taken together this strongly suggests that the stimuli are not processed identically during repeated presentation. **b) stimulus domain (Mooney image or artwork):** the image category can be reliably decoded right from the beginning both from EEG, pupil data, and heart rate, again best from the time-domain EEG. Notably, classification works more reliably during the second image presentation than during initial presentation (except in pupil data, where we barely see a difference). The frequency-domain EEG (except gamma band) did also show noisy but significant decoding results for both categories; see Supp. A.7 for plots of all trials and conditions. Decoding timecourses were computed for each individual participant in a stratified 10-fold cross validation; significance was determined by second level statistical testing (cluster based permutation t-test, significance level $p < .05$). X-axis reflects the chance level (0.5 for the used AUC metric).

Ratings of aesthetic liking showed a brief peak of significant decoding in the time-domain EEG signal in posttitle trials (0.46-0.68 s after image onset). The pupil signal, though not significant, generally delivered above chance decoding with a marked peak in early pretitle trials (about the first second after image onset), while not exhibiting this peak in posttitle trials but rather ramping up over time.

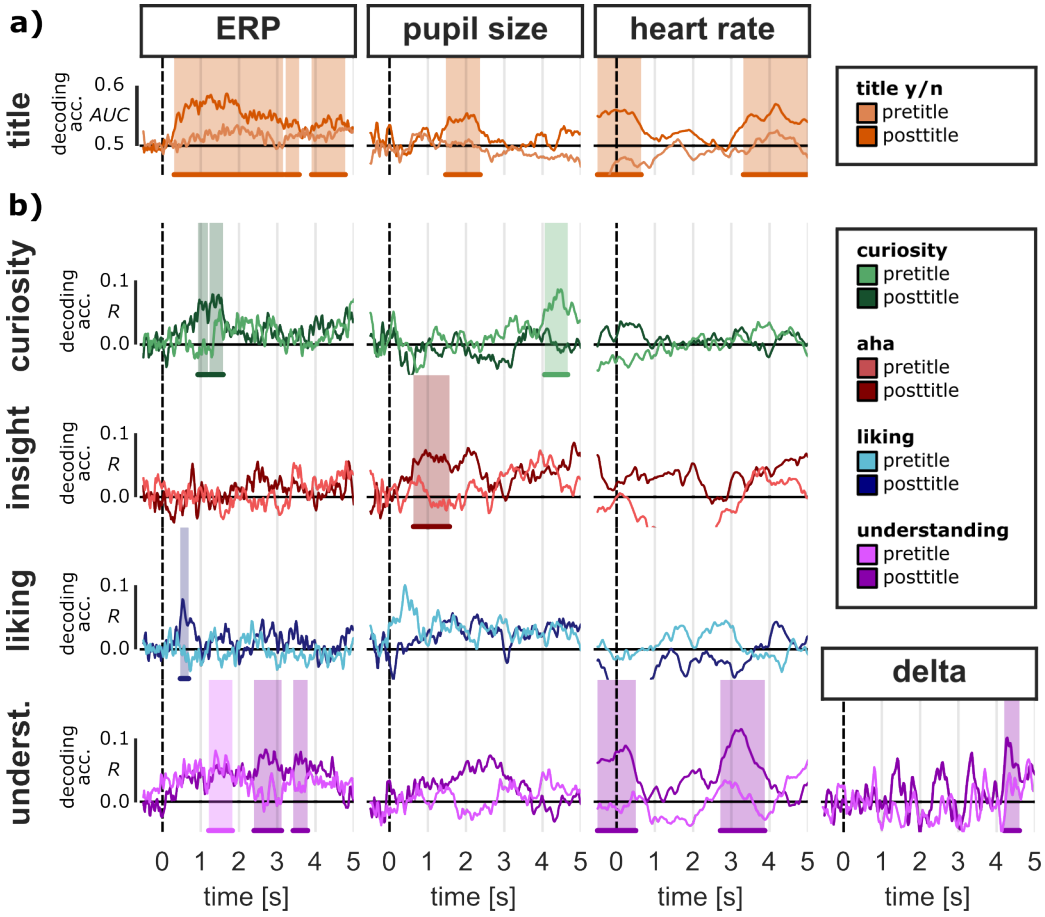


Figure 4: **Decoding physiological correlates of meaning making (artwork trials only) – a) Decoding the "title-effect":** we were able to significantly decode whether a painting stimulus was paired with their real title or an uninformative dummy label ("untitled"). This worked best from time-domain EEG both during second viewing (posttitle), but also from pupil size and heart rate (heart rate during 2nd presentation seems still affected from earlier processing, e.g. during title presentation, and decoding results are best before and shortly after image onset). As expected, the title condition can not be decoded in during first presentation of the images (before the title was presented). **b) Decoding trial wise ratings of curiosity for the title, *aha*, aesthetic liking, and felt understanding:** we could decode each type of rating from one or more of the recorded physiological signals. Felt understanding could be best decoded, from time-domain EEG (both in pre- and posttitle trials), from late delta band power (in posttitle trials), and from heart rate (in posttitle trials, with a timecourse similar to those for decoding the title condition, though a little earlier). Pupil dilation exhibits above chance decoding during post title trials too (with a timecourse similar to ERPs), but does not become significant. Curiosity for the title could be significantly decoded from the late pretitle pupil signal and the early posttitle time-domain EEG. Insight (*aha*) ratings could be significantly decoded from pupil size early in posttitle trials. Ratings of aesthetic liking showed a short peak of significant decoding in the posttitle time-domain EEG. While not becoming significant, the pupil size signal delivered generally above chance decoding with a marked peak in early pretitle trials. All other data modalities did show non-significant results in pre or posttitle trials; see Supp. A.7 for plots of all trials and conditions. Decoding timecourses were computed for each individual participant in a stratified 10-fold cross validation; significance was determined by second level statistical testing (cluster based permutation t-test, significance level $p < .05$). X-axis reflects the chance level (0.5 for the AUC metric in binary decoding, and 0 for the spearman R metric in continuous ratings).

4 Discussion

We aimed to study the relationship between curiosity, insight, felt understanding, and aesthetic liking during meaning making with visual art, both in the domain of behavioral ratings and (neuro)physiological correlates. We were able to replicate a number of earlier findings from both the behavioral literature on insight, and on aesthetic valuation. Yet crucially we could shine a light on the intricate interplay between the investigated component processes, arriving at a better understanding of the role of information processing in aesthetic valuation. Using a time-resolved decoding approach and multi-modal physiological data (EEG, pupillometry, heart rate) we aimed to investigate the temporal profile of the cognitive component processes and found evidence for both fast responses as well as slower processes that build up over time. While the findings did not confirm all our hypotheses, we believe that this multi-modal approach can lead to a more mechanistic understanding of the aesthetic evaluation process and its interplay with meaning making.

4.1 The relationship between ratings of curiosity, insight, felt understanding, and aesthetic valuation

The behavioral part of this study (Exp. 1) supported most of our preregistered hypotheses. We were able to both replicate a strong effect of title-manipulation on perceived understanding ratings with visual art and to establish the same (even stronger) effect of titles on *aha* ratings. We did not find a significant title effect on aesthetic preference ratings, though, thereby adding to an already inconclusive literature. As reported many times before, we saw differences in aesthetic preference ratings across different stimulus categories, with generally higher ratings for figurative art than for abstract art, with semi-abstract art lying in between. This difference was even greater for perceived understanding. In *aha* ratings however, there was a mere trend towards slightly higher *aha* for semi-abstract art than for figurative or abstract art.

Next to visual art we had included a set of Mooney images and matching "solutions". Those are known to often cause perceptual shifts and evoke strong feelings of *aha*, and can hence act as a type of benchmark. Furthermore, several previous studies have worked on similar questions as the current study using Mooney images (but not visual art; Muth & Carbon, 2013; Van de Cruys et al., 2021). As expected, *aha* ratings were higher for Mooney images than for all artwork subcategories except semi-abstract art with explanatory titles – a category in some ways similar to the Mooney images as they can both lead to *perceptual* shifts after an explanation was given (e.g., in cubist paintings). Aesthetic preference ratings, on the other hand, were on average slightly lower for Mooney images than for art. With respect to felt understanding Mooney images lay between figurative and abstract titled art, on par with the semi-abstract stimuli. Curiosity for a title or solution predicted subsequent *aha* (as previously reported for Mooney images; Van de Cruys et al., 2021), though our data suggests that strength and linearity of this association depends on the information content of the title cues. *Aha* predicts liking (quite linearly, as previously reported for Mooney images; Muth & Carbon, 2013) but also felt understanding, non-linearly interacting with the degree of abstraction of the stimuli.

These are interesting findings, as taken together the observed rating patterns extend our previous understanding of the role of insight in meaning making and aesthetic preference formation. The association of both curiosity and understanding with *aha* (especially in the high *aha* range) supports the hypothesis that *aha* might be a valid indicator of information gain or uncertainty reduction (in line with the conclusions of Van de Cruys et al., 2021). Trial-wise curiosity ratings, in this logic, might reflect the expected *potential* for information gain – its actualization, however, depends on the additional semantic information provided, as shown by the nonlinear effect of title-style (i.e. the amount of newly provided semantic information) on the relationship between curiosity and *aha* after the title (see fig 2d): the more uncertainty reduction is possible or expected (indicated by high curiosity ratings), the stronger the effect of the given cue and hence the association between curiosity and *aha* (with the strongest relation between curiosity and *aha* observed in Mooney images, and the weakest in untitled art; see Fig. 2d).

Low felt understanding, on the other hand, might be interpreted as an indicator of remaining or residual uncertainty: strong *aha* (large information gain) was consistently associated with high understanding, but depending on the stimuli

(art category and title style) high understanding ratings also appeared in low *aha* trials. This was particularly the case for figurative art and Mooney images (see Fig. 2b) where low *aha* followed particularly low curiosity ratings (see also Supp. Fig. A.3). It is possible that for a good fraction of these stimuli participants were less curious about the title/solution because there was little uncertainty in the first place, and hence less potential for *aha* – a figurative painting of a lush meadow or a Mooney image that was already recognized on first sight might not make one as eager for the title or base image, while abstract and semi-abstract artworks have a relatively low likelihood of appearing entirely unambiguous. The EEG decoding results lend support to this interpretation, as understanding was best decodable towards the end of 2nd trial (both from time-domain EEG and theta power), when all information had been brought together.

Liking ratings, on the other hand, were positively and rather linearly predicted by all other ratings – curiosity, *aha*, and understanding. These associations were largely unaffected by other factors such as art style or title style (see Fig. 2d and Supp. A.3). At first glance this might suggest that the role of aesthetic valuation (or at least our rating operationalization in this study) was in fact quite plain: aesthetic liking simply followed the other positively valenced judgments, while these ratings themselves can have a more intricate genesis. Yet, our data also contains arguments against such a simplistic causal interpretation: while we (as many groups before us) found that figurative art received on average higher aesthetic ratings than semi-abstract and abstract art, this was not the case for *aha* and curiosity – the artwork category did not matter much for *aha* ratings (with only semi-abstract art scoring higher than the others) and curiosity was even markedly lower for figurative art than for the more abstract categories. Furthermore, there was no significant effect of our title manipulation on aesthetic liking, while it substantially affected both felt understanding and *aha* (for curiosity the comparison does not make sense as the titles were presented after collecting this rating). Although all three ratings linearly predicted liking on trial-wise bases, liking does not appear to simply follow the other ratings, or the grand average patterns would be more similar. Furthermore, in the decoding timecourses liking exhibited a peak of decodability early in the trials, well before the other ratings (further discussion below). Building on our findings, future research might design intervention studies to have a closer, ideally more causal, look at these interrelations.

4.2 Physiological correlates and the temporal sequence of information processing

This part of the study was more exploratory, and we did only preregister a few coarse hypotheses based on previous work – namely an association between EEG alpha and gamma band activity with *aha* ratings and liking ratings. We did not find evidence supporting these hypotheses, though. While our multivariate decoding approach was capable to well identify the basic experimental manipulations based on time-domain EEG, frequency-domain EEG (except gamma band), heart rate and pupil data, the graded ratings proved harder to recover, and better from the various time-domain signals than from the frequency-domain EEG.

Decoding from EEG

Among the ratings, felt understanding (a rating we did not find any EEG work on and hence did not formulate hypotheses) had the strongest signature in the EEG. Interestingly there seemed to be an initial phase of increased decodability in both pre- and posttitle trials (roughly the first 2 seconds of image presentation, partly significant only in the pretitle condition and hence apparently not dependent on the title information provided). In the posttitle trials however, the decoding rate did not drop after this initial phase but rather built up and became significant later in the trial – this would be consistent with accumulating evidence leading to a sustained understanding signal in the data that can be captured. This interpretation is further supported by the relatively similar (though not significant) decoding timecourse in the pupil trace, and the late significant peak of decoding rate in the instantaneous heart rate (see Fig. 4b).

Decoding timecourses for aesthetic liking showed a salient peak in the first second after image onset in the time-domain EEG of posttitle trials. This offers further neurophysiological evidence for the presence of a very fast aesthetic

response component, consistent with previous observations that some ERP components might be linked to aesthetic processing (OliverRodríguez et al., 1999; Marzi & Viggiano, 2010; Werheid et al., 2007; de Tommaso et al., 2008; Augustin et al., 2011; Wang et al., 2012; Jiang & Cai, 2013; Noguchi & Murota, 2013; Ma et al., 2015; Bölte et al., 2017; Righi et al., 2017; Carbon et al., 2018). The early peak in the pupil data of pretitle trials, though not significant, might further back this claim. Importantly, liking ratings could only be decoded from the EEG of the 2nd presentation of images, post title, but not from the initial presentation. Apparently, the early cortical response to the image differed between 1st and 2nd presentation (see also grand average response in Supp. Figs. A.4, A.8 and discussion below) and this difference was relevant in decoding aesthetic liking. This might speak against a dominant role of low-level image features on final liking ratings. It is possible that the EEG response is present in the pre-title trial too (see pupil response) but overshadowed by noise and the far greater temporal distance to the rating. However, it could also be that the liking response collected after the second viewing deviated from a hypothetical liking rating that would have been given after the initial viewing. Given our behavioral findings we should almost expect the title intervention and related meaning making processes to cause such divergence, or evolution of liking, and other contributors like the repeated or longer exposure (cf. Zajonc, 1968; Burgess & Sales, 1971), unknown factors, or simply noise in the ratings seem plausible, too. Unfortunately, our data does not allow for a more thorough investigation of this effect, as we lack a measure of change in liking from before to after seeing the title – while collecting such a measure would be methodologically challenging, future research could follow up on this. We find it noteworthy that we did not see significant decoding of liking later in the trials – previous EEG (Strijbosch et al., 2021) and behavioral findings (Brielmann & Pelli, 2017), as well as our own behavioral observation in this study would lead us to expect also slower contributions to a final liking response. Further research should validate or rectify these null findings.

The early peak for liking was not immediately mirrored for *aha*; while *aha* could be significantly decoded from posttitle trials (in the pupil signal) it became significant *after* the peak for liking. Perhaps this finding is misleading, though: the pupil response is a comparatively slow signal (and might be unreliable immediately after image onset; see discussion below), so it seems possible that the actual *aha* causing the pupil response happened slightly earlier, around the same timerange as the liking response. Still, our findings do not strongly support the hypothesis that moments of aha would causally lead to increased liking.

Curiosity proved a very interesting signal: while the significant cluster in the late pretitle pupil signal (immediately prior to the rating) might be unsurprising, the robust decodability from the time-domain EEG of posttitle trials certainly is. Participants have registered their rating several seconds prior to this timepoint, so it cannot be a case of reverse attribution. Furthermore, they had to process an entirely different stimulus in the meantime (the title cues). One possible interpretation would be, that the curiosity rating does in fact accurately reflect a distinct state of cortical activity, and that this activity is sustained over several seconds from the initial presentation of the image, over seeing and processing the title cue, and well into the posttitle trial. This would be a very important finding, as these altered states could also affect the cortical processing and contextualization of the new title information with the repeated image stimulus. The fact that the curiosity rating can be significantly decoded only after the peak in liking ratings (also in the time-domain EEG) is also relevant, as it makes it appear less likely that liking is a direct consequence of curiosity, despite their linear relation in the rating space (see discussion above).

In general, liking, curiosity and *aha* ratings proved hard to recover from the EEG data. There are several potential reasons why this might be the case. Many of the previously proposed EEG biomarkers for insight and/or aesthetic valuation are patterns in the frequency-domain EEG. In this light, it seems surprising that in our data the EEG band power is such a poor feature to predict the ratings. This does not necessarily mean that these correlates are not present, though. Poor decoding might also be caused by poor temporal synchronicity of physiological responses across trials: with the current decoding pipeline, short bursts of activity, as proposed for *aha* (especially gamma band activity; Oh et al., 2020), would be most reliably detected if temporally aligned (i.e. immediately time-locked to the stimulus) and otherwise easily lost in noise. Previous EEG studies on insight typically addressed this issue by reverse time-locking analyses to a participant's button press indicating moments of insight – we had decided against this procedure for our

study, as 1) it disrupted the flow of the paradigm, 2) we were interested in a graded rating, and 3) we explicitly wanted to relate this continuous rating of *aha* to other ratings like curiosity and liking. While we knew in advance that there would not be clear alignment for these internally triggered processes (insight or aesthetic liking can evolve anytime or not at all), we nevertheless expected that the alignment would be good enough to detect more sustained or partly overlapping correlates. Other analysis approaches might be better suited to find asynchronous patterns and could be tried in the future. Interesting candidate pipelines might be based on unsupervised learning (e.g. Schneider et al., 2023), temporal deconvolution (e.g. Ehinger & Dimigen, 2019), or simply decoding with flexible timelags. Another way would be to further increase the signal to noise level in the recorded data. The most effective option would be to use intracranial recordings, as e.g. done by Lopez-Persem et al. (2020); however, to date this also implies transitioning to smaller and very different sample populations (i.e. patients). More phasic frequency responses on the other hand should be detectable. In our data this seems to be the case in the alpha, beta, and even gamma band for some of the categorical decoding targets (1st vs 2nd presentation and artwork vs. Mooney; see Supp. A.7).

We could further reliably decode the categorical experimental factors. Particularly the fact that the title effect can be decoded from time-domain signals of posttitle trials (see Fig. 4a) is more than a sanity check and very encouraging. This contrast, in which the only difference was in the text information shown prior to image onset, might remind the reader of work on contextual priming (e.g. Wirth et al., 2008; Horner & Henson, 2012) though differences in the stimulus material and priming procedures prevent a direct comparison. In this field repetition suppression can play a conceptual role with different priming stimuli thought to moderate the gained ease of processing upon repeated presentation (either further facilitating, or interfering). Our title cues can certainly be seen as primer stimuli and both processing fluency (Reber et al., 1998) as well as repeated exposure (though inconclusively; see e.g. Zajonc 1968; Burgess and Sales 1971, but also Park et al. 2010) are though to play a role in aesthetic judgment. But would the titles facilitate or complicate processing of a repeated visual art stimulus? This most likely depends on the specific image stimuli and the title (see our behavioral findings). To widen the picture again, cortical activity related to curiosity carried over into the 2nd image presentation regardless of which type of title stimulus was presented, speaking against a monopolistic influence of the priming title cues. The interplay between repeated presentation effects, ease of processing, and additional information in processing visual art (and meaning making more general) is indeed very interesting – we might have opened a view into these intricate processes and their interrelations, but further research is necessary to progress our understanding.

Interestingly, pre- vs post-title presentation was better decodable in art trials than in Mooney images (judging by length of significant timewindows; see Fig. 3). One explanation could be that repetition effects are particularly strong in art (see discussion above). Alternatively, Mooney images might show particularly weak repetition suppression – either the prominent low level visual features of Mooney images might govern physiological responses, allowing for relatively smaller effects of repeated presentation, or the perceptual shift that can happen after understanding the solution effectively renders the second presentation of a Mooney image a different stimulus, resulting in less repetition suppression. Previous work has in fact linked the strength of repetition suppression to surprise and expectations (Summerfield et al., 2008; Summerfield et al., 2011). However, qualitatively the decoding timecourses do not differ that much and it is also possible that we simply observed false negative findings. There were fewer trials in the Mooney-only condition than in the art condition (16 Mooney vs 48 artworks per participant), resulting in lower statistical power.

Decoding from pupillometry and heart rate

As expected, pupillometry turned out to be a very promising data modality: it could be used to reliably decode the categorical factors (1st or 2nd presentation, Mooney or artwork), showed significant timewindows in the title effect, curiosity and *aha* ratings, and non-significant but above chance decoding with interesting temporal profiles for liking and understanding (see Figs. 3 and 4). While findings for aesthetic liking showed no significant effects, we saw above chance decoding throughout most of the trial and a marked peak of decodability in the first second of pretitle trials (not present in posttitle trials). If true, this would be further evidence for a first and fast component in preference

formation (cf. posttitle ERP data). Importantly, pupil size was also the only recorded data modality that we could significantly predict *aha* ratings from (besides a short cluster in theta band EEG of the differential signal; see Supp. A.6). Perhaps the pupil, as a comparatively slow or inert signal (though not as slow as heart rate), is less affected by the lack of synchronization across insight moments. This might be consistent with the decoding timecourse in the differential signal (see Supp. A.6 and Supp. Fig. A.8) that exhibited a slow rather monotonic increase over the duration of the trial. In posttitle trials there was also a significant window early on in the trials, becoming significant about half a second after image onset. It is interesting to compare this with the decoding of curiosity ratings, which showed marked significant effects late in the pretitle trials, in the last second before image offset. Taken together this pattern makes the pupil a very promising candidate for future research on the interplay of curiosity, insight and understanding in information processing (and perhaps also for work on preference formation).

Unfortunately, pupil size early in the trial might not be a reliable signal for higher-level cognitive processes. In our data, we see that no rating nor the title condition can be decoded in the early part of trials. At this time the pupil adjustments to luminance changes on the screen (about the first 1 s; see grand average pupil traces, in Supp. Fig. A.4). Note however, that the interesting non-significant pattern in pretitle liking ratings falls exactly in this temporal profile, and significant decodability of *aha* in posttitle trials starts immediately afterwards, potentially indicating that the readjustment of the pupil is affected by *aha*. The contrast between Mooney images and artwork, on the other hand, can be particularly well decoded in this timewindow; this is not surprising given that the Mooney images are high contrast black/white images with a large fraction of pure white canvas space – hence the pupil should adapt stronger to those images, compared to artworks which are on average less luminous. Further supporting this interpretation, the 1st vs 2nd presentation *within Mooney images* can not be significantly decoded from the early pupil data (only later in the trial) while this works very reliably for artwork stimuli, or from the other data modalities like EEG (see Fig. 3 and Supp. A.7). These patterns are consistent with the explanation that low level stimulus features are driving the pupil's response in this part of the trials.

Heart rate, a signal closely linked to the autonomous nervous system via sympathetic/parasympathetic pathways, was another candidate data modality to be affected by aesthetic processing. As in previous work we observed a marked heart rate deceleration following the onset of visual stimuli (Palomba et al., 1997; Welke & Vessel, 2022). This deceleration was stronger for the first (pretitle) presentation than the second presentation, mainly because in the second presentation heart rate started on an already lower level (see Supp. A.4). Previous events, during the pretitle trial or the title presentation itself, apparently spilled over into the posttitle trials due to the sluggishness of the signal. This explains why we can decode events like the image category or title condition, but also the graded understanding ratings, already in the pre-stimulus period before the second image onset (importantly, unlike most other data modalities, we did not baseline correct the heart rate data). In the pretitle trials, no condition or rating could be decoded from heart rate.

4.3 Conclusion

In two preregistered experiments we employed the "title effect" – the fact that text-based semantic information accompanying an artwork, such as titles, can change how an observer judges or understands a piece of visual art – to investigate the epistemic arc during visual aesthetic experiences. In a first largely hypothesis driven behavioral part we established a paradigm to robustly evoke insight and could progress our understanding of the intricate interplay between curiosity, insight, felt understanding, and aesthetic preference, revealing partly nonlinear relations shaped by the type and amount of information a specific stimulus provides. In a second more exploratory neurophysiological part we characterized the temporal profile of these interrelated component processes using a time-resolved decoding pipeline and a variety of data features from EEG, heart rate and pupillometry. While our study did not confirm all our hypotheses, we believe that this multi-modal data approach opened new angles to further investigate the topic in the future and can lead to a more mechanistic understanding of meaning making and its role in the aesthetic evaluation process.

5 Additional information

Data and code availability

The experimental design and hypotheses reported here were preregistered (https://doi.org/10.17605/OSF.IO/PAGVH and https://doi.org/10.17605/OSF.IO/6KW9N). The dataset for statistical analysis, as well as python scripts replicating the results, and python scripts used in data collection were made available in a public online repository (https://osf.io/j42g5). The stimulus set is entirely in the public domain and included in the online repository; all artwork stimuli are further depicted in the online Supplement of this paper. An earlier stage of this article was uploaded to a preprint server (https://www.biorxiv.org/TBD).

Acknowledgments

This work was supported by the Max Planck Society.

Author statement (CRediT)

DW: Conceptualization, Methodology, Investigation, Formal Analysis, Software, Data curation, Project administration, Visualization, Writing - Original Draft, Writing - Review & Editing EV: Conceptualization, Writing - Original Draft, Writing - Review & Editing, Resources, Supervision, Funding acquisition

References

Appelhoff, S., Sanderson, M., Brooks, T., van Vliet, M., Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., Tavabi, K., Höchenberger, R., Welke, D., Brunner, C., Rockhill, A., Larson, E., Gramfort, A., & Jas, M. (2019). MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. *Journal of Open Source Software*, *4*(44), 1896. https://doi.org/10.21105/joss.01896

Augustin, M. D., Defranceschi, B., Fuchs, H. K., Carbon, C.-C., & Hutzler, F. (2011). The neural time course of art perception: An ERP study on the processing of style versus content in art. *Neuropsychologia*, *49*(7), 2071–2081. https://doi.org/10.1016/j.neuropsychologia.2011.03.038

Babiloni, F., Cherubino, P., Graziani, I., Trettel, A., Infarinato, F., Picconi, D., Borghini, G., Maglione, A. G., Mattia, D., & Vecchiato, G. (2013). Neuroelectric brain imaging during a real visit of a fine arts gallery: A neuroaesthetic study of XVII century dutch painters. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6179–6182. https://doi.org/10.1109/EMBC.2013.6610964

Babiloni, F., Rossi, D., Cherubino, P., Trettel, A., Picconi, D., Maglione, A. G., Vecchiato, G., de Vico Fallani, F., Chavez, M., & Babiloni, F. (2015). The first impression is what matters: A neuroaesthetic study of the cerebral perception and appreciation of paintings by titian. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 7990–7993. https://doi.org/10.1109/EMBC.2015.7320246

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00456

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Belke, B., Leder, H., Strobach, T., & Carbon, C.-C. (2010). Cognitive fluency: High-level processing dynamics in art appreciation. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(4), 214–222. https://doi.org/10.1037/a0019648

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00016

Bölte, J., Hösker, T. M., Hirschfeld, G., & Thielsch, M. T. (2017). Electrophysiological correlates of aesthetic processing of webpages: A comparison of experts and laypersons. *PeerJ*, *5*, e3440. https://doi.org/10.7717/peerj.3440

Brielmann, A. A., & Pelli, D. G. (2017). Beauty requires thought. *Current Biology*, *27*(10), 1506–1513.e3. https://doi.org/10.1016/j.cub.2017.04.018

Bubi, A., Suac, A., & Palmovi, M. (2017). Observing individuals viewing art: The effects of titles on viewers eye-movement profiles. *Empirical Studies of the Arts*, *35*(2), 194–213. https://doi.org/10.1177/0276237416683499

Bühler, E., Lachenmeier, D. W., & Winkler, G. (2014). Development of a tool to assess caffeine intake among teenagers and young adults. *Ernahrungs Umschau*, *61*(4), 58–63. https://doi.org/10.4455/eu.2014.011

Burgess, T. D. G., & Sales, S. M. (1971). Attitudinal effects of mere exposure: A reevaluation. *Journal of Experimental Social Psychology*, *7*(4), 461–472. https://doi.org/10.1016/0022-1031(71)90078-3

Carbon, C.-C., Faerber, S. J., Augustin, M. D., Mitterer, B., & Hutzler, F. (2018). First gender, then attractiveness: Indications of gender-specific attractiveness processing via ERP onsets. *Neuroscience Letters*, *686*, 186–192. https://doi.org/10.1016/j.neulet.2018.09.009

Cela-Conde, C. J., Marty, G., Maestu, F., Ortiz, T., Munar, E., Fernandez, A., Roca, M., Rossello, J., & Quesney, F. (2004). Activation of the prefrontal cortex in the human visual aesthetic perception. *Proc. Natl. Acad. Sci. U.S.A.*, *101*(16), 6321–6325.

Cheung, M.-C., Law, D., Yip, J., & Wong, C. W. Y. (2019). Emotional responses to visual art and commercial stimuli: Implications for creativity and aesthetics. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00014

Commitee. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology*, *10*(2), 370–375. https://doi.org/10.1016/0013-4694(58)90053-1

Cristofori, I., Salvi, C., Beeman, M., & Grafman, J. (2018). The effects of expected reward on creative problem solving. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(5), 925–931. https://doi.org/10.3758/s13415-018-0613-5

Delorme, A. (2023). EEG is better left alone [Publisher: Nature Publishing Group]. *Scientific Reports*, *13*(1), 2372. https://doi.org/10.1038/s41598-023-27528-0

de Tommaso, M., Pecoraro, C., Sardaro, M., Serpino, C., Lancioni, G., & Livrea, P. (2008). Influence of aesthetic perception on visual event-related potentials. *Consciousness and Cognition*, *17*(3), 933–945. https://doi.org/10.1016/j.concog.2007.09.003

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *NeuroImage*, *207*, 116–117. https://doi.org/10.1016/j.neuroimage.2019.116117

Ehinger, B. V., & Dimigen, O. (2019). Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *PeerJ*, *7*, e7838. https://doi.org/10.7717/peerj.7838

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gerger, G., & Leder, H. (2015). Titles change the esthetic appreciations of paintings. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00464

Gwilliams, L., & King, J.-R. (2020). Recurrent processes support a cascade of hierarchical decisions (T. Serre & M. J. Frank, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *9*, e56603. https://doi.org/10.7554/eLife.56603

Horner, A. J., & Henson, R. N. (2012). Incongruent abstract stimulusresponse bindings result in response interference: fMRI and EEG evidence from visual object classification priming. *Journal of Cognitive Neuroscience*, *24*(3), 760–773. https://doi.org/10.1162/jocn_a_00163

Hristova, E., Georgieva, S., & Grinberg, M. (2011). Top-down influences on eye-movements during painting perception: The effect of task and titles [Series Title: Lecture Notes in Computer Science]. In A. Esposito, A. M. Esposito, R. Martone, V. C. Müller, & G. Scarpetta (Eds.), *Toward autonomous, adaptive, and context-aware multimodal interfaces. theoretical and practical issues* (pp. 104–115). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-18184-9_10

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Jiang, X., & Cai, L. (2013). Evaluation of aesthetic response to clothing color combination: A behavioral and electrophysiological study. *Journal of Fiber Bioengineering and Informatics*, *6*(4), 405–414. https://doi.org/10.3993/jfbi12201305

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., Reber, P. J., & Kounios, J. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biology*, *2*(4), 11. https://doi.org/https://doi.org/10.1371/journal.pbio.0020097

Kang, J.-H., Kim, S. J., Cho, Y. S., & Kim, S.-P. (2015). Modulation of alpha oscillations in the human EEG with facial preference (M. Ptito, Ed.). *PLOS ONE*, *10*(9), e0138153. https://doi.org/10.1371/journal.pone.0138153

Kapoula, Z., Daunys, G., Herbez, O., & Yang, Q. (2009). Effect of title on eye-movement exploration of cubist paintings by fernand léger. *Perception*, *38*(4), 479–491. https://doi.org/10.1068/p6080

Kim, H. (2017). Brain regions that show repetition suppression and enhancement: A metaanalysis of 137 neuroimaging experiments. *Human Brain Mapping*, *38*(4), 1894–1913. https://doi.org/10.1002/hbm.23492

Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology*, *65*(1), 71–93. https://doi.org/10.1146/annurev-psych-010213-115154

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Laukkonen, R. E., Webb, M., Salvi, C., Tangen, J. M., Slagter, H. A., & Schooler, J. W. (2023). Insight and the selection of ideas. *Neuroscience & Biobehavioral Reviews*, *153*, 105363. https://doi.org/10.1016/j.neubiorev.2023.105363

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

Lee, I. E., Latchoumane, C.-F. V., & Jeong, J. (2017). Arousal rules: An empirical investigation into the aesthetic experience of cross-modal perception with emotional visual music. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00440

Liao, H.-I., Makoto, Y., Kashino, M., & Furukawa, S. (2018). Pupillary dilation response reflects surprising moments in music. *Journal of Eye Movement Research*, *11*(2). https://doi.org/10.16910/jemr.11.2.13

Lopez-Persem, A., Bastin, J., Petton, M., Abitbol, R., Lehongre, K., Adam, C., Navarro, V., Rheims, S., Kahane, P., Domenech, P., & Pessiglione, M. (2020). Four core properties of the human brain valuation system demonstrated in intracranial signals [Publisher: Springer US]. *Nature Neuroscience*, *23*(5), 664–675. https://doi.org/10.1038/s41593-020-0615-9

Ma, Q., Hu, L., & Wang, X. (2015). Emotion and novelty processing in an implicit aesthetic experience of architectures: Evidence from an event-related potential study. *NeuroReport*, *26*(5), 279–284. https://doi.org/10.1097/WNR.0000000000000344

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

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

Mastandrea, S., & Umiltà, M. A. (2016). Futurist art: Motion and aesthetics as a function of title. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00201

Millis, K. (2001). Making meaning brings pleasure: The influence of titles on aesthetic experiences. *Emotion*, *1*(3), 320–329. https://doi.org/10.1037/1528-3542.1.3.320

Mohammad, S., & Kiritchenko, S. (2018). WikiArt emotions: An annotated dataset of emotions evoked by art. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Moon, J., Kim, Y., Lee, H., Bae, C., & Yoon, W. C. (2013). Extraction of user preference for video stimuli using EEG-based user responses. *ETRI Journal*, *35*(6), 1105–1114. https://doi.org/10.4218/etrij.13.0113.0194

Mooney, C. M., & Ferguson, G. A. (1951). A new closure test. *Canadian Journal of Psychology / Revue canadienne de psychologie*, *5*(3), 129–133. https://doi.org/10.1037/h0083540

Mullennix, J. W., Pilot, K. M., Steeves, T. A., & Burns, J. C. (2018). The effects of cognitive load on judgments of titled visual art. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(2), 166–176. https://doi.org/10.1037/aca0000128

Munar, E., Nadal, M., Castellanos, N. P., Flexas, A., Maestú, F., Mirasso, C., & Cela-Conde, C. J. (2012). Aesthetic appreciation: Event-related field and time-frequency analyses. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00185

Munar, E., Nadal, M., Rosselló, J., Flexas, A., Moratti, S., Maestú, F., Marty, G., & Cela-Conde, C. J. (2012). Lateral orbitofrontal cortex involvement in initial negative aesthetic impression formation (L. M. Martinez, Ed.). *PLoS ONE*, *7*(6), e38152. https://doi.org/10.1371/journal.pone.0038152

Muth, C., & Carbon, C.-C. (2013). The aesthetic aha: On the pleasure of having insights into gestalt. *Acta Psychologica*, *144*(1), 25–30. https://doi.org/10.1016/j.actpsy.2013.05.001

Nicholls, M. E., Thomas, N. A., Loetscher, T., & Grimshaw, G. M. (2013). The flinders handedness survey (FLANDERS): A brief measure of skilled hand preference. *Cortex*, *49*(10), 2914–2926. https://doi.org/10.1016/j.cortex.2013.02.002

Noguchi, Y., & Murota, M. (2013). Temporal dynamics of neural activity in an integration of visual and contextual information in an esthetic preference task. *Neuropsychologia*, *51*(6), 1077–1084. https://doi.org/10.1016/j.neuropsychologia.2013.03.003

Oh, Y., Chesebrough, C., Erickson, B., Zhang, F., & Kounios, J. (2020). An insight-related neural reward signal. *NeuroImage*, *214*, 116757. https://doi.org/10.1016/j.neuroimage.2020.116757

OliverRodríguez, J. C., Guan, Z., & Johnston, V. S. (1999). Gender differences in late positive components evoked by human faces. *Psychophysiology*, *36*(2), 176–185. https://doi.org/10.1111/1469-8986.3620176

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713–719. https://doi.org/10.1016/S1388-2457(00)00527-7

Palomba, D., Angrilli, A., & Mini, A. (1997). Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli. *International Journal of Psychophysiology*, *27*(1), 55–67. https://doi.org/https://doi.org/10.1016/S0167-8760(97)00751-4

Park, J., Shimojo, E., & Shimojo, S. (2010). Roles of familiarity and novelty in visual preference judgments are segregated across object categories. *Proceedings of the National Academy of Sciences*, *107*(33), 14552–14555. https://doi.org/10.1073/pnas.1004374107

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pedroni, A., Bahreini, A., & Langer, N. (2019). Automagic: Standardized preprocessing of big EEG data. *NeuroImage*, *200*, 460–473. https://doi.org/10.1016/j.neuroimage.2019.06.046

Peirce, J. W. (2007). PsychoPypsychophysics software in python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, *6*(1). https://doi.org/10.1038/s41597-019-0104-8

Pernet, C. R., Garrido, M., Gramfort, A., Maurits, N., Michel, C., Pang, E., Salmelin, R., Schoffelen, J. M., Valdes-Sosa, P. A., & Puce, A. (2018, August 9). *Best practices in data analysis and sharing in neuroimaging using MEEG* (preprint). Open Science Framework. https://doi.org/10.31219/osf.io/a8dhx

Pernet, C. R., Garrido, M. I., Gramfort, A., Maurits, N., Michel, C. M., Pang, E., Salmelin, R., Schoffelen, J. M., Valdes-Sosa, P. A., & Puce, A. (2020). Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. *Nature Neuroscience*. https://doi.org/10.1038/s41593-020-00709-0

Quax, S. C., Dijkstra, N., van Staveren, M. J., Bosch, S. E., & van Gerven, M. A. (2019). Eye movements explain decodability during perception and cued attention in MEG. *NeuroImage*, *195*, 444–453. https://doi.org/10.1016/j.neuroimage.2019.03.069

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2020). Validation of the short and extra-short forms of the big five inventory-2 (BFI-2) and their german adaptations. *European Journal of Psychological Assessment*, *36*(1), 149–161. https://doi.org/10.1027/1015-5759/a000481

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments [Publisher: SAGE Publications Inc]. *Psychological Science*, *9*(1), 45–48. https://doi.org/10.1111/1467-9280.00008

Righi, S., Gronchi, G., Pierguidi, G., Messina, S., & Viggiano, M. (2017). Aesthetic shapes our perception of everyday objects: An ERP study. *New Ideas in Psychology*, *47*, 103–112. https://doi.org/10.1016/j.newideapsych.2017.03.007

Robbins, K. A., Touryan, J., Mullen, T., Kothe, C., & Bigdely-Shamlo, N. (2020). How sensitive are EEG results to preprocessing methods: A benchmarking study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(5), 1081–1090. https://doi.org/10.1109/TNSRE.2020.2980223

Roth, H. L. (2002). Effects of monocular viewing and eye dominance on spatial attention. *Brain*, *125*(9), 2023–2035. https://doi.org/10.1093/brain/awf210

Russell, P. A., & Milne, S. (1997). Meaningfulness and hedonic value of paintings: Effects of titles. *Empirical Studies of the Arts*, *15*(1), 61–73. https://doi.org/10.2190/EHT3-HWVM-52CB-8QHJ

Russell, P. A. (2003). Effort after meaning and the hedonic value of paintings. *British Journal of Psychology*, *94*(1), 99–110. https://doi.org/10.1348/000712603762842138

Salvi, C., Simoncini, C., Grafman, J., & Beeman, M. (2020). Oculometric signature of switch into awareness? pupil size predicts sudden insight whereas microsaccades predict problem-solving via analysis. *NeuroImage*, *217*, 116933. https://doi.org/10.1016/j.neuroimage.2020.116933

Sandkühler, S., & Bhattacharya, J. (2008). Deconstructing insight: EEG correlates of insightful problem solving. *PLOS ONE*, *3*(1). https://doi.org/10.1371/journal.pone.0001459

Schacht, A., Werheid, K., & Sommer, W. (2008). The appraisal of facial beauty is rapid but not mandatory. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(2), 132–142. https://doi.org/10.3758/CABN.8.2.132

Schlotz, W., Wallot, S., Omigie, D., Masucci, M. D., Hoelzmann, S. C., & Vessel, E. A. (2020). The aesthetic responsiveness assessment (AReA): A screening tool to assess individual differences in responsiveness to art in english and german. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000348

Schneider, S., Lee, J. H., & Mathis, M. W. (2023). Learnable latent embeddings for joint behavioural and neural analysis [Publisher: Nature Publishing Group]. *Nature*, 1–9. https://doi.org/10.1038/s41586-023-06031-6

Shen, W., Yuan, Y., Liu, C., & Luo, J. (2016). In search of the aha! experience: Elucidating the emotionality of insight problem-solving. *British Journal of Psychology*, *107*(2), 281–298. https://doi.org/10.1111/bjop.12142

Sheth, B. R., Sandkühler, S., & Bhattacharya, J. (2009). Posterior beta and anterior gamma oscillations predict cognitive insight. *Journal of Cognitive Neuroscience*, *21*(7), 1269–1279. https://doi.org/10.1162/jocn.2009.21069

Skaar, Ø. O., & Reber, R. (2020). The phenomenology of aha-experiences [Place: US Publisher: Educational Publishing Foundation]. *Motivation Science*, *6*(1), 49–60. https://doi.org/10.1037/mot0000138

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the big five inventory2: The BFI-2-s and BFI-2-XS. *Journal of Research in Personality*, *68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Sternberg, R. J., & Davidson, J. E. (Eds.). (1995). *The nature of insight*. MIT Press.

Strijbosch, W., Vessel, E. A., Welke, D., Mitas, O., Gelissen, J., & Bastiaansen, M. (2021). On the neuronal dynamics of aesthetic experience: Evidence from electroencephalographic oscillatory dynamics. *Journal of Cognitive Neuroscience*, 1–19. https://doi.org/10.1162/jocn_a_01812

Struk, A. A., Carriere, J. S. A., Cheyne, J. A., & Danckert, J. (2017). A short boredom proneness scale: Development and psychometric properties. *Assessment*, *24*(3), 346–359. https://doi.org/10.1177/1073191115609996

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. https://doi.org/10.1038/nn.2163

Summerfield, C., Wyart, V., Johnen, V. M., & De Gardelle, V. (2011). Human scalp electroencephalography reveals that repetition suppression varies with expectation. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00067

Szubielska, M., Ganczarek, J., Pietras, K., & Stoliska, A. (2021). The impact of ambiguity in the image and title on the liking and understanding of contemporary paintings. *Poetics*, *87*, 101537. https://doi.org/10.1016/j.poetic.2021.101537

Temme, J. E. V. (1992). Amount and kind of information in museums: Its effects on visitors satisfaction and appreciation of art. *Visual Arts Research*, *18*(2), 28–36.

Thielen, J., Bosch, S. E., van Leeuwen, T. M., van Gerven, M. A. J., & van Lier, R. (2019). Evidence for confounding eye movements under attempted fixation and active viewing in cognitive neuroscience. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-54018-z

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology*, *38*(2), 227–242. https://doi.org/10.1177/0022022106297301

Tschacher, W., Greenwood, S., Kirchberg, V., Wintzerith, S., van den Berg, K., & Tröndle, M. (2012). Physiological correlates of aesthetic perception of artworks in a museum. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(1), 96–103. https://doi.org/10.1037/a0023845

Vallat, R. (2018). Pingouin: Statistics in python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

Van de Cruys, S., Damiano, C., Boddez, Y., Król, M., Goetschalckx, L., & Wagemans, J. (2021). Visual affects: Linking curiosity, aha-erlebnis, and memory through information gain. *Cognition*, *212*, 104698. https://doi.org/10.1016/j.cognition.2021.104698

Wang, X., Huang, Y., Ma, Q., & Li, N. (2012). Event-related potential p2 correlates of implicit aesthetic experience: *NeuroReport*, *23*(14), 862–866. https://doi.org/10.1097/WNR.0b013e3283587161

Waskom, M. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Welke, D., & Vessel, E. A. (2022). Naturalistic viewing conditions can increase task engagement and aesthetic preference but have only minimal impact on EEG quality. *NeuroImage*, *256*, 119218. https://doi.org/10.1016/j.neuroimage.2022.119218

Werheid, K., Schacht, A., & Sommer, W. (2007). Facial attractiveness modulates early and late event-related brain potentials. *Biological Psychology*, *76*(1), 100–108. https://doi.org/10.1016/j.biopsycho.2007.06.008

Wirth, M., Horn, H., Koenig, T., Razafimandimby, A., Stein, M., Mueller, T., Federspiel, A., Meier, B., Dierks, T., & Strik, W. (2008). The early context effect reflects activity in the temporo-prefrontal semantic system: Ev-

idence from electrical neuroimaging of abstract and concrete word reading. *NeuroImage*, *42*(1), 423–436. https://doi.org/10.1016/j.neuroimage.2008.03.045

Yao, D. (2001). A method to standardize a reference of scalp EEG recordings to a point at infinity. *Physiological Measurement*, *22*(4), 693–711. https://doi.org/10.1088/0967-3334/22/4/305

Zajonc, R. B. (1968). Attitudinal effects of mere exposure [Place: US Publisher: American Psychological Association]. *Journal of Personality and Social Psychology*, *9*(2), 1–27. https://doi.org/10.1037/h0025848

Zhao, Z., & Zhang, Y. (2018). SQI quality evaluation mechanism of single-lead ECG signal based on simple heuristic fusion and fuzzy comprehensive evaluation. *Frontiers in Physiology*, *9*, 727. https://doi.org/10.3389/fphys.2018.00727

A Supplementary Material

A.1 Participant demographics

See Tabs. A.1 and A.2 for full sample demographics.

Table A.1: Categorical demographic factors

Variable	Exp. 1	Exp. 2
n	56	49***
Sex		
Female	37	28
Male	19	19
Other	–	2
Handedness*		
Right	49	41
Left	3	8
Ambidextrous	3	–
N/A	1	–
Eye dominance*		
Right	7	28
Left	14	17
No dominance	4	4
N/A	31	–
Highest degree of education**		
Realschulabschluss	1	–
Hochschulreife	25	26
Hochschulabschluss	23	23
Promotion	6	–
N/A	1	–
Mental disorder		
No	51	38
Yes	5	11
Neurological disorder		
No	54	49
Yes	2	1***

Note: * in Exp. 1 handedness and eye dominance were collected as self-reports, while in Exp. 2 the FLANDERS questionnaire was used to assess handedness, and the PORTA test was used to assess eye dominance; ** German qualification levels – RS: finished school, HSR: qualified for/currently in higher education, HSA: bachelor/master degree, PR: PhD ; *** in Exp. 2, data from N=50 participants were collected, but 1 participant with neurological disorders was excluded from the sample.

Table A.2: Continuous demographic factors

Variable	n	M	SD	min	max	25%	50%	75%
Exp. 1								
Age	55	40.0	16.97	20	82	25	35	51
Years of education	53	17.4	3.89	7	26	15	18	20
BFI (range 3-15)								
extraversion	56	9.36	2.178	4	15	8	9	11
open mindedness	56	11.52	1.935	8	15	10	12	13
agreeableness	56	11.00	2.132	5	15	9	11	13
conscientiousness	56	10.93	2.213	5	14	9	10.5	12
negative emotionality	56	8.59	2.661	3	14	7	9	11
AReA (range 0-4)								
total	56	2.02	0.692	0.64	3.36	1.57	2.07	2.5
aesthetic appreciation	56	2.60	0.736	1	4	2.25	2.75	4
intense experience	56	1.50	0.831	0	3.5	1	1.5	3.5
creative behavior	56	1.15	0.938	0	3	0.25	1.33	3
BPS (range 8-56)	56	19.71	8.329	8	41	13	17	25
PANAS (range 10-50)								
positive	56	29.38	6.323	17	46	14	18	21
negative	56	12.97	4.311	10	31	6	6	7.5
Exp. 2								
Age	49	28.0	7.17	18	52	24	26	30
Years of education	48	17.3	3.90	8	27	15	17	20
BFI (range 3-15)								
extraversion	49	9.47	2.256	4	13	8	9	11
open mindedness	49	11.96	1.707	7	15	11	12	13
agreeableness	49	10.82	1.997	7	15	9	11	12
conscientiousness	49	9.80	2.491	5	14	8	10	12
negative emotionality	49	9.29	2.458	4	14	8	10	11
AReA (range 0-4)								
total	49	1.93	0.624	0.64	2.93	1.43	2	2.43
aesthetic appreciation	49	2.47	0.619	1	3.62	2	2.5	3
intense experience	49	1.30	0.802	0	2.75	0.75	1.25	1.75
creative behavior	49	1.26	1.037	0	3.33	0.33	1.33	2
BPS (range 8-56)	49	21.86	7.979	9	45	16	21	26
Caffeine intake (in mg/kg)*	48	1.06	0.923	0	3.61	0.34	0.89	1.50
PANAS (range 10-50)								
positive	49	25.80	6.423	12	41	22	26	30
negative	49	12.35	3.345	10	27	10	11	13

Note: * caffeine intake was only assessed in Exp. 2

A.2 Stimulus subselection for EEG study

Besides the main analyses, the rating data from the online study (Exp. 1) were used to preselect the most promising candidate stimuli to be presented in the subsequent EEG study (Exp. 2). Selection was based on the individual strength of the title effect on aha for each image (random effect slope coefficients of an LMM; the details of the procedure are described in the methods). Fig. A.1 shows the ranked slope coefficients - we selected the top 16 images for each stimulus category (figurative art, semi-abstract art, abstract art, Mooney image). While there was a high variance in the slope coefficients across all individual stimuli (with a few even showing *lower* aha ratings when the title was presented) the selected image subsets for Exp. 2 have high average coefficients on a comparable level (Mooney: average coefficient $= 38.6$; abstract art: average coefficient $= 25.7$; semi-abstract art: average coefficient $= 31.9$; figurative art: average coefficient $= 27.2$).

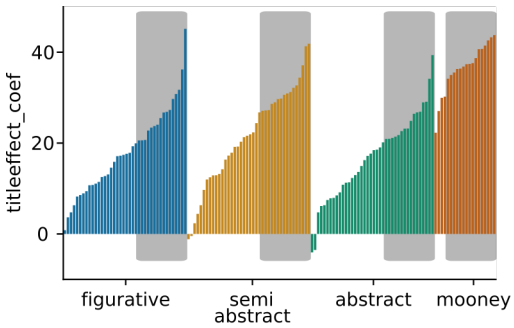


Figure A.1: **Stimulus subselection for EEG study.** gray boxes indicate the selected stimuli (16 per category)

A.3 Correlation of trial wise ratings.

Repeated measures correlation (Fig. A.2) and paired local regression (Fig. A.3) of all collected ratings (data from Exp. 1). Repeated measures correlation (Bakdash & Marusich, 2017) implemented in the pingouin python package (Vallat, 2018, version 0.5.5).

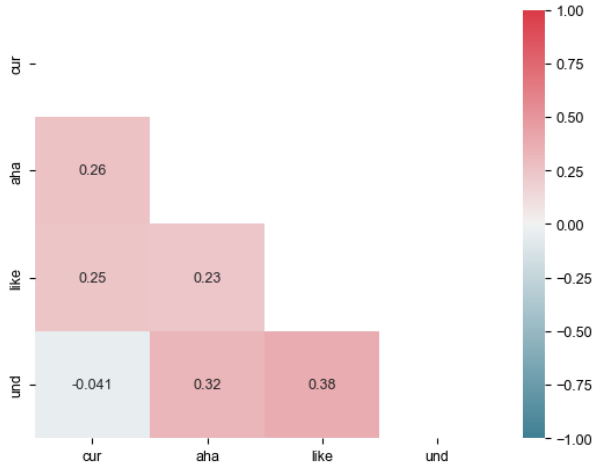


Figure A.2: **Correlation of ratings** Repeated measures correlation; values indicate R; all $p < .05$ (data from Exp. 1)

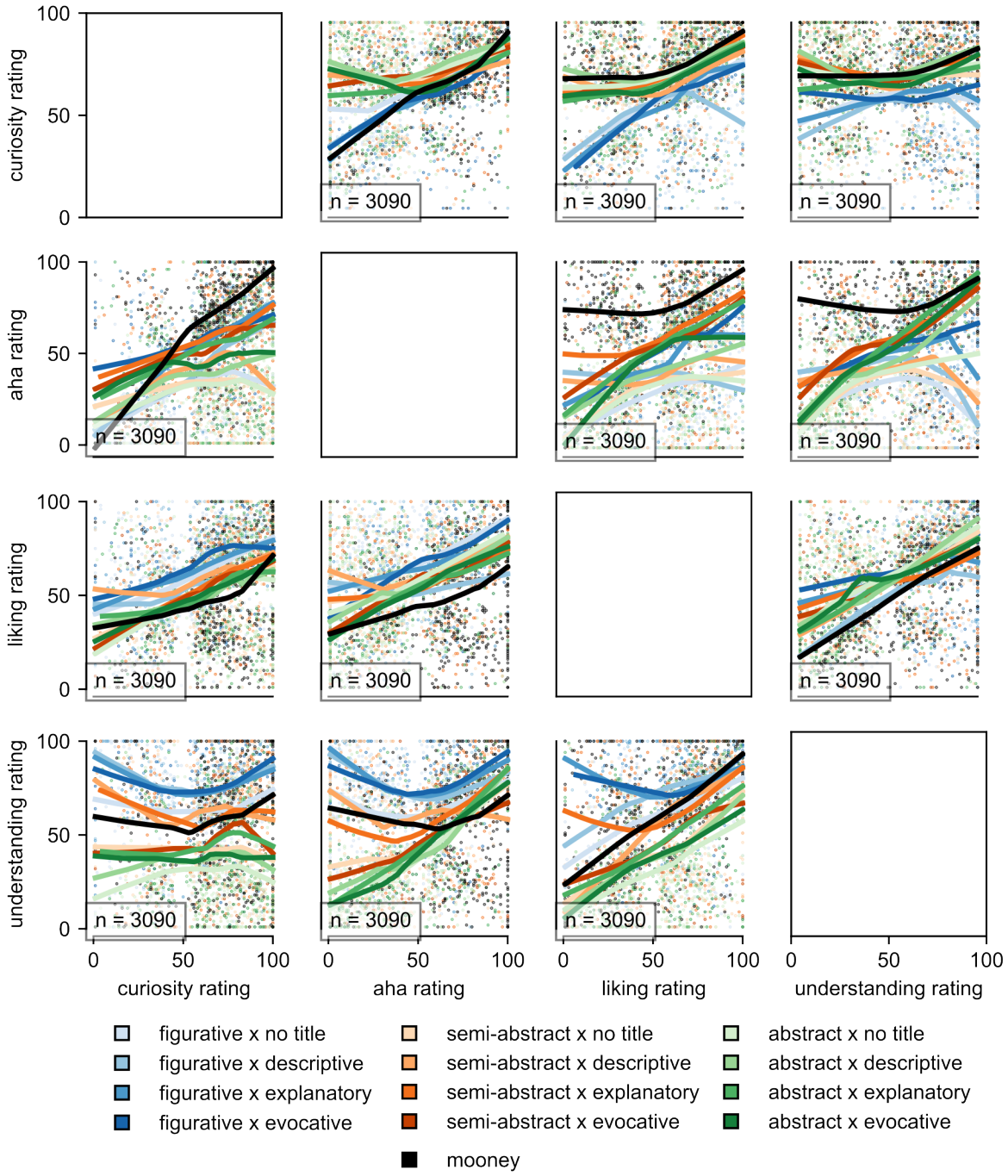


Figure A.3: **(Non)linear relation of ratings.** Scatter plots with local regression (LOWESS) curves grouped by stimulus categories (data from Exp. 1)

A.4 Grand average physiological responses

Below we provide plots of the grand average data traces (only artwork trials, split by 1st or second presentation of each stimulus). While these plots do not map directly to the decoded data features (for these results see Fig. 3a) they might serve to give a general impression of the signal. In the case of ERP we see a prominent frontal/occipital gradient throughout pre and pottitle trials, that degrades over time. In the differential signal (subtraction done on the individual trial level) we see that the most prominent differences are in central vertex channels.

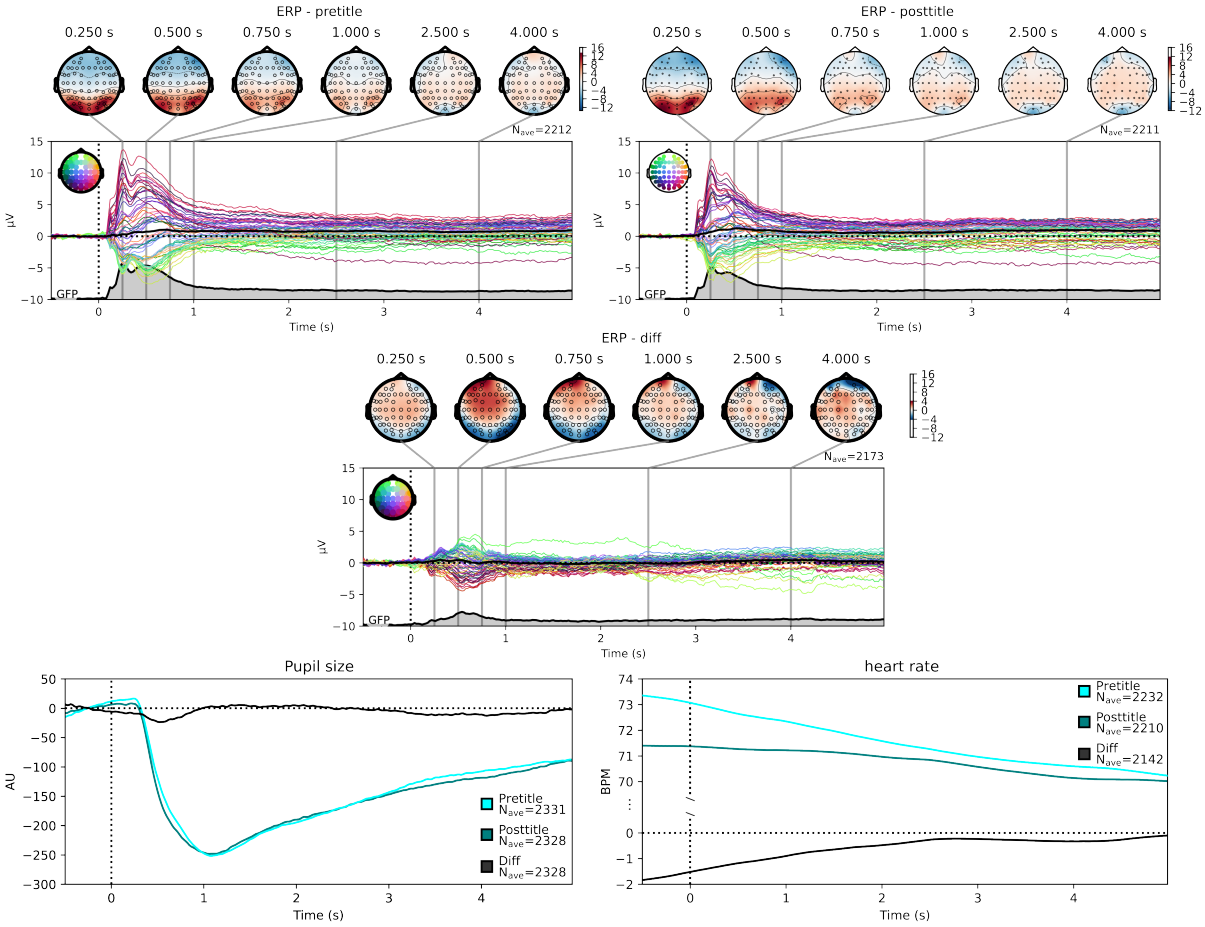


Figure A.4: **Grand average physiological responses in time domain signals – 1st vs 2nd presentation** depicted are grand average responses across all artwork trials, split by trial condition (pretitle – 1st presentation before title, posttitle – 2nd presentation after title, diff – trace during 2nd - trace during 1st presentation). Colored lines in EEG depict individual channels, black line the grand average, and at the bottom is the global field power (GFP). Above the time course are topography plots at 6 representative timepoints (selection not data driven). Pupil size and heart rate have only 2, respectively 1 data channel, hence the cleaner plot.

A similar comparison for the ratings is not straight forward, as these are continuous graded responses.

A.5 Categorical decoding results

The categorical experimental manipulations we applied could be decoded quite robustly from the recorded data. Decoding worked very well for the contrast between the 1st and 2nd presentation of any given image stimulus (see Fig. 3a). Decoding was reliably possible from almost all data modalities. It was strongest from the time-domain EEG (significantly above chance throughout the full trial with a marked peak around 0.5–1 s after image onset, very high accuracy throughout). Time-frequency EEG in all frequency bands allowed above chance decoding during most of the trial duration, though not significant throughout. Alpha band EEG gave overall the best results, while decoding worked better for art stimuli than for Mooney images. There were no clear peaks, but in theta EEG decoding was better at the beginning of trials. The pupil trace allowed for significant above chance decoding throughout the full trial in art trials and all trials combined. When looking at Mooney trials alone, decoding became significant only at the end of the trial, starting around 3.3 s after image onset. The heart rate signal allowed significant decoding throughout the full trial – including the period before image onset – in art trials and all trials combined. When looking at Mooney trials alone, decoding was significant in the early (-0.5–1.33 s) and medium late trial (1.73–3.72 s). Looking at the grand average responses this does not come as a surprise (cf. Supp. Fig. A.4). Taken together this strongly suggests that the stimuli are not processed identically during repeated presentation; rather, cognitive processing differs in a reliably detectable way. Whether this is caused by mere repetition suppression (cf. Kim, 2017) or other factors specific to the study such as the interim contextualization with additional information (i.e. the real or fake title, or the Mooney solution respectively) or the differing implicit or explicit tasks (e.g., generating ratings) can not be clearly determined at this point.

The presented image category (Mooney image or painting) could also be reliably decoded right from the beginning from both EEG and pupil data (see Fig. 3b). In the time domain EEG data it could be clearly decoded during the first image presentation (pretitle: starting 120 ms after image onset), the second image presentation (posttitle: starting 40 ms after image onset). Decoding accuracy was well above chance throughout the full trial, but with a clear and marked peak at the beginning, about 1 s after image onset. Notably, classification worked more reliably during the second image presentation (i.e. higher decoding performances, and more sustained significance); this was the case for most of the investigated decoding tasks in this study. In the frequency domain EEG, findings were mixed. Theta and alpha band power were overall the best predictors; both delivered above chance decoding throughout the trials with several significant patches. Theta delivered good decoding in pretitle trials (significant: 0.56-1.05 s, 2.27-2.61 s, 2.83-3.03 s, 3.82-4.00 s, 4.32-4.56 s, and 4.72-4.94 s) and throughout the first 2 seconds of posttitle trials (significant: 0.26-0.91 s, and 1.11-1.77 s). Alpha was worse in pretitle trials (only significant: 2.79-3.03 s) but quite reliable throughout posttitle trials (significant: 0.26-0.44 s, 1.41-1.77 s, 1.85-2.31 s, 3.27-3.64 s, and 4.28-5.00 s). Beta band power was also a quite reliable predictor in posttitle trials (significant: 0.48-0.93 s, 1.09-1.21 s, 1.35-1.63 s, 1.81-1.99 s, 2.03-2.27 s, 3.03-3.15 s, 3.53-3.74 s, 3.84-3.98 s, 4.42-4.66 s, 4.88-5.00 s), but not in pretitle trials. Delta showed a significant blob in posttitle trials (1.79-2.49 s) but again not in pretitle trials. No significant results were found in the gamma band. In the pupil data the image category could be decoded during pretitle (starting at 300 ms) and posttitle presentation (starting 220 ms after image onset). As with ERPs, there was a clear peak at the beginning, around 1 s after image onset, then decreasing but staying well above chance for the rest of the trial. Interestingly, there was no big difference between decoding traces in pretitle and posttitle trials. Heart rate only predicted image category in the second presentation (significant from -0.5–0.44 s and from 1.59 s onward with a short non-significant intermission) but not in the pretitle condition. Besides the early peak in the decoding rate from pupil data (see discussion), single trial ERPs yielded the best results.

A.6 Decoding using a differential physiological signal

We repeated the decoding pipeline on the differential physiological responses (posttitle–pretitle traces). See Fig. Supp. A.8 for all decoding time-courses.

From this differential signal we could significantly decode the presented image category (Mooney image or painting) throughout the trial (starting 160 ms after image onset) in the time-domain EEG, and there were two short significant clusters in EEG alpha power (0.80-1.07 s and 4.24-4.44 s).

Title condition: time domain EEG - differential: significant 0.42–2.63 s after image onset

Insight: late differential pupil signal (2nd – 1st presentation; becoming significant from 4.18 s after image onset until the end of trial) There was further a short significant cluster in the differential theta-band power (2.83-2.99 s after image onset).

A.7 Full decoding results

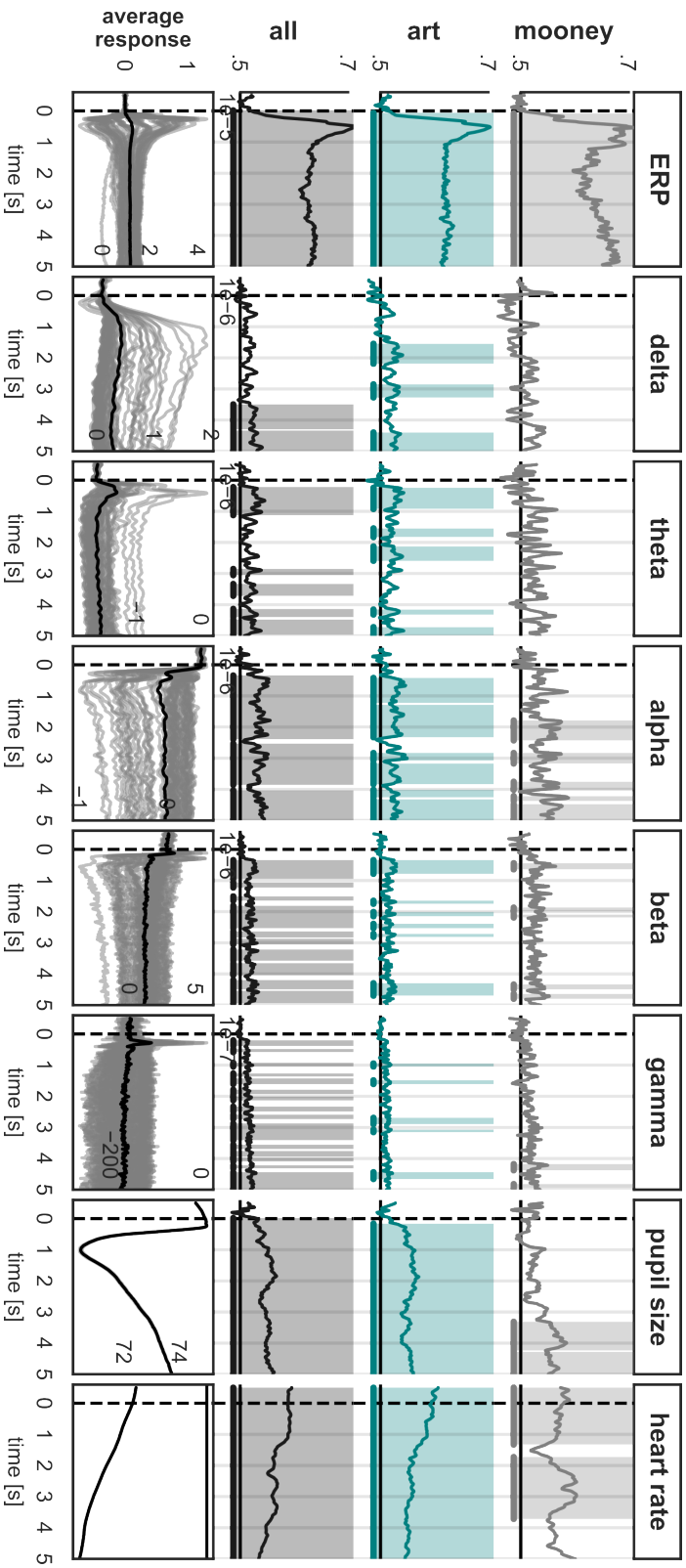


Figure A.5: **Full decoding results for 1st vs 2nd presentation of images** decoding rate over time, significant timewindows are marked by color underlay (significance level $p < .05$ based on second level cluster permutation T-test).

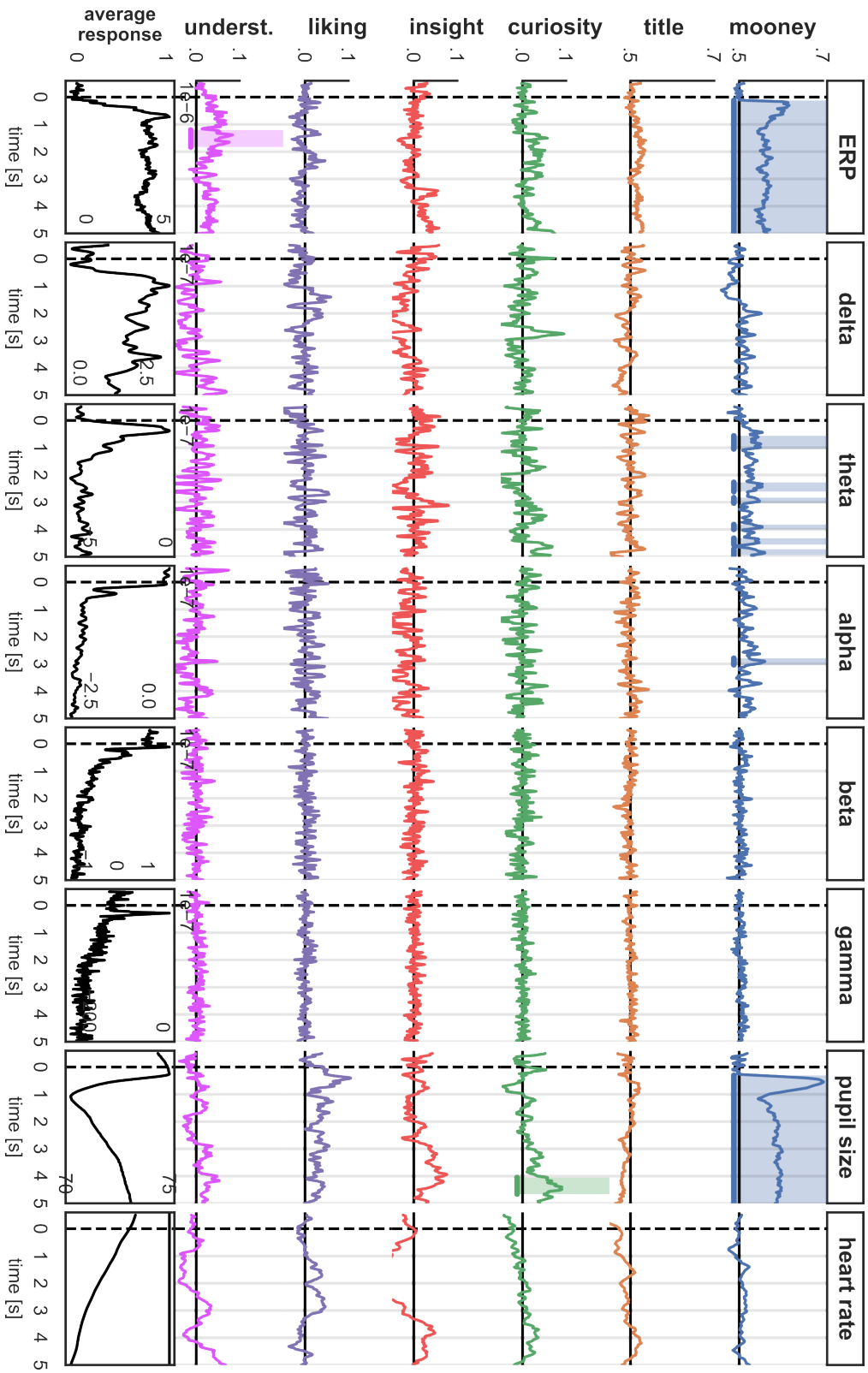


Figure A.6: **Full decoding results based on pretitle signal (1st presentation of images)** decoding rate over time, significant timewindows are marked by color underlay (significance level $p < .05$ based on second level cluster permutation T-test).

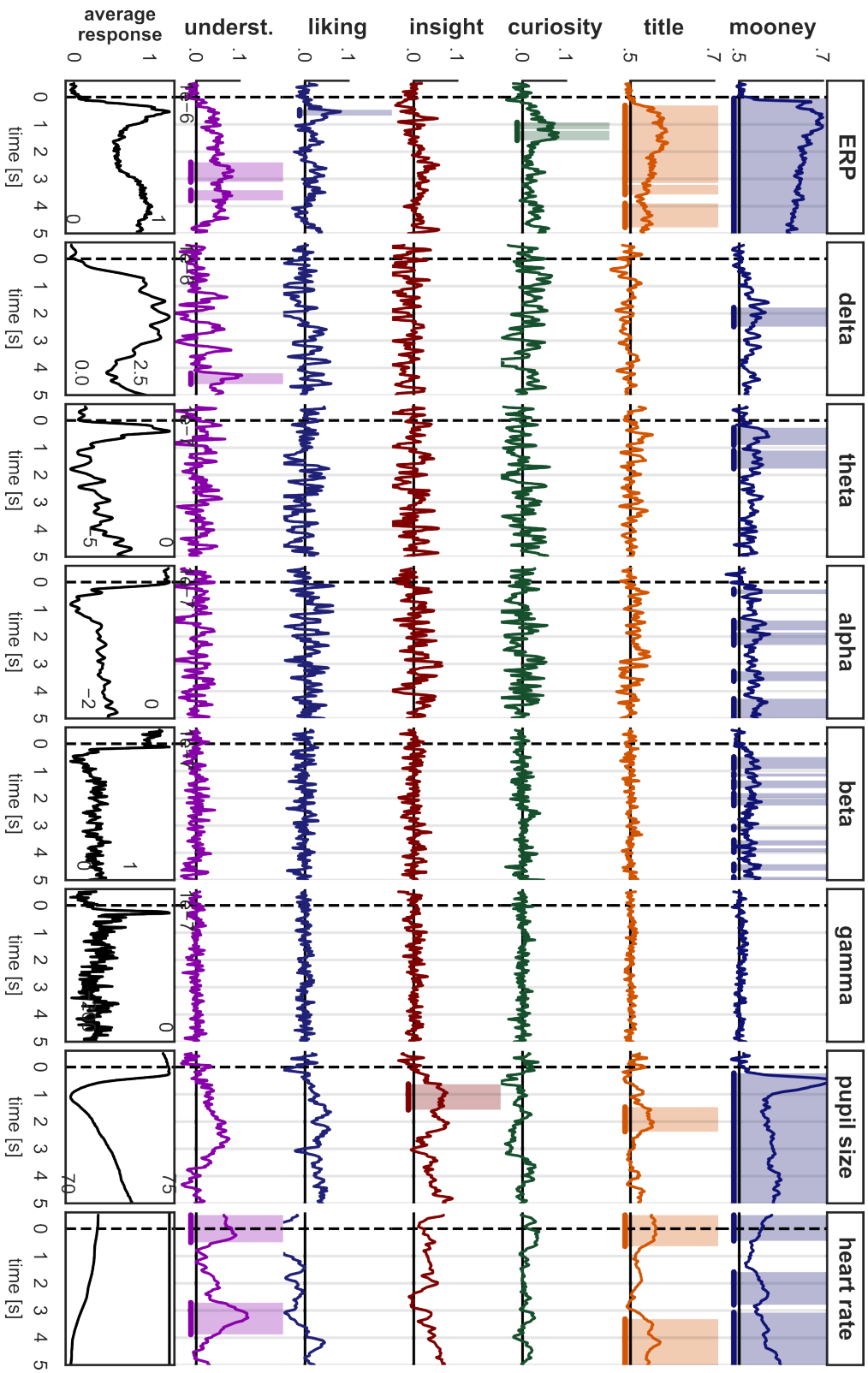


Figure A.7: **Full decoding results based on posttitle signal (2nd presentation of images)** decoding rate over time, significant timewindows are marked by color underlay (significance level $p < .05$ based on second level cluster permutation T-test).

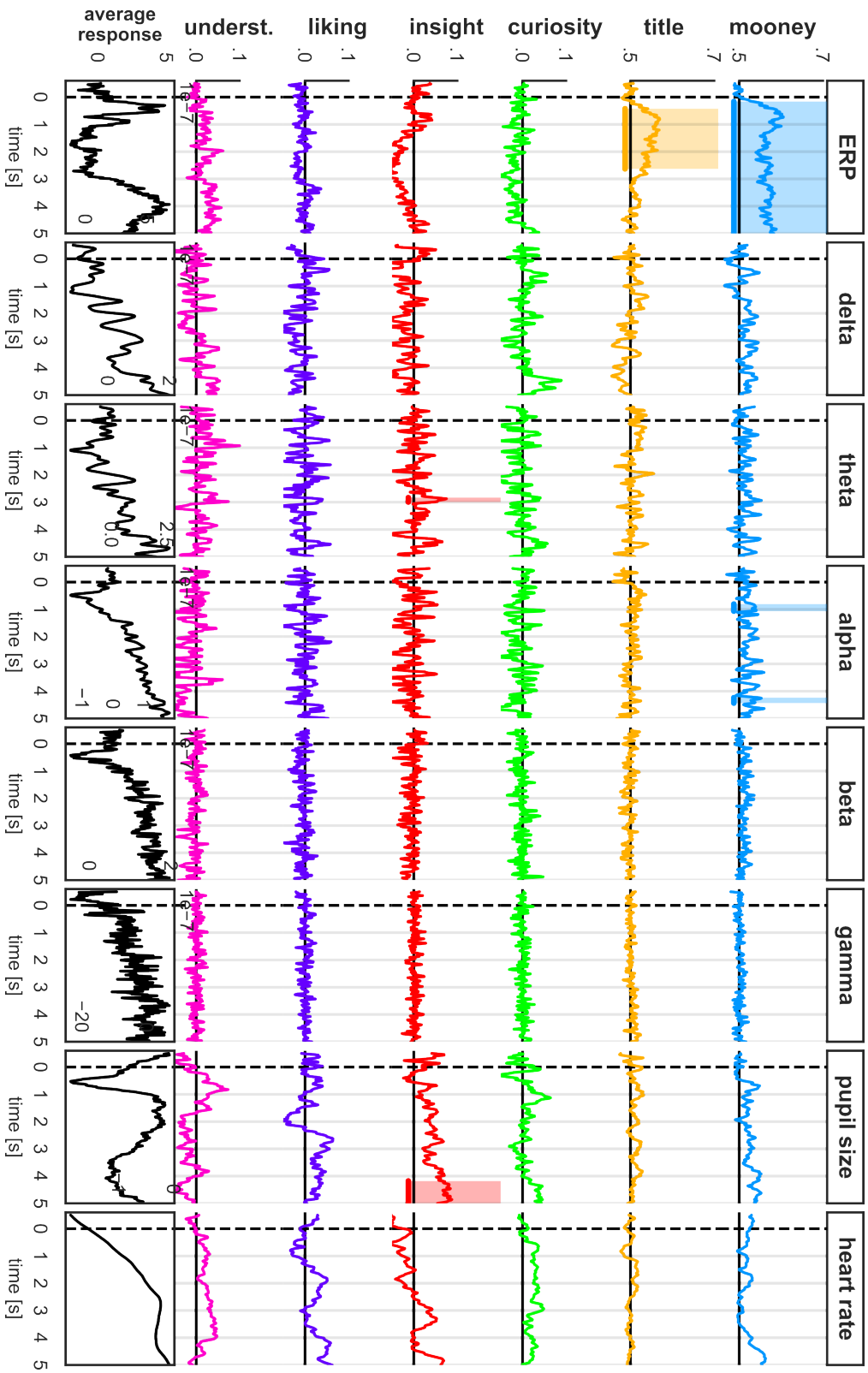


Figure A.8: **Full decoding results based on differential signal (2nd presentation – 1st presentation)** decoding rate over time, significant timewindows are marked by color underlay (significance level $p < .05$ based on second level cluster permutation T-test).

A.8 Artwork stimuli

See below depictions of the used art stimuli; the full stimulus set can be downloaded in the OSF online repository under https://osf.io/dafyv/ and https://osf.io/s7jtx/.

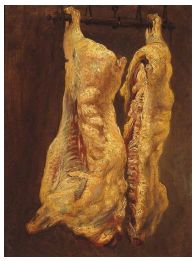
Exp. 1: Descriptive title, Figurative art

Floesse

Blume mit Vogel und Echsenkopf

Stillleben

Muscheln und Meeresfruechte



Rindfleisch

Landschaft mit Bachlauf, Bruecke, Baeumen und zwei Personen

Sonnenuntergang an der Kueste

Die Windmuehle

Kuehe im Fluss

Stillleben mit Kuerbis, Pflaumen, Kirschen, Feigen und Krug

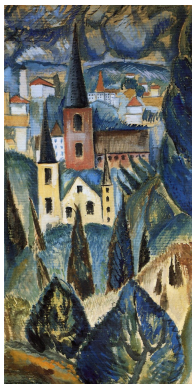
Exp. 1: Descriptive title, Semi-abstract art

Blauer Berg

Landschaft mit Sonnenblumen

Komposition mit Vieux-Strasbourg Vase

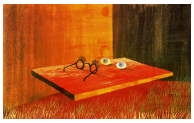
Komposition mit 7 Roehren



Landschaft mit Kirchtuermen und Baeumen

Kiste mit Klavier und Stier

Landschaft mit Birkenstaemmen



Augen auf dem Tisch

Das Ei

Strahlender Sonnenuntergang

Exp. 1: Descriptive title, Abstract art

Komposition

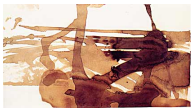
Komposition

Rot, Blau

Blaues Feld

Komposition

Zell-Komposition mit Rotem Kreis



Abstrakte Komposition

Komposition I

Kubistische Komposition

Komposition in Orange, Schwarz und Weiss

Exp. 1: Explanatory title, Figurative art

O'Neills Landeplatz

Der Hafen von Kopenhagen

Der Ausbruch des Vesuv

Kap und Duenen in Saint-Quentin (Frankreich)

Die Schlacht von Krasnaja Gorka

Die Moseley Regatta

Sonnenuntergang in Ivry

Die Muehle in Flatford, von einer Schleuse am Fluss Stour

Sentry-box Reach, Hawkesbury River, New South Wales

Im Giji Park

Landschaft mit Baeumen an der Themse in Calcot, Berkshire

Unruhige See

Moorkanal mit Torf-Barken

Der Leuchtturm von Marseille

Waldbraende

Cervantes

Die Geburt der Venus

Ein Schiff im Hafen bei Ebbe

Die Salinieres von Trouville

Markt in der Normandie

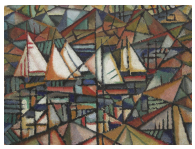
Exp. 1: Explanatory title, Semi-abstract art

Insignien

Propellers 2nd state (zweiter Zustand des Propellers)

New York 1 (Road #4)

Taenzer in einem Cafe



Boote

Helgoland

Bronze Ballett

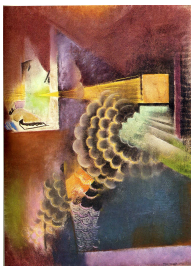
Fisch Zirkus

Hemlock Tanne

Kuhstall

Die Stadt

Hafenszene



Bildvortrag im Metropolitan Museum

Kriegsschiffe

Der Esstisch

Das bestellte Feld

Figuren am Strand

Moskau I

Russisches Ballett

Penthesilea

Exp. 1: Explanatory title, Abstract art

Figur

Pasiphae

Haus am Strand

Wuestenlandschaft

Stone Flower (Schues-selflechte)

Der Zauberkuenstler

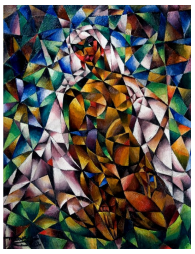
Berge und Meer

Fluctuating Planes (Schwankende Ebenen)

Football #3

Akt (Nackte)

Der Tisch

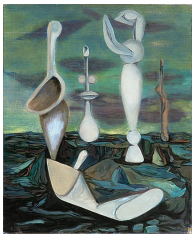


Akt mit Schleier

Sandwich

Das Riff

Rote Bluetenblaetter



Die Leuchttuerme

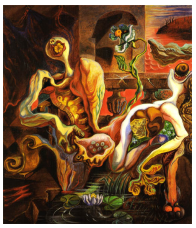
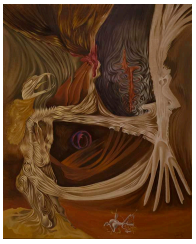
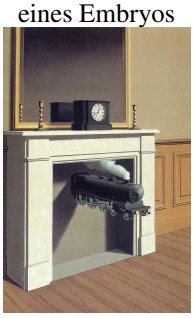
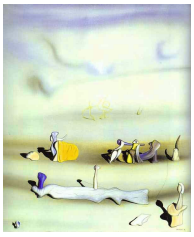
Hoffmannsche Szene

Schneckenspur

Der Fallschirm-springer

Die schwarze Stadt I (New York)

Exp. 1: Evocative title, Figurative art				
Fruehling. Die Kraehen sind angekommen	Der Kraehenbaum	Impression, Sonnenaufgang	Vor dem Sturm	Der Ursprung der Welt
Nach dem Regen	Sonntagmorgen	Die Strasse des Blutes, Delhi	Windiger Tag	Lange Schatten - Vieh auf der Insel Saltholm

Exp. 1: Evocative title, Semi-abstract art				
 Metamorphose der Liebenden	Wunderkammer	Die Autobiographie eines Embryos	Bleifarbenes Meer	Das Signal
 Der Keim der Nacht	Mein Zimmer hat zwei Tueren	 Durchbohrte Zeit	Morphologie	 Morgen

Exp. 1: Evocative title, Abstract art

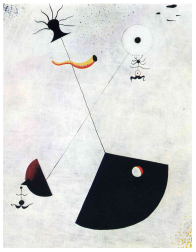
The Liver is the Cock's Comb (Die Leber ist des Hahnen Kamm)

Etendue (Ausdehnung)

Die Woge

Futurismo

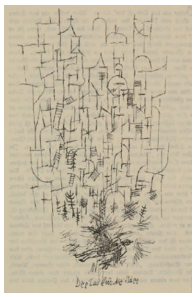
Der prachtvolle Vogel offenbart einem Liebespaar das Unbekannte



Mutterschaft

Tanabata

Rush Hour



Der Tod fuer die Idee

Boundsplayedpause (Grenz-Spiel-Pause)

Exp. 2: Figurative art

Die Strasse des Blutes, Delhi

Vor dem Sturm

Der Ausbruch des Vesuv

Der Leuchtturm von Marseille

Die Geburt der Venus

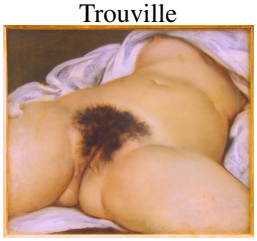
Blume mit Vogel und Echsenkopf

Die Salinieres von Trouville

Frühling. Die Krähen sind angekommen

Der Krähenbaum

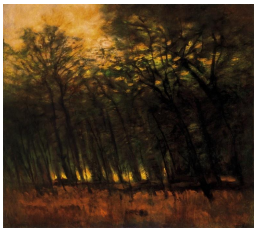
O'Neills Landeplatz



Der Ursprung der Welt

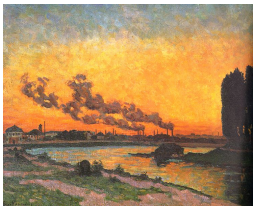
Moorkanal mit Torf-Barken

Die Schlacht von Krasnaja Gorka



Waldbrände

Nach dem Regen



Sonnenuntergang in Ivry

Exp. 2: Semi-abstract art

Tänzer in einem Cafe

Landschaft mit Sonnenblumen

Insignien

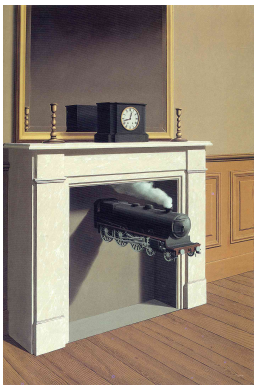
Kuhstall

Die Stadt

Hemlock Tanne

Metamorphose der Liebenden

Figuren am Strand



Durchbohrte Zeit

Das Signal

Moskau I

Kiste mit Klavier und Stier

Die Autobiographie eines Embryos

Helgoland

Das bestellte Feld

Kriegsschiffe

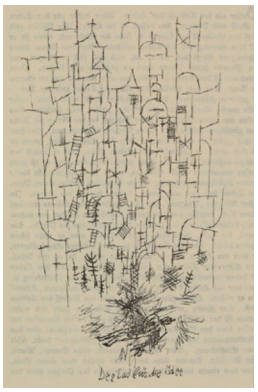
Exp. 2: Abstract art

Komposition in Orange, Schwarz und Weiss

Rote Blütenblätter

Das Riff

Fluctuating Planes (Schwankende Ebenen)



Der Tod für die Idee

Der Zauberkünstler

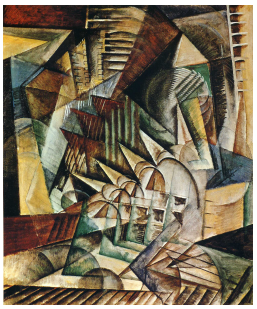
Die Leuchttürme

Der Tisch

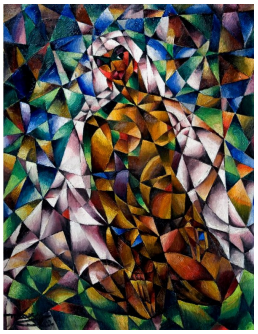
Haus am Strand

Wüstenlandschaft

Die Woge



Rush Hour



Akt mit Schleier

Die schwarze Stadt I (New York)

Der Fallschirmspringer

Blaues Feld