EXAMINING THE EFFECTS OF FREE GAZE AND DYNAMIC VIDEO STIMULI ON ENGAGEMENT, EYE MOVEMENTS, AND EEG SIGNAL QUALITY IN A VISUAL AESTHETIC RATING TASK

A PREPRINT

Dominik Welke
Department of Neuroscience
Max Planck Institute for Empirical Aesthetics
Frankfurt (Main), Germany
dominik.welke@ae.mpg.de

Edward A. Vessel
Department of Neuroscience
Max Planck Institute for Empirical Aesthetics
Frankfurt (Main), Germany
ed.vessel@ae.mpg.de

September 18, 2021

ABSTRACT

Free gaze and the use of dynamically changing video stimuli are typically avoided in EEG experiments to avoid artefacts and confounds related to uncontrolled eye movements. Yet, often it is unclear whether these artificial secondary manipulations might have unwanted effects on the primary measures of interest and for a growing number of research questions removing them would be beneficial: Among those is the investigation of visual aesthetic experiences, which typically involve open-ended exploration of highly variable stimuli. Here we aimed to quantify the effect of fixation task and using still vs. movie stimuli on EEG signal quality and several behavioral and physiological measures of interest during an aesthetic rating task. Participants observed scenes from landscapes and dance performances and rated each stimulus for both aesthetic appeal and their state of boredom while watching it. The scenes were presented either as dynamic video clips or static pictures, and participants observed them either with unconstrained gaze or under attempted fixation We recorded EEG, ECG and eyetracking from 43 participants. An auditory stream of 40Hz amplitude modulated pink noise was played during each trial and signal-to-noise-ratio (SNR) of the auditory steady-state response measured at the scalp was extracted as a proxy measure for overall EEG signal quality. The study including hypotheses and a priori power analysis was preregistered. We found that both behavioral ratings were influenced by the experimental conditions: boredom and aesthetic ratings were positively affected by dynamic video stimuli, indicating that these are experienced as more engaging; both these effects were stronger in dance. As already reported before, landscape stimuli were experienced as more appealing. Fixaton task, on the other hand, had no significant effect on the ratings which is encouraging given how canonically it is applied. Eye movements were significantly affected not only by viewing task, but by stimulus dynamics and content as well: we observed fewer eyeblinks, saccades and microsaccades in video stimuli, and fewer saccades but more microsaccades in dance than in landscape stimuli, with several significant interactions. EEG SNR, to our surprise, was barely affected by fixation task - despite only minimal preprocessing and no trial rejection. We nevertheless believe that the new metric is sensitive to capture noise: it was significantly correlated with the number of eye blinks, and after cleaning the dataset with an ICA based preprocessing pipeline the significant effect of fixation task and the correlation with blink rate vanished. We see these as promising results indicating that at least in the lab more liberal experimental conditions could be achieved without significant loss of signal quality. Specifically the use of dynamic video material bears a lot of potential for future investigations in human neurophysiological studies.

Keywords experimental trade-offs · task-engagement · video stimuli · neuroaesthetics

1 Introduction

Empirical research often requires a careful trade-off between experimental control and ecological validity. Perhaps nowhere is this balancing act more apparent than when investigating the electrophysiological correlates of higher cognitive functions. Many efforts are taken to design rigorously controlled stimuli that do not introduce potential confounds, and to avoid sources of exogenous and endogenous measurement noise that might obscure the neuronal signals of interest. This has led to the development of a set of experimental constraints that are applied almost canonically in human neurophysiology, particularly in the visual domain. Most prominent amongst these is the nearly ubiquitous fixation task, implemented to reduce the effects of eye movements in electroencephalographic (EEG) signals. Almost as common is the use of static images rather than video or otherwise dynamic visual material. Whereas such stimuli have become much more common in the functional magnetic resonance imaging (fMRI) literature (e.g. Hasson, 2004; Hasson and Honey, 2012; Vodrahalli et al., 2018; Isik and Vessel, 2021), the adoption of video stimuli in electroencephalography and magnetoencephalography (MEG) studies lags far behind.

Here we seek to guide future research by systematically investigating the effects of relaxing these experimental constraints in an EEG study of visual aesthetic preference. We present participants with either static pictures or dynamic movie stimuli, viewed either under attempted fixation or with free gaze. We then quantify the effects of these two canonical experimental constraints on task relevant behavior, engagement and EEG recording quality. While there can be very good reasons to apply a fixation task or refrain from presenting dynamic visual stimuli, the decision to do so is often a mere default based on common practice, even when it is unclear whether these restrictions might affect the behavior of interest.

1.1 Relaxing experimental constraints to study naturalistic behaviors

Allowing participants to actively explore stimuli, and adopting the use of time varying visual stimuli in EEG studies would be beneficial for a growing number of research questions related to naturalistic behavior.

In particular, the neuroscientific study of aesthetic experiences would benefit from greater flexibility in experimental design. While fMRI research has identified a set of brain areas and networks involved in aesthetic processing (Vessel, 2020; Chatterjee and Vartanian, 2014), the electrophysiological correlates of aesthetic processing are to date poorly understood. This is unfortunate because the superior temporal precision of EEG and MEG is well suited to investigate the fast neural dynamics of aesthetic processing, and the relative mobility of EEG devices could even allow for investigations of aesthetic encounters *in situ*. Yet visual aesthetic experiences in the real world often involve open-ended exploration of highly complex artistic or natural objects. For example, many works of visual art are too large to be captured in one glance and rely on eye, head, or even body movements for active exploration. Other art forms present the observer with moving images or inherently rely on temporal information (e.g. performances, dance, film and time-based media). In addition, the context in which art is encountered matters (see Tinio et al., 2013), and sterile laboratory environments and rigorous experimental constraints might lower the likelihood of strongly moving aesthetic experiences (Brieber et al., 2014, 2015), thereby negatively impacting statistical power. Investigating the electrophysiological correlates of moving aesthetic experiences would thus benefit from relaxing typical constraints on EEG paradigms.

Beyond empirical aesthetics, such constraints also present problems for other research domains. Static visual stimuli are clearly insufficient for studying other time- or movement-dependent psychological processes, such as motion perception, human movement observation, error detection tasks, and visual narrative. Enforcing fixation is problematic for mental phenomena in which active visual exploration of a stimulus is required, such as visual search, scene perception, the study of reference frames, visual coordination, wayfinding, and decision paradigms that involve foraging. While there are approaches that may allow one to investigate components of these processes within more rigorous frameworks (e.g. using sequential presentation of static pictures), such paradigms remain an approximation and might distort key aspects of the processes of interest. As the fixation task is mostly specific to EEG and MEG studies, it furthermore hinders straight-forward comparison with research done using fMRI, fNIRS or purely behavioral methods. Hence, the *a priori* prohibition of the use of moving images and free gaze limits the questions that can be tackled using MEG and EEG, especially in the context of spontaneous or naturalistic behavior.

Some research areas have already begun developing more liberal experimental paradigms for EEG, though wider adoption of these methods has been limited. In visual neuroscience, the fixation task has been at least partly abandoned when studying scene exploration and visual search (Kamienkowski et al., 2012; Dias et al., 2013; Kaunitz et al., 2014) or reading (Dimigen et al., 2011). Video, moving dot, or other dynamic stimuli are used by researchers interested in motion observation and embodiment (e.g. Heimann et al., 2014), or in hyperscanning studies investigating correlation of neuronal activity between participants that see the same stimuli (Dmochowski et al., 2014; Dikker et al., 2017). There are also a growing number of mobile EEG, MEG, and mobile brain body imaging (MoBI, see Makeig et al., 2009; Gramann et al., 2014) studies in which people's neural responses are recorded while walking (Gwin et al., 2011), navigating complex environments (Debener et al., 2012) or during social interaction (Dikker et al., 2017, 2021). To address the potential loss of data quality, such studies usually try to counteract the introduction of noise by using better recording hardware, by employing advanced techniques to reconstruct or reject noisy data, or by developing entirely new analysis approaches to target features of the EEG data that are less susceptible to specific sources of noise (e.g. saccade related potentials (Ossandon et al., 2010; Ehinger and Dimigen, 2019) for free viewing, or inter-subject correlation (ISC Dmochowski et al., 2012; Ayrolles et al., 2021) or representational similarity analysis (RSA Kriegeskorte, 2008) for less controlled stimuli).

To accelerate the acceptance of new paradigms, we sought to explicitly assess the effect of fixation task and still vs. dynamic video material, as very few such datasets exist. We asked participants to view scenes from landscapes and dance performances and to rate each stimulus for both aesthetic appeal and their state of boredom while watching it. The scenes were presented either as dynamic video clips or static pictures, and participants observed them either with unconstrained gaze or under attempted fixation in a fully-crossed, within-subject factorial design. Rather than tracking specific correlates of a given cognitive process, we focused on EEG data quality as a prerequisite to being able to detect such correlates. We expected EEG signal quality to be impacted by the viewing task and by motion modality (still vs. video). At the same time, relaxing these constraints might benefit behavioral or physiological responses of interest and we sought to quantify the magnitude of these effects and weigh them against each other. Therefore, we also investigated the effects of fixation task and still vs. dynamic video material on aesthetic appeal, task engagement, eye movements and heart rate – measures that are of interest in the study of aesthetic preference and higher cognition in general, but that likely also impact each other in important ways (see below). The content condition (dance or landscape stimuli) functioned both as a control (as the expected rating behavior was well characterized in earlier studies, see below), and to assess the generalizability of any observed effects on EEG data quality.

1.2 Quantifying EEG data quality

How can we measure the effects of experimental constraints on EEG data quality? Distortions in EEG recordings can be categorized by their origin into either physiological (endogenous) noise such as eye movements, heart beats, muscle activity (e.g. swallowing, chewing), and task-irrelevant brain activity, or environmental (exogenous) noise such as line-noise, touch/shock on the sensors, electrode impedance, and issues with broken sensors or cables. The standard approach is to detect and remove noisy segments or components from the data. EEG preprocessing routines identify broken or noisy EEG channels by employing metrics such as the overall signal amplitude or the correlation with neighboring channels (Luck, 2014), and more specific sources of noise are typically addressed by a researcher's individual processing pipeline. Visual inspection remains a common way to identify bad data segments that can be very sensitive if done by experts, but this procedure is time-consuming, has low reproducibility, and can introduce (unintended) bias (Rosenthal, 1966). Many advanced techniques have been introduced, including independent components analysis (ICA), canonical correlation analysis, (CCA), artefact subspace reconstruction (ASR), or regression with data from additional modalities (e.g. EOG, EMG, accelerometer), and fully automatic cleaning routines have been proposed to tackle the problem of reproducibility and experimenter bias (e.g. Bigdely-Shamlo et al., 2015; Pedroni et al., 2019).[1]

In the present study, however, we need a more versatile quantitative measure of EEG recording quality that can change over time, thereby allowing for a *within-subject* comparison of different trials or segments of data in a continuous EEG recording. The desired metric has to be sensitive to the typical sources of endogenous and exogenous noise in EEG recordings described above. While such a quantitative measure of EEG quality would arguably be a very practical tool, there is a striking shortage of proposed metrics in the literature and the ones established usually only target an expected evoked response rather than artefact pollution more generally (Luck, 2014; Luck et al., 2021; Picton, 2011; Wong and Bickford, 1980).

Here we utilize the auditory steady-state response (ASSR) as a proxy measure for overall EEG recording quality. We chose an auditory process to minimize interference with the primary visual task. ASSR is an automatic neuronal response to a periodically modulated auditory stream (Stapells et al., 1984). The ASSR is generated throughout the auditory system with contribution from both brainstem and cortical regions and can be elicited by a broad range of auditory stimuli such as click trains, beats, tone bursts, amplitude modulated sine waves or noise, and frequency modulated signals (Picton et al., 2003). The brain tracks the auditory signal and its response can be quantified from the frequency domain EEG signal as peaks at the stimulation frequency and its corresponding harmonics (Picton et al., 2003; Norcia et al., 2015; Meigen and Bach, 1999). ASSR is stable over time (Van Eeckhoutte et al., 2018) and has been shown to not interfere with visual processing (Keitel et al., 2013) In this study an ASSR-eliciting auditory stream

[1]Note that algorithm based data cleaning can leave residuals in the data (see Robbins et al., 2020; Dimigen, 2020, for ICA), which can be especially problematic if subsequent analyses are based on machine learning approaches (Quax et al., 2019; Thielen et al., 2019).

accompanied all trials as a passive background manipulation. The stimulus was optimized to be as unobtrusive as possible, with a relatively low volume adapted to the participants individual hearing threshold. It is important to note that there is a difference between the brain response (ASSR *sensu strictu*) and the ASSR signal that is measured at the scalp. If the EEG recording quality drops while an auditory ASSR stream is presented to a listener (i.e. if the ASSR is overlayed by endogenous or exogenous noise), we expect the strength of the continuously measured ASSR signal to decrease. In order to exploit this for our metric, the EEG data collected in this study were only minimally preprocessed and were not cleaned to remove noisy segments.

1.3 Measuring aesthetic appeal, engagement and physiological responses under varying task conditions

Our primary behavioral measure of interest is rated aesthetic appeal. It appears reasonable to explore the potential effects of viewing restrictions on aesthetic valuation as previous research has shown that the presentation context matters (e.g. Brieber et al., 2014) and we think that this might be partly due to the observers' ability to freely explore the stimuli, as gaze patterns and viewing time can be linked to aesthetic valuation (Mitrovic et al., 2020). To our knowledge, we are the first to investigate potential effects of dynamic video material compared to static stimuli. We also expect higher aesthetic ratings for nature stimuli compared to dance videos. While aesthetic preferences for specific stimuli can be highly variable across observers (Vessel et al., 2018, 2012), many studies have shown an overall effect of the content domain on aesthetic ratings, with higher average ratings and larger agreement in natural environment, landscape scenes, and faces than in other categories (e.g. Vessel and Rubin, 2010; Vessel et al., 2018). We operationalize our measure as *aesthetic* appeal, but very similar ratings are also collected in other research on preference and reward (e.g. Lopez-Persem et al., 2020). Hence we believe that the results will also apply to more general measures of preference, liking or sensory pleasure.

We sample the participants' state of boredom in order to operationalize task engagement vs. disengagement. Both, disengagement and states of boredom have been consistently related to task-unrelated thought, and fluctuations of attention (Smallwood et al., 2004; Raffaelli et al., 2018). We expect lower boredom ratings in the free-viewing task compared to fixation task, and a moderate negative correlation between the boredom and aesthetic preference rating. To our knowledge, no study to date has collected both aesthetic and boredom ratings at the same time and hence the interplay of these two behavioral ratings is of particular interest to us. Periods of spontaneous task unrelated thought, e.g. mind wandering, have been shown to involve the default-mode network (DMN) (Fox et al., 2015), and the few available fMRI studies on boredom also consistently reported activation of this network (Raffaelli et al., 2018). Interestingly, the DMN has also been implicated in aesthetic processing (Vessel et al., 2012, 2019). Low engagement can also impact behavioral data quality (e.g. false responses or reaction times McVay and Kane, 2012) and EEG recordings. Decreased task engagement marked by periods of mind wandering was shown to negatively affect the strength of ERP components (Smallwood et al., 2008; Kam et al., 2011) and is marked by time-frequency components in the alpha and theta band (Braboszcz and Delorme, 2011; Kam et al., 2021). Further, disengagement with the task at hand might also lead to increased physical restlessness or higher eye blink frequency (thereby introducing artefacts) and increased fatigue.

We recorded eye tracking data both to test whether participants maintained fixation during the fixation task, and also to explore whether any of the other experimental conditions significantly impacted the observers' eye movements. In EEG research, eye movements are mainly regarded as problematic artefacts (Iwasaki et al., 2005) that should be avoided or suppressed. Eyeblinks and larger saccades can be identified in the time domain EEG and would most likely lead to a rejection or reconstruction of data by any of the cleaning routines described above. Smaller fixational eye movements, also called microsaccades, cannot be easily identified from the EEG but might be problematic as well, since they have been shown to induce transient gamma-band responses in the EEG (Yuval-Greenberg et al., 2008), that could be misinterpreted as brain activity. While a fixation task significantly reduces participants' saccade counts (Otero-Millan et al., 2008), this mainly inhibits voluntary saccades and blinks – microsaccades, and their potential negative effects on EEG signal, also occur under attempted fixation (Thielen et al., 2019; Otero-Millan et al., 2008).

However, eye movements must not be conceptualized as "noise" alone. In fact, they are meaningful goal directed behavior that controls the visual input stream to the observers brain. Furthermore, previous research indicates that eye movement patterns may carry relevant information in the context of boredom, engagement, and aesthetic processing: viewing times and fixation heat maps are frequently used in behavioral work on empirical aesthetics (e.g. Mitrovic et al., 2020), microsaccades can carry information about internal states during music listening (Fink et al., 2019), and average fixation duration has even been proposed as a general marker for engagement and external focus during visual tasks (Ramos Gameiro et al., 2017).

Heart rate (HR) was included in our study on an exploratory basis. Changes in HR have been related to boredom and engagement (Raffaelli et al., 2018), eye movements (Liu et al., 2020), and are generally investigated when interested in contributions of the autonomous nervous system (ANS), especially in the context of emotional processing (Winton et al., 1984; Vrana et al., 1988; Patrick et al., 1993; Palomba et al., 1997). Also eye movements have been linked to HR and other ANS responses (Liu et al., 2020). Boredom can cause increased heart rate, decreased skin conductance and increased cortisol level, hinting at high arousal and difficulties in sustaining attention (Merrifield and Danckert, 2014), slowing of the heart beat has been associated with higher engagement (i.e. increased focus and concentration) (Coles, 1972). Heart rate might thus be a promising measure in both directions of a hypothesized boredom-engagement continuum. Heartbeats are also relevant in the context of EEG quality: heartbeat evoked potentials (HEP) (Schandry et al., 1986) are not typically controlled for in EEG cleaning routines (if not captured by ICA), and hence contribute to the measured and analysed signal.

2 Methods

The experimental design and our hypotheses were preregistered (see `https://osf.io/bkep4`). Any detail of the final methods that differed from the preregistered plan is clearly described and discussed as such. Information regarding Participants and Recording device are reported following the standards proposed for M/EEG studies by the Committee on Best Practice in Data Analysis and Sharing (COBIDAS) of the Organization for Human Brain Mapping (OHBM) (Pernet et al., 2018, 2020). Data were collected from Oct. 2019 through Jan. 2020.

2.1 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). Calculation for a fixed effects ANOVA with 8 groups, and a numerator df = 1 (all main effects and possible interactions individually) for an effect size of Cohen's $f = 0.5$, alpha = .05, and power = .8 resulted in an target sample size of 34 participants. Due to expected dropouts we planned to collect data from at least 40 participants. During the preparation of the manuscript we realized, that we had erroneously computed the statistical power for an across-subject rather than a within-subject design as fit for the collected data (we wrongly chose "ANOVA: Fixed effects, special, main effects and interactions"). This lead to a significant overestimation of the required sample size: with the correct design our final sample offers sensitivity to detect effect sizes of Cohen's $f > 0.021$ with a critical $F = 2, 04$ ("ANOVA: Repeated measures within factors" with alpha = .05, power = .8, total sample size = 43, 1 group, 8 measurements, and conservative expectation of 0 correlation among repeated measures).

Recruitment

Participants were convenience-sampled from the general public in the Rhein-Main metropolitan region in Germany. Recruitment was performed via advertisement on the institute website and direct mailings to members of an institute hosted participant database (>2000 members, 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, no impaired hearing, normal or corrected to normal vision and no need to wear glasses during the study (as this might decrease the quality of the eye tracking).

Final sample

47 participants enrolled for the experiment and 45 participants finished data collection. Data from 43 participants were included in the final analysis.

Two types of non-systematic and non-reproducible software errors appeared in some of the recording sessions: 1) a sound-driver related error leading to absence of the auditory stimulus, and 2) a screen freeze, probably linked to the interaction of the eye tracking system with the presentation PC. Two recording sessions were aborted when these software errors appeared for the first time, and data were excluded from analysis because a reconstruction of the time of occurrence was impossible. Two further participants were excluded from EEG analysis due to erroneous amplifier settings that resulted in a sampling frequency of only 250 Hz (incongruent with the preregistered design).

Five of the 43 remaining participants had missing trials due to either of the above mentioned software issues but were included in the analysis. From these 43, one participant reported minor neurological problems (peripheral nerve damage after an accident). Nine other participants reported current or past episodes of mental or psychological disorders. One of these nine indicated current and long-term medical treatment.

The participants were between 19 - 52 years old (mean = 27.1 years, std = 7.1 years), 26 of them female, 17 male, and 0 indicated "other." Participants had received between 9 - 25 years of education (mean = 17.8 years, std = 3.6 years), with the sample showing a strong bias towards highly educated people: 40 out of 43 held the German Abitur or a university degree as their highest qualification. 39 of the participants were right handed, 2 left handed, and 2 ambidextrous (based on self report). Eye dominance was assessed by the experimenter (see below): 33 participants exhibited right eye, 6 left eye, and 4 no eye dominance. 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 - 4.28 mg/kg body weight (mean = 0.84 mg/kg, std = 1.04 mg/kg). 15 out of 43 participants indicated zero caffeine intake before participation. See Supp. Tabs. 2 and 3 for full details on participant demographics.

All participants received monetary compensation of 14 euro per hour 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).

2.2 Experimental design

This study was laid out as a 2x2x2 fully crossed factorial design with each of the three stimulus factors having two levels: factor 1 – fixation task (attempted fixation or free gaze); factor 2 – stimulus dynamics (dynamic video stimuli or static picture stimuli); and factor 3 – stimulus content domain (dance or landscape scenes). An overview of the design is shown in Fig. 1.

We presented 80 trials per participant in a fully balanced fashion resulting in 40 trials within each level of a given stimulus factor (e.g. video stimuli, regardless of content and fixation task) and 10 trials for each possible combination of factors (e.g. static landscape pictures with fixation task). Rating responses were collected following each stimulus presentation trial.

Participants finished all parts of the study in a single session. Electroencephalographic (EEG), electrocardiographic (ECG), and electromyographic (EMG; data not analysed) data were recorded continuously, together with a marker channel for subsequent epoching. Eye tracking data were recorded in epochs time-locked to each trial start.

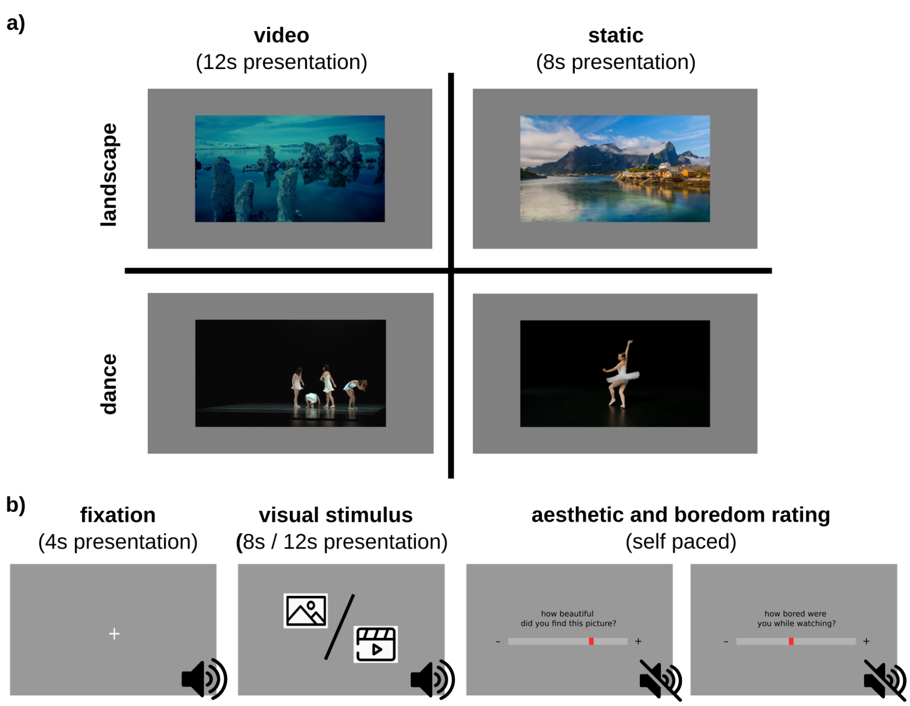

Figure 1: Overview of the experimental paradigm. a) fully crossed 2x2x2 factorial design with either dance or landscape stimuli, and either dynamic video or static picture stimuli; the third condition, either unconstrained gaze or attempted fixation on a fixation dot, is not shown in the Figure. b) general sequence of each of the 80 experinental trials: 4s of fixation dot on grey background are followed by the visual stimulus (8s or 12s respectively). an auditory stream eliciting an auditory steady-state response (ASSR) accompanies these parts. afterwards, to behavioral ratings of aesthetic appeal and state of boredom during the trial are collected (self-paced).

Task procedure

First the participant answered a set of ancillary questionnaires (see below) in digital form on a laptop computer, while the experimenter prepared the EEG cap. Ocular dominance was determined by the experimenter using a variation of the Porta test (as described in Roth, 2002).

After that the participant entered the cabin, the EEG cap was mounted, and peripheral electrodes (ECG and EMG) 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.

To begin the experiment, the participant's personalized hearing threshold was determined with a self paced staircase 1-up-1-down task. A short burst of the same auditory stimulus presented in the experiment was used for the threshold detection, starting with an unnoticeable low volume. Participants were instructed to increase the volume until they heard the stimulus for the first time, then decrease volume, until they did not hear the stimulus anymore, then changed direction again and so forth. After 7 reversals the task ended, and the individual sensory threshold was set as the average of the last 3 reversal intensities.

After the end of the threshold task 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.

Instructions for the main experiment were displayed, followed by a training session consisting of a calibration routine for the eye tracker (horizontal and vertical calibration at 5 positions) and four practice trials. The four practice trials contained stimuli from all possible conditions: dance and landscape, video and static, fixation task and free gaze. These stimuli were not presented again in the main experiment. The main experiment then started with a calibration of the eye tracker followed by the first block of trials.

The general sequence of each trial is shown in Fig. 1b. A trial began with 3 s of a blank screen, accompanied by auditory stimulation. Auditory stimulation continued during presentation of a fixation cross (4 s) and presentation of either a video clip (12 s) or static picture stimulus (8 s). Auditory stimulation stopped with the end of the image or video stimuli, and was followed by a blank screen (4 s), a self-paced aesthetic rating (see below), another blank screen (1.5 s), a self-paced boredom rating (see below), and a final blank screen (2 s). A mid-grey background ($rgb[128, 128, 128]$) was used for the entire experiment.

The presentation durations were chosen as a compromise: for video clips, shorter than 12s is quite short for an experience to evolve, while long periods of static pictures might be experienced as boring. EEG analyses were performed on a common time window of 0-8 s (last 4 s of video presentation omitted).

Stimuli were presented in 16 blocks of 5 trials, consistent in modality (dance or landscape, video or static), and observation task (fixation or free gaze). The presentation order of stimuli within the blocks was randomized. The same random order was applied for all participants, yet counterbalanced by reversing it for every other participant. Source stimuli were not repeated until each was shown once (i.e. the first 40 trials consisted of 20 pictures and 20 videos from the 40 different source clips; see below).

At the beginning of each block the type of observation task (fixation or free view) and a counter of the remaining blocks was displayed to the participant. Block conditions were intermixed; the block order was counterbalanced across participants. The eye tracker was re-calibrated at the beginning of every other block. A longer self-paced break was offered after half of the blocks (40 trials).

Behavioural measures and questionnaires

First, participants answered a set of questionnaires: basic demographic information (including age, gender, education, and information on mood or neurological disorders), caffeine intake of the day (Bühler et al., 2014), the big-five personality inventory (BFI-2-XS: Soto and John, 2017; Rammstedt et al., 2020), the short boredom proneness scale (sBPS: Struk et al., 2017), the positive negative affect schedule (PANAS-SF: Thompson, 2007), and the Snaith-Hamilton pleasure scale (SHAPS: Snaith et al., 1995) in their German versions.

Ratings of aesthetic appeal and state boredom 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 lack of 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 question: "Wie sehr hat diese Szene Sie angesprochen?" ("How much did this scene appeal to you?") for the aesthetic rating, or "Haben Sie sich beim betrachten der Szene gelangweilt?" ("Were you bored while watching the scene?") for the boredom rating respectively. The order of the ratings was fixed with the aesthetic rating always coming first.

In the more detailed task description preceding the study, the aesthetic rating was related to the concept of "being moved" by an aesthetic stimulus (Menninghaus et al., 2015): This psychological construct has been used in several behavioral and MRI studies both by our lab and others (Armstrong and Detweiler-Bedell, 2008; Menninghaus et al., 2015; Silvia, 2009; Vessel et al., 2012). Previous research on boredom typically applies induction paradigms via specific boring tasks or video material (Raffaelli et al., 2018). Since here we are interested in dynamic changes of state boredom over the course of the experiment we apply repeated sampling with a self-report rating analogous to the assessment of aesthetic appeal. In the task descriptions we emphasized that this question was targeting the participant's state of perceived boredom, not their evaluation of the stimulus or any of its features.

2.3 Stimuli

Video stimuli

The video stimuli were generated from a larger set of 60 video clips (30 dance performances, 30 landscapes) of 30 s length compiled for a previous study (Isik and Vessel, 2019). These clips were screened by a lab assistant naive to the purpose of the study: all shots (continuous segments without cuts) of 12 s or longer were identified, and 12 s excerpts starting with the first frame after a cut were extracted using Adobe Premiere (Adobe Inc.). If any 30 s clip had no cuts at all or contained shots much longer than 12 s the assistant was instructed to extract 12 s excerpts that did not start in the middle of a fast pan (camera shift) or dance move. This procedure yielded a set of 86 excerpt clips from 47 of the 60 original 30 s clips (0-3 per original clip; 53 dance performances, 33 landscape). From these 86 clips 40 were chosen as stimuli for the experiment (20 landscapes, 20 dance scenes; random choice, with only one clip per source video). 1 of the 40 videos was in greyscale and the remaining 39 were color videos. All video clips had an aspect ratio of 16:9, an initial resolution of 1280x720 px, and were compressed in the same video compression method (H.264).

Picture stimuli

The static picture stimuli were still-frames from each of the 20 chosen video stimuli (one picture stimulus per video stimulus). Frames were taken from the last 2 seconds of each clip and had to be free of motion smearing (especially in dance performances). Other than these restrictions, frame selection was based on experimenter choice.

Videos and pictures were scaled not to exceed $20° deg$ of visual angle in the vertical or horizontal dimension, resulting in a stimulus size of 502 x 282 px on the screen.

Auditory stream to elicit ASSR

During each trial, pink noise with a continuous 40 Hz amplitude modulation was played to participants via binaural in-ear headphones with a loudness of 35dB SL (i.e. 35 db louder than their individual sensory detection threshold for the stimulus; see above for description of the threshold detection task). The final loudness was controlled via the presentation software. We chose 40 Hz stimulation frequency because ASSR has been shown to be particularly strong in response to this frequency range (Galambos et al., 1981), possibly due to superposition of brainstem and middle latency response (Bohórquez and Özdamar, 2008).

To generate the auditory stimulus a 30 s pink noise waveform was created using MATLABs *pinknoise* function and dotwise multiplied with an equally long sine wave as the modulating factor; the depth of the amplitude modulation was hence 100%. The stimulus was saved as an uncompressed wav file with 44.1 kHz sampling rate. It can be retrieved from the associated online repository.

Presentation of the auditory stream started 3 s before the first frame of the pre-stimulus fixation cross and ended with the last frame of the visual stimulus, resulting in a total of 15 s auditory stimulation in static picture trials, and 19 s in video trials, respectively. In this design the initial transient response period of the ASSR lies outside the analysed time window of visual stimulation (see below).

2.4 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 x 2,54 x 2,3 m). 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 adjustable, such that participants could sit in a comfortable upright position which they were able to sustain for the time of the study.

The study was run by three different experimenters, with the vast majority (44 out of 47 participants) run by one. 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 framerate of 144 Hz, resolution 1920 x 1080 px (mirrored for the experimenter), and auditory stimuli were presented using an RME Fireface UCX Audiointerface (Audio AG, Germany) with ER3C Tubal Insert Earphones (Etymotic Research Inc., USA).

EEG and peripheral physiology acquisition

EEG data were collected using a 64 channel actiCAP system with active Ag/AgCl electrodes with no active shielding (Brain Products GmbH, Germany), placed according to extended international 10-20 localization system (com, 1958; Oostenveld and Praamstra, 2001) with FCz recording reference and AFz ground. The cap was positioned by centering the Cz electrode on the axes *Nasion* to *Inion*, and left ear to right ear. The channel layout included 2 bipolar auxiliary channels for ECG (electrodes placed on the right mid clavicle and lower left rib cage, in correspondence with the II Einthovens derivation) and EMG (electrodes placed on left forearm, on top of *Musculus brachioradialis* and *Musculus extensor carpi radialis longus* with a distance of approximately 5-10 cm depending on the participants size; EMG data were not analyzed). The skin at electrode sites was cleaned with alcohol in advance.

Signals were amplified using a BrainVision actiCHamp 128 Amplifier with 5 BIP2Aux Adapters with 24-bit digitization and an analogue bandpass filter from DC - 280 Hz, 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 GmbH, 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.

Eye tracking acquisition

Eye position was 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 controlled using the PyLink Python module (SR Research Ltd., version 1.11.0.0) Epoched recordings were made for each trial, with a sampling frequency of 500 Hz. The eye tracker was re-calibrated every other block using a horizontal and vertical calibration routine at 5 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.

2.5 Data preprocessing and processing

Data preprocessing, processing, visualization, and analysis was done using Python (version 3.7), Matlab (The MathWorks Inc., version R2018a), and R (Team, 2018, version 3.5.0).

EEG Data Preprocessing

EEG Data were transferred into BIDS format (Pernet et al., 2019) using the MNE-BIDS Python module (version 0.4 Appelhoff et al., 2019).

EEG data were only minimally preprocessed with the PREP pipeline (Bigdely-Shamlo et al., 2015), implemented in EEGLAB (Delorme and Makeig, 2004), using default parameters. The pipeline is an automated and fully reproducible multistage preprocessing routine consisting of line noise removal, highpass filtering, robust rereferencing to average reference, detection of noisy channels, and subsequent interpolation of these channels. Other than mentioned in the preregistration, we applied no rejection of data based on visual inspection to enhance reproducability of the study.

To further evaluate the proxy metric, we generated an alternative version of the EEG data, more thoroughly cleaned with the automagic preprocessing pipeline (Pedroni et al., 2019). This pipeline combines the early stage preprocessing of the PREP pipeline with a fully automatic ICA based artefact removal to clean the EEG data from common sources of measurement noise, including eye movements (MARA; Winkler et al., 2011).

EEG processing pipeline - ASSR

The ASSR analysis was performed using MNE-Python (Gramfort, 2013, version 0.22.0) and custom code. All analyses were done in sensor space. 8 s epochs were extracted starting from the beginning of each visual stimulus presentation (the last 4 s of the longer movie trials were omitted to yield the same amount of data across all trial types).

The metric we extracted for the present study is the signal-to-noise ratio (SNR) of the ASSR, a local ratio of the power at the stimulation frequency versus the average power in neighboring frequencies (Meigen and Bach, 1999). For each combination of sensor and trial, power spectral density (PSD) was calculated using FFT on the full 8 s of visual stimulation. The PSD spectrum was then transformed into a spectrum of signal-to-noise ratio (SNR) using custom code: The SNR measure we use is defined as the power in a given frequency band relative to the average power in the neighboring frequency bins. Therefor we convolved the PSD arrays along the frequency axis with the following convolution kernel: 6 surrounding bins (3 on each side), skipping the 2 directly neighboring bins. This yields the SNR spectrum for every trial and channel, composed of unit-less values. A SNR-value for a given frequency, channel, and trial can take any positive value (SNR > 0) but in the absence of narrowband rhythmic activity in this frequency should be approximately 1. Values much bigger than 1 indicate narrow band rhythmic components in the EEG, as expected for ASSR (but also e.g. power line noise, if not removed). As the dependent measure of interest for this study we extract only SNR at stimulation frequency, i.e. 40 Hz for every trial and EEG channel. The code created for this task was also made publicly available as part of a tutorial in the MNE-Python documentation.

For the subsequent statistical analysis, the resulting SNR data arrays were averaged in two dimensions: over all channels of the montage resulting in one SNR value per trial per participant, and subsequently (after log transformation) over subsets of trials, resulting in one SNR value per stimulus condition per participant. Other than in typical studies working with ASSR we did not confine the analysis to an auditory region of interest (ROI) on the scalp where the responses are strongest, but averaged data from all registered channels; while an ROI analysis would result in higher average SNR values, such a design could only reflect distortions in these channels while spatially localized noise outside the ROI would be neglected.

Other than preregistered, we decided to log transform SNR values to account for the variable's skewed gamma distribution which we hadn't considered in advance see Supp. 5.3. Transformation took place on the trial level, after averaging SNR values over all EEG channels. After transformation, the distribution indeed assumed a more gaussian shape and this step did not change the significance pattern of the main ANOVA model.

Eye tracking Data Preprocessing and processing

Recorded data consisted of the raw binocular gaze path (as x/y coordinates for both eyes), but also contained eye blink, saccade, and fixation annotations detected online by the Eyelink system's proprietary algorithms. The recording files were parsed into an accessible format using the proprietary software EDF Converter (SR Research, version 4.0) and a modified version of the ParseEyeLinkAsc module for Python (`https://github.com/djangraw/ParseEyeLinkAscFiles`, code based on version 7/4/19).

Blink annotations were taken from the recorded files (detection based on SR default algorithm) but custom code was used to select only binocular blinks (overlapping blinks in traces of both eyes), and only blinks longer than 30 ms.

Saccades and microsaccades were detected from the raw gaze coordinates using the velocity based detection algorithm proposed by Engbert and Mergenthaler, implemented in the Microsaccade Toolbox for R (Engbert and Mergenthaler, 2006, version 0.9). Parameter were set to binocular saccades, a minimal saccade duration of 6 ms (3 samples), and a relative velocity threshold (VFAC parameter) of 5. This procedure was preregistered as an optional step to increase data quality. Saccades with a peak velocity $\geq 900°/s$ were rejected as artefacts. Saccades with an amplitude $\leq 1°deg$ were categorized as microsaccades.

To align with the EEG data, timing of blinks and saccades was expressed relative to trial onset (as identical triggers were sent from the presentation PC to both the EEG amplifier and eye tracker.) Blinks and saccades outside the trial limits were discarded. For the statistical analysis, counts per trial for blinks, microsaccades and saccades were computed as dependent trial measures (8s beginning with the onset of the visual stimulus).

ECG Preprocessing and processing

ECG data handling and initial preprocessing of the unsegmented recording was done in MNE Python, processing and detection of R-peaks and heart rate in trial segments was done using HeartPy module for python (van Gent et al., 2019, version 1.2). The data were visually inspected to detect whether polarity of the recording electrodes was correct, if not this was accounted for during preprocessing.

The initial processing pipeline consisted of a bandpass filter between 0.01 - 100 Hz (FIR filter, zero phase), a 50 Hz notch filter to remove line noise (zero phase), cutting epochs of -8 to 8 s around visual stimulus onset. Then Heartrate (HR) was computed for every trial within a baseline window (8 s immediately preceeding the visual stimulus onset) and a trial window (8s starting with visual stimulus onset) by the following pipeline: In each segment the polarity of the recorded ECG signal was switched if necessary, a notch filter was applied to remove baseline drift (0.05 Hz notch, zero phase), and the signal was scaled to a have a more or less constant amplitude over time (to help the detection algorithm, HeartPy default parameter). The R-peak was detected from the preprocessed ECG, outlier were detected via IQR method (default parameter) and rejected. The interbeat intervals (of accepted R peaks) were averaged and converted to heartrate (HR) in beats per minute (bpm).

For each trial HR values for baseline and trial segments were extracted for all participants. The difference between HR of the baseline segment and HR of the stimulus segment was computed for every trial as the measure of interest. Outlier in HR deceleration values were removed by a conservative visually derived threshold (lower limit: -30 bpm, upper limit: +30 bpm), which rejected data from 13 trials. We note that this analysis of the ECG data was not preregistered in detail, but only generally listed as an exploratory investigation.

2.6 Statistical analyses and visualization

ANOVA model

Main effects and interactions of the experimental conditions on the dependent measures were tested using a categorical three-way fully crossed ANOVA with 8 groups with 3 factors in a within subject design (repeated measures ANOVA). The three factors were viewing task (fixation/free), stimulus motion (video/static) and content (dance/landscape). We note that in the preregistration, while we described the intended three-way ANOVA we erroneously wrote about a "two-way" ANOVA; we apologise for any confusion.

This repeated measures ANOVA model was applied to the following dependent measures individually: ASSR SNR (log transformed average over all sensors for each trial, see below and Supp. 5), aesthetic ratings, boredom ratings, blink rate, saccade rate, and microsaccade rate (as counts per trial) and heart rate deceleration (in bpm). Responses were therefore aggregated and averaged within participants for each of the 8 possible factor combinations (i.e. averaging over 10 trials within each participant). Unfortunately a bug in the randomized allocation of the stimuli to the viewing task condition lead to a slight unbalance in some of the groups. This error occurred for half of the participants: While overall full balance was kept for the main effects (n=1674-1675 trials each in all participants) and for two of the two-way combinations (stimulus motion x content and fixation task x content; n=834-840 trials in each combination), the combination of fixation task x stimulus motion (n=614-615 trials in dynamic x free gaze and static x fixation, but n=1060 trials in dynamic x fixation and static x free gaze) as well as the three-way combinations were unbalanced (n=530 trials or n=304-310 trials). However, each participant was presented with at least 5 trials per combination (instead of 10 as laid out), and as the ANOVA model was applied to averaged values per participant the effect should be manageable.

The models were created using least squared regression implemented in the AnovaRM class from the statsmodels module for Python (Seabold and Perktold, 2010, version 0.12.1). An effect size (η_p^2: partial eta squared) for all main effects and interactions was calculated from the ANOVA table using the formula for fixed effect designs

$$\eta_p^2 = \frac{F \times df_{\text{effect}}}{F \times df_{\text{effect}} + df_{\text{error}}}$$

implemented in custom code (after Lakens, 2013). Full ANOVA tables can be found in the supplemental material 5.3).

Correlation analysis

We analysed the trial wise correlation between several of the response measures: aesthetic and boredom ratings, SNR and both ratings, SNR with eye blink count, saccade count and microsaccade count, and both ratings with eye blink count, saccade count and microsaccade count.

Therefor we applied repeated measures correlation (Bakdash and Marusich, 2017) to take into account that individual observations were clustered by participant; we used the method implemented in the rm_corr function from the pingouin module for Python (Vallat, 2018, version 0.3.8). To construct the rmcorr model of the blink data we had to remove data from 2 additional subjects who exhibited 0 detectable blink during all of the trials (the model did not converge). All correlation results were corrected for multiple comparison using Holm's method (Holm, 1979) implemented in the multipletests function from the statsmodels module for Python (Seabold and Perktold, 2010, version 0.12.1). Correlation pattern of all dependent variables can be found in the supplemental material 5.3).

Local regression in Fig. 2c was generated using locally weighted scatterplot smoothing (LOWESS) as implemented in the seaborn module for Python (Waskom, 2021, version 0.11.1)

Visualization

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

3 Results

Observers viewed videos and images of natural landscapes and dance performances and were either allowed to freely view the stimuli or asked to fixate a central fixation dot in a fully crossed factorial design (2 stimulus motion x 2 content x 2 tasks), while EEG, eye tracking and ECG data were recorded. Following each trial, observers rated the stimulus for aesthetic appeal and for their state of boredom. Main effects and interactions of the experimental conditions on the various dependent measures were tested using a categorical three-way repeated measures ANOVA with 8 groups and 3 factors: viewing task (fixation/free), stimulus motion (video/static) and content (dance/landscape). Average ratings of the trials within each of the 8 groups per participant were used. Full ANOVA tables can be found in the supplemental material 5.3.

3.1 Behavioral measures of aesthetic appeal and boredom

Average ratings of both aesthetic appeal and of boredom were strongly affected by experimental condition (Fig. 2a and b). The overall 3-way ANOVA for aesthetic ratings was highly significant ($F(301, 42) = 5.87$, $p < .001$), accounting for approximately 81% of the variance in participant's ratings ($adj.R^2 = 0.81$). The 3-way ANOVA for boredom ratings was also highly significant ($F(301, 42) = 3.84$, $p < .001$), and accounted for approximately 71% of variance in aggregate participant ratings ($adj.R^2 = 0.71$. Both models show several significant main effects and interactions.

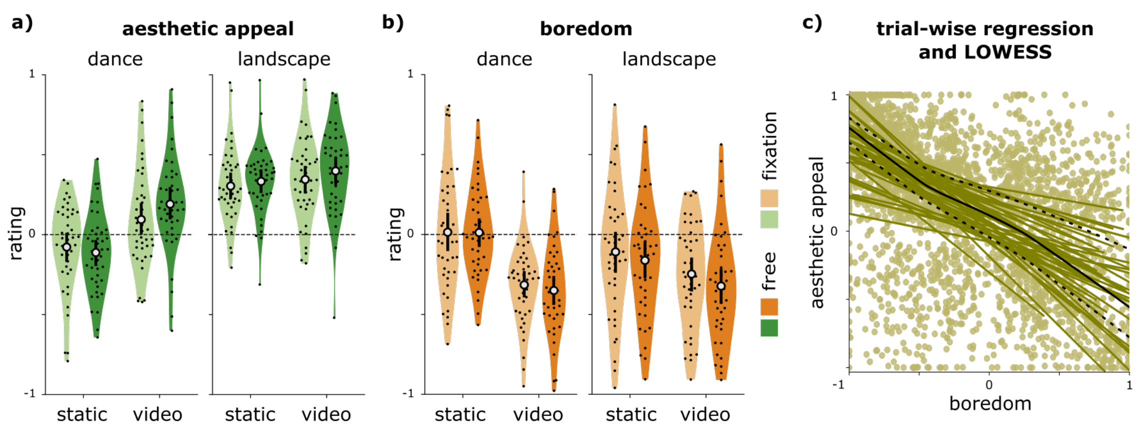

Figure 2: Effect of presentation category on stimulus ratings and correlation of aesthetic appeal and boredom ratings (N=43 participants). a) average aesthetic rating by category (one dot for each participant): landscapes were rated significantly higher than dance performances (p < .001), dynamic video stimuli were rated significantly higher than static images (p < .001), and stimulus motion (video/static) interacted significantly with content (dance/landscape; p < .001) suggesting that dynamic video stimuli affected the appeal of dance to a greater extent compared to landscapes. b) average boredom rating by category (one dot for each participant): significantly lower boredom ratings for videos compared to static images (p < .001), and stimulus motion again interacted significantly with stimulus content (p < .001) with the same effect direction. c) correlation between aesthetic and boredom ratings (one dot for each trial): ratings of aesthetic appeal and boredom were negatively correlated ($r(3305) = -.58$, $p < .001$). This relation was very robust for individual observers (individual lines). A locally weighted fit of all data (LOWESS; solid black line) suggests that the relationship has a degree of nonlinearity and that this non-linearity is stronger in landscape (upper dotted line) than in dance stimuli (lower dotted line): a larger number of landscape trials were rated as boring yet still moderately aesthetically appealing.

As hypothesized, aesthetic ratings of natural landscape stimuli were significantly higher compared to the dance performances (Fig. 2a; $F(1, 42) = 50.21$, $p < .001$, $\eta_p^2 = 0.54$). This replicates findings from previous research that nature scenes tend to be rated higher than various other visual stimulus categories (Isik and Vessel, 2019; Vessel et al., 2018)

In addition, dynamic video stimuli were rated as significantly more appealing than static images ($F(1, 42) = 29.72$, $p < .001$, $\eta_p^2 = 0.41$) and stimulus motion (video/static) interacted significantly with content (dance/landscape; $F(1, 42) = 18.50$, $p < .001$, $\eta_p^2 = 0.31$). In particular, moving from video to static images affected the appeal of dance to a greater extent than for landscapes. No other main effect or interaction reached significance.

Turning to the boredom ratings, we observed significantly lower boredom ratings for videos compared to static images (Fig. 2b, $F(1, 42) = 41.18$, $p < .001$, $\eta_p^2 = 0.50$), confirming that restricting stimuli to static images has a negative effect on engagement. As for aesthetic ratings, stimulus motion again interacted significantly with stimulus content (dance/landscape; $F(1, 42) = 18.92$ $p < .001$, $\eta_p^2 = 0.31$), with the shift from video to static images having a greater effect with dance stimuli than with landscape stimuli. Although we had hypothesized that the free-viewing task would result in lower boredom ratings compared to the fixation task, this effect, though in the predicted direction, was not significant ($F(1, 42) = 2.16$, $p = .15$, $\eta_p^2 = 0.05$). No other main effect or interaction reached significance.

As expected, aesthetic ratings and boredom ratings were negatively correlated ($r(3305) = -.58, 95CI = [-.60, -.56], p < .001$; Fig. 2c and Supp. Tab. 5). The relationship was extremely robust for individual observers, whom all showed a negative relationship (Fig. 2c, individual lines). A locally weighted fit of all data (LOWESS; solid black line in Fig. 2c) suggests that the relationship has a degree of nonlinearity, being more steep at the extremes but shallower for middle ratings. Separate LOWESS fits for landscape (upper dotted line in Fig. 2c) and dance stimuli (lower dotted line in Fig. 2c) reveal that this might be attributed to a less linear relationship for landscape stimuli compared to dance: a larger number of landscape trials were rated as boring yet still moderately aesthetically appealing.

3.2 Physiological measures

Eye tracking

In order to better understand how relaxing experimental constraints might affect behavior and EEG signal, average blink rate, saccade rate, and microsaccade rate were extracted from eye tracking data (aggregated to counts per trial; see Fig. 3).

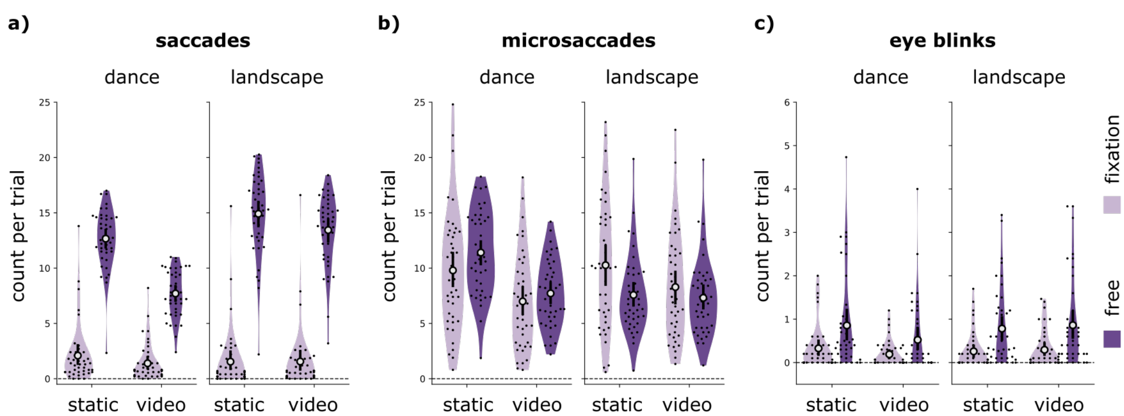

Figure 3: Effect of presentation category on gaze (N=43 participants). a) larger saccades ($> 1°deg$) were significantly decreased by fixation task ($p < .001$), but there were also significantly fewer saccades for video stimuli compared to static pictures ($p < .001$), and for dance compared to landscape ($p < .001$). All interaction effects reached significance. b) there were significantly fewer microsaccades ($< 1°deg$) when viewing dynamic videos than when viewing static scenes ($p < .001$), and when viewing landscapes compared to dance ($p = .025$). Further there was a significant interaction between stimulus content and stimulus dynamics, with a more pronounced effect of video in the dance stimuli compared to landscapes ($p < .001$), a significant interaction between stimulus content and viewing task ($p < .001$), and a significant three-way interaction ($p = .008$). c) eye blink rate was significantly decreased by fixation task ($p < .001$) and video stimuli ($p = .031$), with a significant interaction effect between stimulus content and dynamics ($p < .001$).

Not surprisingly, eye movements were significantly affected by viewing task, but also by stimulus dynamics and content. The ANOVA models for all three measures were significant, though the amount of variance accounted for varied across the three measures. For the number of larger saccades, the model captured 96% of the variance (Fig. 3a, $F(301, 42) = 30.79$, $p < .001$, $adj.R^2 = 0.96$). For microsaccade counts, the model accounted for only 78% of variance (Fig. 3b, $F(301, 42) = 4.93, p < .001, adj.R^2 = 0.78$), and for eye blink counts, the model accounted for approximately 86% of variance (Fig. 3c, $F(301, 42) = 7.75, p < .001, adj.R^2 = 0.86$).

Viewing task modulated larger saccades and blink count (both of them are more frequent in the free viewing condition), with a particularly strong effect on larger saccades ($F(1, 42) = 420.17$ $p < .001$), $\eta_p^2 = 0.91$; Eye blinks: $F(1, 42) = 27.46$ $p < .001$, $\eta_p^2 = 0.40$). Thus enforcing fixation did indeed significantly reduce eye movements, as intended. Interestingly however, we observed no significant effect of the fixation task on microsaccade count.

Saccade count was also strongly affected by both stimulus motion and stimulus category, and all interaction effects reached significance. Perhaps counterintuitively, there were more saccades when viewing static images than videos ($F(1, 42) = 152.88$ $p < .001$, $\eta_p^2 = 0.78$), and more saccades for landscape stimuli than for dance stimuli ($F(1, 42) = 135.02$, $p < .001$, $\eta_p^2 = 0.76$). Stimulus motion interacted with stimulus content, such that there was a stronger effect of motion for the dance stimuli ($F(1, 42) = 58.36$, $p < .001$, $\eta_p^2 = 0.58$), and stimulus motion also interacted with task, with a stronger effect of task for static images ($F(1, 42) = 125.92$, $p < .001$, $\eta_p^2 = 0.75$). Stimulus content and task also interacted significantly, with a stronger effect of task for landscape stimuli ($F(1, 42) = 113.23$ $p < .001$, $\eta_p^2 = 0.73$). Finally, the three-way interaction was also significant (content x motion x task), with the task by motion interaction being larger for dance stimuli, resulting in the comparatively smallest effect of task for dance videos ($F(1, 42) = 28.85$ $p < .001$, $\eta_p^2 = 0.41$).

Interestingly, microsaccade counts were significantly affected by stimulus motion, with fewer microsaccades when viewing dynamic videos than when viewing static scenes ($F(1, 42) = 79.49$ $p < .001$, $\eta_p^2 = 0.65$), and by stimulus content, with more microsaccades in dance than in landscape ($F(1, 42) = 5.42$ $p = .025$, $\eta_p^2 = 0.11$). There was also a significant interaction between stimulus content and viewing task, with a differently oriented effect of fixation task in the two contents (fewer microsaccades under fixation in dance, more so in landscapes; $F(1, 42) = 31.25$ $p < .001$, $\eta_p^2 = 0.43$). Further there was again a significant interaction between stimulus content and stimulus motion, with a stronger effect of motion on number of microsaccades for the dance stimuli compared to landscapes ($F(1, 42) = 24.13$ $p < .001$, $\eta_p^2 = 0.36$). The three-way interaction was also significant ($F(1, 42) = 7.83$ $p = .008$, $\eta_p^2 = 0.16$).

Besides the effect of viewing task eye blinks were marginally significantly affected by stimulus motion, with fewer blinks during dynamic video clips ($F(1, 42) = 4.96$ $p = .031$, $\eta_p^2 = 0.11$), and a significant interaction between stimulus dynamics and stimulus content ($F(1, 42) = 9.89$ $p = .003$, $\eta_p^2 = 0.19$).

We observed several significant trial wise correlations between eye movements and the behavioral ratings: both aesthetic ratings and boredom ratings were significantly correlated with eye blink count (aesthetic appeal: $r(3305) = -.06$, $95CI = [-.09, -.02]$, $p < .001$; boredom: $r(3305) = .08$, $95CI = [.04, .11]$, $p < .001$) and microsaccade count (aesthetic appeal: $r(3305) = -0.09$, $95CI = [-.12, -.05]$, $p < .001$; boredom: $r(3305) = .13$, $95CI = [.09, .16]$, $p < .001$), with boredom exhibiting a slightly stronger correlation. Saccade count did not significantly correlate with the ratings.

ASSR SNR

One of our primary goals was to assess the effects of task and stimulus motion on EEG signal quality. To do so, a continuous auditory stimulus (pink noise with 40 Hz amplitude modulation; See methods) was played during each trial, and SNR of the auditory steady-state response (ASSR) was computed as a proxy measure for overall EEG recording quality.

The distribution of raw SNR values was found to resemble a highly skewed gamma distribution with long tails and SNR values were therefore log-transformed (see Supp. Fig. 5).

Despite their strong effects on behavioral ratings of aesthetic appeal and boredom, the experimental manipulations had only minimal effects on measured EEG SNR (Fig. 4a). Average SNR-values were slightly higher under attempted fixation compared to free viewing. The overall ANOVA model was significant ($F(301, 42) = 8.57$, $p < .001$, $adj.R^2 = 0.87$), with fixation task as the only significant factor ($F(1, 42) = 6.24$ $p = .017$, $\eta_p^2 = 0.13$).

Importantly, while we barely found a significant effect of the experimental conditions on SNR values we believe that the measure is sensitive to adequately capture bad quality data and noisy trials. The trialwise SNR measure did correlate with the number of blinks per trial ($r(3305) = -.05$, $95CI = [-.09, -.02]$, $p = .002$) and with the number of saccades ($r(3305) = -.04$, $95CI = [-.08, -.01]$, $p < .018$; see Fig. 4b). However, only the correlation with blink count survived the multiple comparison correction ($p.corr = .036$). The SNR measure did not correlate with the number of microsaccades per trial (repeated-measures correlation $r(3305) = .00$, $95CI = [-.03, .03]$, $p = .98$).

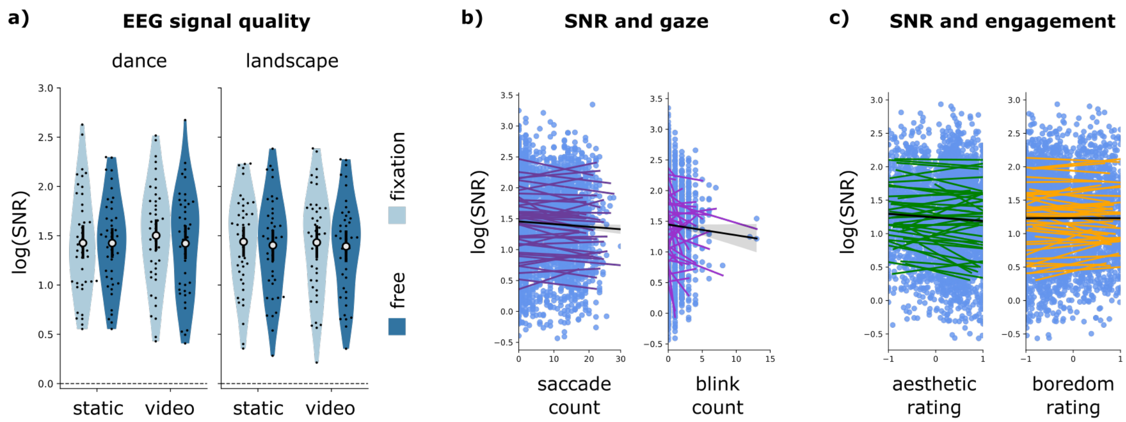

Figure 4: Effect of presentation category on ASSR SNR and correlation of SNR with saccade count, eye blink count, and behavioral ratings (N=43 participants). a) significantly higher SNR for fixation task compared to free gaze (p = .017). b) significant trial wise correlation (repeated measures correlation) of log transformed SNR with the participants' number of eye blinks per trial ($r(3305) = -.05$, $p = .04$) but not with number of saccades per trial ($r(3305) = -.04$, $p = .18$). c) no significant trial wise correlation (repeated measures correlation) between log transformed SNR and aesthetic ratings ($r(3305) = -.05$, $p = .10$), and marginally significant correlation between log transformed SNR and boredom ratings ($r(3305) = .04$, $p = .26$).

When directly compared with the behavioral ratings (Fig. 4c), there was trial wise correlation between log transformed SNR and boredom ratings ($r(3305) = .04$, $95CI = [.00, .07]$, $p = .029$) as well as with aesthetic ratings ($r(3305) = -.05$, $95CI = [-.08, -.01]$, $p = .008$) but also these correlations were not strong enough to remain significant after multiple comparison correction (see Supp. Tab. 5 for full trial wise correlation structure).

To further evaluate the proxy metric, we repeated the entire analysis on an alternative version of the EEG data that were cleaned in a fully automated preprocessing pipeline with ICA based artefact removal (see Methods). There was now no effect of any of the investigated factors on log transformed SNR: the overall ANOVA model was significant ($F(301, 42) = 7.74$, $p < .001$, $adj.R^2 = 0.86$) but none of the main effects or interactions (effect of fixation vs. free gaze $F(1, 42) = 0.67$ $p = .42$), $\eta_p^2 = 0.02$). Other than in the minimally preprocessed data, log transformed SNR in the cleaned data did not significantly correlate with saccade count ($r(3305) = -.02$, $95CI = [-.05, .01]$, $p = .24$) or blink count ($r(3305) = -.02$, $95CI = [-.06, .01]$, $p = .18$), and also not with microsaccade count (repeated-measures correlation $r(3305) = .02$, $95CI = [-.02, .05]$, $p = .35$).

ECG

We performed an exploratory analysis of event related heart rate (HR) changes to investigate the participants' autonomous responses to the stimuli. We observed a significant HR deceleration after the onset of the visual stimuli, consistent with previous reports (e.g. for IAPS images Palomba et al., 1997). HR during the 8 s time window preceding onset of the visual stimulus (baseline) was significantly larger than HR in the 8 s window after stimulus onset (2-sided paired $t = 33.00$, p < .001, for averaged HR values). However, the investigated experimental factors did not strongly influence HR deceleration. The overall ANOVA model was significant, but accounted for only 33% of the variance ($F(301, 42) = 1.55$, $p = .043$, $adj.R^2 = 0.33$). The only significant main effect was stimulus motion ($F(1, 42) = 11.24$ $p = .002$, $\eta_p^2 = 0.21$) indicating that HR deceleration was slightly stronger for videos than for static pictures.

We further observed no significant correlation of HR deceleration with either aesthetic ratings ($r(3292) = .01$, $95CI = [-.03, .04]$, $p = .76$) or boredom ratings ($r(3292) = -.01$, $95CI = [-.04, .02]$, $p = .57$). Yet, interestingly HR deceleration significantly correlated with blink rate ($r(3292) = 0.05$, $95CI = [0.01, 0.08]$, $p = 0.005$), saccade rate ($r(3292) = .06$, $95CI = [0.02, 0.09]$, $p < 0.001$), and SNR ($r(3292) = -.04$, $95CI = [-0.07, -0.00]$, $p = 0.039$), but not with microsaccade count ($r(3292) = -.00$, $95CI = [-0.03, 0.03]$, $p = 0.98$). Only the correlation with saccade count was strong enough to remain significant after multiple comparison correction ($p.corr = .012$; see Supp. Tab. 5 for full trial wise correlation structure).

4 Discussion

In the present study we assessed trade-offs between ecological validity and EEG signal quality as participants viewed complex, aesthetically engaging stimuli. We could show that the use of video stimuli does not necessarily result in lower EEG quality but can in fact significantly reduce eye movements under free viewing conditions, especially if human agents are depicted (see Table 1 for a comprehensive overview of the main effects). The use of video stimuli further resulted in significantly higher aesthetic ratings and lower perceived boredom, indicating higher engagement with task and stimulus material. This constitutes another possible benefit of using video material, given that a lack of engagement might in turn decrease data quality or even cause untimely termination of data collection. In our data the fixation task was confirmed to significantly reduce eye movements, and to have an overall positive effect on EEG quality. Small fixational eye movements, on the other hand, can not be inhibited by attempted fixation. The fixation task did not significantly influence the investigated behavioral ratings, which is encouraging given the predominance of this experimental constraint in studies using EEG and MEG. Finally, the stimulus content domain - primarily included as a control condition for the behavioral responses - had a remarkably far reaching influence. Beyond its already known effect on aesthetic ratings it also significantly affected saccade count and microsaccade count, and showed consistent interaction effects with the stimulus dynamics in several of the behavioral measures. By presenting these data and findings we hope to help inform future trade-offs during the design and piloting phase of neuroimaging experiments and to encourage researchers to reconsider the default application of canonical constraints.

Table 1: Main effects of experimental conditions on dependent variables

	Free gaze vs Fixation	Video vs Static pictures	Landscape vs Dance
ASSR SNR	decrease 0.13	no effect	no effect
Ratings			
aesthetic	no effect	increase 0.41	increase 0.54
boredom	no effect	decrease 0.50	no effect
Gaze			
eyeblinks	increase 0.40	decrease 0.11	no effect
saccades	increase 0.91	decrease 0.78	increase 0.76
microsaccades	no effect	decrease 0.65	decrease 0.11
Heart rate	no effect	decrease 0.21	no effect

Note: Values below significant main effects indicate the effect size η_p^2

4.1 Using ASSR SNR as a proxy measure for EEG signal quality

We had hypothesized that, without extensive preprocessing of the EEG data SNR of the auditory steady-state response would be affected by the viewing task (i.e. lower SNR in free-viewing condition than in the fixation task). Indeed, we observed that fixation task enhanced grand average SNR across all conditions. The same effect was present in eye blinks and saccades (significantly fewer of all these with fixation task) and SNR was significantly correlated with the number of blinks on a trial-by-trial bases. In order to evaluate the reliability of the proposed proxy measure and look for significant effects of the presentation conditions we did not optimize the preprocessing pipeline for highest possible SNR response. We used the full scalp montage rather than a ROI based analysis (i.e. no spatial cleaning that would differentially remove or reduce noise components with certain sources or orientations), we applied only minimal frequency domain filtering, and most importantly we did not exclude or reconstruct "noisy" data segments (except for rejection and interpolation of outlier channels): for the primary analysis we applied no visual inspection of trials, nor any ICA or EOG based cleaning.

Any of these cleaning routines would likely result in higher SNR measures and a better detection of ASSR, but keep in mind that in our framework the ASSR generated by the brain is considered a fix factor, while we are interested to detect changes in ASSR SNR measured at the scalp that are caused by noise. Thorough cleaning of the data would thus be counterproductive. This was confirmed by our secondary analysis of a version of the EEG data thoroughly cleaned via an ICA based artefact rejection pipeline. In these data, the significant effect of the fixation task as well as the correlation of the SNR measure with saccade count and blink count vanished.

This leads us to the conclusion, that the SNR measure indeed captured eye movement related signal distortions that would typically lead to a rejection of trials or data segments in common EEG cleaning routines (Luck, 2014). However, the effect size associated to the fixation task is small and there are large differences across participant in their average SNR, indicating that that the SNR values may be meaningfully compared only within one subject, but not across participants. Hence, even though ASSR SNR shows some sensitivity and can capture noise in the EEG, its promise as a universal quantitative metric for EEG quality is questionable. We will discuss this point in more detail below.

4.2 Using video stimuli improves engagement, reduces eye movements, and only minimally affects EEG SNR

Importantly we found evidence for our hypothesis that the fixation task and the other experimental conditions compared in this study not only affected EEG signal quality, as intended, but also several other dependent measures. We observed higher aesthetic and lower boredom ratings for videos than for static stimuli. While the two measures might focus on slightly different psychological evaluation processes, both should reflect an engagement component and hence this finding suggests that video stimuli were experienced as more engaging than the static images generated from the exact same visual content.

While high engagement with stimulus material is certainly important for fields that directly study it, such as empirical aesthetics, we want to emphasize that other fields might also want to design their experiments as engaging as possible. Many researchers might have experienced that soft factors such as a participant's commitment, their motivational state, or the engagement with the task can have an influence on data quality, in the worst case leading to early termination of recording sessions. In the EEG, however, such effects are to date very difficult to quantify: extreme cases in which subjects are very tired, marked by increased alpha activity, can be identified and labelled by field experts (Lacourse et al., 2020). However, this is not a part of typical preprocessing routines in cognitive neuroscience. Another possible mechanism by which EEG quality might be affected by a participants state of engagement is physical restlessness, which can lead to more frequent body- (muscle artifacts) and eye movements (Ramos Gameiro et al., 2017). If anything, bad quality data due to a supposed lack of engagement with the task can be identified based on behavioral performance (e.g. response accuracies, false response rates, or response times). However, in order to do so researchers need prior knowledge to set meaningful performance thresholds and risk excluding participants who are engaged but simply bad at the specific task (thereby potentially introducing bias to their population and data by rejecting lower tails of a distribution). Importantly, cleaning data based on participants' responses is less straight forward for exploratory studies and tasks that do not have "correct" answers, such as ratings of preference).

In the light of the present findings, we want to point out that EEG signal quality (and especially eye movement related artefacts) should not be a major concern when deciding whether to use videos or static images. On the contrary, using video stimuli might even come with further practical advantages besides higher engagement: with video material we observed significantly fewer blinks, microsaccades and saccades, and our proxy measure for EEG quality was not affected by stimulus dynamics. Interestingly, the boredom ratings were significantly correlated with the participants' eyeblink rate and especially their microsaccade rate, indicating that boredom can increase these potentially problematic eye movements, even despite a fixation task.

Hence, dynamic stimulus material might in fact be a promising tool to naturally reduce eye movements and blinks without having to apply a fixation task with all its potentially negative side effects (see above). Our data suggests that this effect might be especially strong when humans are depicted. These positive effects of video might be related to a phenomenon called center bias - a tendency to shift and keep the gaze focused at the center of a visual stimulus -

that appears to be stronger in videos than in static stimuli (for a review see Smith, 2013), and potentially also to the fact that dynamically changing visual stimuli can exert some level of exogenous control on observers' gaze, thereby creating temporal attentional synchrony across viewers (Goldstein et al., 2007; Smith and Henderson, 2010; Smith and Mital, 2013). This latter point further suggests that video stimuli might be used to align both the sensory input stream and endogenous eye-related noise (Nikolaev et al., 2016) across participants, potentially with a similar efficiency as a fixation task. Indeed, hyperscanning studies that investigate how similar brain responses are across observers frequently use video stimuli (Hasson and Frith, 2016; Poulsen et al., 2017).

In current practice, though, most researchers would refrain from using video stimuli, not only in electrophysiological work but in cognitive science in general. There are several reasons for this level of caution, mainly the reductionist view that stimulus dynamics is an (in many cases) unnecessary degree of freedom in the stimulus that can be taken out of the equation by using static pictures. However, while initial studies suggest that some findings from the huge body of EEG work with static stimuli might translate to dynamic video stimuli (e.g. the N170 ERP component was observed using video material, Johnston et al., 2015) a direct transfer of findings between stimulus modalities is not straightforward, but needs to be established for each component. As previous work from the eye movement literature and the present work suggest, there seem to be significant differences between static and dynamic visual stimuli both in the observers' visual exploration and gaze patterns (leading to different low-level input streams to the visual cortex) as well as in their higher level cognitive processing (manifesting in differences in preference and boredom ratings).

Researchers from fields that rely on unconstrained vision have developed tools and experimental paradigms that allow for analysis of electrophysiological data recorded under such uncommon conditions (e.g. Ehinger and Dimigen, 2019; Lu and Ku, 2020; Ayrolles et al., 2021). Indeed, we would suggest that a growing literature of neuroimaging studies using dynamically changing visual stimuli, regardless of the specific topic at hand, would be beneficial to the broader field: to date, such work is underrepresented in the neuroscientific literature and could eventually bridge the gap to the body of behavioral work concerned with unconstrained vision.

4.3 Fixation task improves EEG SNR, and does not significantly impact ratings, but does introduce cognitive load

The effect of viewing task on eye movements was expected, as it is the very reason to apply this restriction in the first place. We confirmed that there were significantly fewer blinks and saccades in trials with a fixation task compared to free gaze trials. Notably the number of blinks was overall very low with less than one blink per trial averaged over all participants (0.3 per 8s trial for fixation, 0.8 per 8s trial for free gaze). This is well below spontaneous blink rates reported in the respective literature, also for the free gaze condition (e.g. Jongkees and Colzato, 2016, 7.2 - 18.9 blinks per minute (equivalent to 0.96 - 2.52 per 8 s) in the reviewed cognitive research studies, and 6 - 34.4 blinks per minute (equivalent to 0.8 - 4.59 per 8 s) over a broad variety of tasks and situation in healthy populations). As to be expected the number of larger voluntary saccades was heavily reduced as well, from a bit less than one per second in the free viewing to less than one saccade per trial in the fixation task (1.6 per 8s trial for fixation, 12.2 per 8s trial for free gaze). This replicates findings from the eye-tracking literature (e.g. Otero-Millan et al., 2008). Again, the numbers observed in this study are very low compared to spontaneous saccade rates. One reason might be that the stimuli were presented in a relatively small format ($< 20° deg$ of visual angle) and could be almost fully captured by near peripheral vision (i.e. without saccadic scanning of the stimulus). We want to emphasize that, as seen in our data and reported before (see e.g. Otero-Millan et al., 2008; Thielen et al., 2019), a fixation task is not suitable to suppress microsaccades. In the present study the number of microsaccades was even slightly higher under attempted fixation (8.8 per 8s trial for fixation, 8.5 per 8s trial for free gaze), but the effect was not significant. Microsaccades can induce transient low gamma band activity in the EEG which can be misinterpreted as brain activity (Yuval-Greenberg et al., 2008) and hence constitutes a source of endogenous noise that can not be counteracted with a fixation task. Unfortunately cleaning the signal using ICA does not reliably remove the effects of microsaccades from the EEG (Craddock et al., 2016; Dimigen, 2020). If this frequency range of the EEG signal matters, one would hence need to coregister microsaccades to fully account for their effects.

Even more important than the absolute reduction of eye movements might be that the fixation task also reduced the variance eye movements across conditions (striking differences in saccade rate across stimulus conditions, see figure 3). Systematic differences in eye movements across conditions can constitute confounds and might lead to false or misattributed findings in downstream analyses (Quax et al., 2019; Thielen et al., 2019).

Apart from eye movements (and SNR) the fixation task did not affect any of the other dependent measures. There was a hint that aesthetic ratings might be slightly higher in the free gaze condition, but this effect was not significant. This is a very encouraging finding, given how ubiquitously the fixation task is applied in EEG and MEG studies regardless of the investigated questions. On the downside however, the predominance of the fixation task in EEG and MEG research renders a comparison to the large body of eye movement literature investigating unconstrained gaze patterns (for a review see Eckstein et al., 2017) or spontaneous blink rates (e.g. Stern et al., 1994; Jongkees and Colzato, 2016), as well as to research done using fMRI or other modalities difficult if not impossible. More importantly, initial research suggests that there might be significant differences in frequently investigated neuronal processing components under free viewing compared to attempted fixation (s.a. in the N170 Auerbach-Asch et al., 2020). When interested in spontaneous gaze patterns and related brain processes, a fixation task is not an option. Additionally, other fields might benefit from a more direct transfer of existing findings or paradigms.

Last but not least, attempting to maintain fixation over a long period adds atypical cognitive load for a study participant, and can lead to fatigue. This might be problematic for some high-level cognitive tasks, such as research on aesthetic processing (Brielmann and Pelli, 2017). In the present study, our stimuli were small and could be almost entirely captured using the initial field of view. In fact, it is possible that this makes fixation easier for the participants: there were anecdotal reports in the debriefing that it is especially hard, or even frustrating, to fixate if there were salient features one would like to explore right outside the focus of visual attention. Hence, it is not clear whether we would have also observed no effects if the stimuli were larger. These factors can decrease a participant's concentration on or engagement with the primary task, thereby introducing a new source of noise to the various dependent measures within an experiment that might be more difficult to track and remove than the eye related artifacts that were to be avoided in the first place.

4.4 Stimulus content also matters, and dance stimuli are differentially affected by video condition

As expected landscape stimuli were, on average, rated more aesthetically pleasing than dance scenes, consistent with previous findings (Vessel and Rubin, 2010; Vessel et al., 2018). Interestingly this main effect was not significant in the boredom ratings, while overall these two types of ratings showed a fairly high negative correlation.

Looking at the LOWESS curves in figure 2c might help to further explore the relation between the two measures. We see a larger fraction of landscape trials with high rated boredom that were nevertheless rated highly aesthetically appealing. This suggests that the aesthetic evaluation of landscapes is guided by a partly different set of features than the evaluation of dance scenes: comparably boring landscapes with less salient features (e.g. meadows compared to a rocky mountain vista) might nevertheless be appealing for their expected affordances (e.g. security, lush vegetation, etc), personal relevance (e.g. because they resemble the region observers grew up in), or other factors. In dance the correlation of aesthetics and boredom was tighter, potentially indicating that the engagement component is more relevant in this domain.

While we cannot offer a conclusive explanation for this phenomenon, the divergent findings are also promising: they posit a proof of concept that the two behavioral rating scales - for boredom and aesthetic appeal - were not used identically by the participants. Further work is necessary to investigate the relation between these measures.

One salient finding in our data is that we consistently observed interaction effects between stimulus content (dance/landscape) and dynamics (video/static). We observed this interaction in behavioral ratings, blink count, saccade counts, and microsaccade counts. The interaction effect always went in the same direction, indicating that the effect of stimulus dynamics was stronger in dance scenes and weaker in landscapes (or that the effect of content was stronger with video and weaker with stills, respectively).

As mentioned above, the correlation between aesthetic and boredom ratings might be affected by the content conditions as well, with a weaker correlation and stronger nonlinearity in landscape scenes.

One potential explanation for this differential effect of content could be that motion matters more if human agents are depicted. Again, affordance theory (Greeno, 1994) might offer an interpretation: A landscape can offer a whole set of affordances like landscape features, relevant objects, or humans, plants, and animals within it that one can interact with. These can be hidden or dispersed within the landscape, and motion would be only one cue among others to identify these (e.g. if movement is happening in a small part of the scene, caused by a small animal that one was not aware of). More likely though, affordances will be identified during active visual exploration. This would fit to our observation that there are significantly more saccades in landscapes than in dance stimuli.

Direct interactions with another human, on the other hand, are arguably largely concerned with communication and understanding the other's intentions (e.g. non-verbal communication cues can convey intentions like sympathy or threat). This might also be implicitly happening while observing dance scenes. In our study's setup vision is the only sensory modality available to the participant, and hence the dancer's motion might constitute the most important stream of information besides overall outer appearance.

4.5 Potentially problematic effects of other experimental conditions in the present study

Here we present a study aiming to assess potential problematic effects of systematic secondary manipulations on primary measures of interest; hence we should also consider whether any secondary experimental manipulations we ourselves introduced might have such influence.

One harsh restriction compared to true natural conditions was the head rest. Without this restriction we would expect increased EEG noise due to muscle movements and physical motion/shock of the cables and sensors, but importantly also brain activity related to the motion and repositioning of the head (Gramann et al., 2021). On the other hand, being forced to not moving the head over a long period might result in fatigue or stiffening neck muscles, thereby also introducing noise in the EEG, but potentially also lowering the engagement with the task or aesthetic appeal. Future research could let participants hold their head freely, and record the head position and movements via accelerometer or neck EMG to quantify these factors as well, but in order to record high quality eye tracking data we unfortunately had to refrain from this in the present study.

The second, and arguably most critical, manipulation introduced in this study was the background ASSR stimulus. Unlike the fixation task and stimulus dynamics the auditory background task was not balanced by any number of trials without auditory stimulation and we cannot directly test for its effect on any of our dependant variables. However, measures were taken to make the stimulus as unobtrusive as possible: the sound stimulus was selected from a set of different candidates in a pilot study (data not shown), and loudness was adapted to the individual sensory threshold of each participant. However, there were individual differences in how the participants perceived the stimulus qualitatively. During debriefing several participants mentioned that they had completely stopped noticing the sound after a while. In contrast, a few other participants mentioned that they had sometimes struggled with a decision whether to incorporate the sound into the summary rating of aesthetic appeal - even though they were aware that the sound was a background manipulation, it was nevertheless experienced as part of an audiovisual aesthetic stimulus. Two comments even pointed at another possible instance of an interaction effect with stimulus content, reporting that they had interpreted the auditory stream as the sound of a helicopter in the landscape videos recorded using drone shots. The auditory stimulus used in this study is substantially pink noise with changing loudness and individual differences in the qualitative perception of audio noise were reported before (Bergamasco et al., 1976). Concerning the brains internal processing of the stimulus, we want to emphasize again that ASSR is primarily an early auditory response (Picton et al., 2003), and it was already shown to not interfere with some types of early visual processing (Keitel et al., 2013). Relatedly, traffic noise and white noise were shown to reduce sensitivity to other auditory stimuli (in this case verbo-acoustic communication), but do not affect sensitivity in the visual domain (Bergamasco et al., 1976). These findings suggest that the ASSR manipulation might not be critical for studies concerned with visual processes. Studies investigating other auditory processing, however, might be more affected, and bottom-up interactions with higher level cognitive processes can not be excluded at this point. We take it as a positive sign, though, that several people did entirely stop noticing the sound.

4.6 Future development of ASSR as a general time-varying SNR measure

With the present work we proposed and tested the use of ASSR as a possible new online marker for EEG recording quality on a trial-wise and potentially continuous bases (e.g. combined with a sliding window function). To our knowledge, no universally accepted marker exists, and it would be promising for applied research (e.g. BCI context) or research using mobile EEG when the external noise level cannot be sufficiently controlled as it is the case in the laboratory. In the light of the present results, however, we would urge caution.

While the measure was sensitive to common sources of endogenous measurement noise (namely eye movements), there are some concerns: 1) The sensitivity was low. Effect sizes and correlation coefficients with the metric were very small (especially in comparison to individual differences across subjects). 2) Comparisons of raw signal level appeared to only be meaningful within, but not across observers. 3) The temporal sensitivity of the ASSR metric was in a medium range - here we used 8 s trials, which was probably at the lower end of the possible range. While this might be enough for an online measure in many applied contexts, it is likely too long for many trial based research applications.

Extensive further research would be necessary to validate applicability and sensitivity of ASSR as an online marker for EEG data quality.

Even though we systematically lifted some experimental constraints, the study was nevertheless conducted in a very controlled lab environment. Hence we can not say much on how the metric would react to typical sources of exogenous measurement noise such as electronic equipment or concussion of the electrodes. Likewise we observed overall very low rates of endogenous noise sources such as eye (see above) and body movements (the head was mounted on a chin rest), or concurrent sensory input (the experiment took place in an acoustically and visually shielded cabin). These conditions fundamentally differ from the largely uncontrollable recording environments in applied and mobile EEG studies that would most profit from a validated online measure of signal quality.

Microsaccades can cause increased low gamma activity in the range of our 40 Hz ASSR stimulation (Yuval-Greenberg et al., 2008). This might be problematic for our study, since microsaccade rate was significantly linked to several of the investigated stimulus categories. However, we did not observe a significant trial-wise correlation of the SNR metric with microsaccade count in either direction. This might imply that either the induced gamma power adds both to signal and to the noise term and hence does not affect SNR, or that the effect of microsaccadic gamma on SNR is too weak (compared to other activity) to manifest in significant changes of SNR. Further research would be neccessary.

Lastly, the question remains whether the brain's actual ASSR (not its measured SNR at scalp level) might be systematically affected by factors not controlled for in the present study. Directed attention towards the auditory stream is controversially discussed to influence the strength of ASSR: while early work found no overall effect of selective attention on ASSR (attending the stimulus vs. reading a book; Linden et al., 1987), more recent work did (eg. for attention shifts between two auditory stimuli [Skosnik et al. 2007], or for attention shift from the visual to the auditory domain or vice versa [Saupe et al. 2009]). If this were true it constitutes a potential confound for the proxy measure only if attention to the auditory stream should be systematically linked to the investigated primary task. In the present study, we took measures to avoid such a link and didn't find evidence for any attention related confound.

5 Additional information

5.1 Code and Data Availability

The experimental design and hypotheses reported here were preregistered (see `https://osf.io/bkep4`). The manuscript was uploaded to a preprint server (see `https://tbd`) The dataset used for statistical analysis, as well as R and python scripts replicating the results, and python scripts used in data collection were made available in a public online repository (see `https://tbd` . The code created for computing ASSR SNR was also made publicly available as part of a tutorial in the MNE-Python documentation; see `https://mne.tools/stable/auto_tutorials/time-freq/50_ssvep.html`. Raw EEG data cannot be publicly shared due to ethics regulations at our institute.

5.2 Acknowledgements

A big Thank you to Ayse Ilkay Isik for the original stimulus set, to Sarah Koldehoff for cutting the final video stimuli, to Elke Lange and Lauren Fink for their valuable comments and feedback on the Eyetracking analysis, and to Sarah Koldehoff, and Johannes Messerschmidt for assistance with the data collection. The work was supported by the Max Planck Society.

5.3 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

(1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology*, 10(2):370–375.

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., and Jas, M. (2019). MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. *Journal of Open Source Software*, 4(44):1896.

Armstrong, T. and Detweiler-Bedell, B. (2008). Beauty as an emotion: The exhilarating prospect of mastering a challenging world. *Review of General Psychology*, 12(4):305–329.

Auerbach-Asch, C. R., Bein, O., and Deouell, L. Y. (2020). Face Selective Neural Activity: Comparisons Between Fixed and Free Viewing. *Brain Topography*, 33(3):336–354.

Ayrolles, A., Brun, F., Chen, P., Djalovski, A., Beauxis, Y., Delorme, R., Bourgeron, T., Dikker, S., and Dumas, G. (2021). HyPyP: a Hyperscanning Python Pipeline for inter-brain connectivity analysis. *Social Cognitive and Affective Neuroscience*, 16(1-2):72–83.

Bakdash, J. Z. and Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology*, 8.

Bergamasco, B., Benna, P., Covacich, A. M., Gilli, M., and Rossi, G. (1976). Behaviour Of CNV During Exposure To Urban Traffic Noise. *Acta Oto-Laryngologica*, 81(sup339):27–29.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., and Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, 9.

Bohórquez, J. and Özdamar, . (2008). Generation of the 40-Hz auditory steady-state response (ASSR) explained using convolution. *Clinical Neurophysiology*, 119(11):2598–2607.

Braboszcz, C. and Delorme, A. (2011). Lost in thoughts: Neural markers of low alertness during mind wandering. *NeuroImage*, 54(4):3040–3047.

Brieber, D., Nadal, M., and Leder, H. (2015). In the white cube: Museum context enhances the valuation and memory of art. *Acta Psychologica*, 154:36–42.

Brieber, D., Nadal, M., Leder, H., and Rosenberg, R. (2014). Art in Time and Space: Context Modulates the Relation between Art Experience and Viewing Time. *PLoS ONE*, 9(6):e99019.

Brielmann, A. A. and Pelli, D. G. (2017). Beauty requires thought. *Current Biology*, 27(10):1506–1513.e3.

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

Chatterjee, A. and Vartanian, O. (2014). Neuroaesthetics. *Trends in Cognitive Sciences*, 18(7):370–375.

Coles, M. G. H. (1972). Cardiac and respiratory activity during visual search. *Journal of Experimental Psychology*, 96(2):371–379.

Craddock, M., Martinovic, J., and Müller, M. M. (2016). Accounting for microsaccadic artifacts in the EEG using independent component analysis and beamforming: ICA, beamforming, and artifactual gamma. *Psychophysiology*, 53(4):553–565.

Debener, S., Minow, F., Emkes, R., Gandras, K., and de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk?: EEG to go. *Psychophysiology*, 49(11):1617–1621.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Dias, J. C., Sajda, P., Dmochowski, J. P., and Parra, L. C. (2013). EEG precursors of detected and missed targets during free-viewing search. *Journal of Vision*, 13(13):13–13.

Dikker, S., Michalareas, G., Oostrik, M., Serafimaki, A., Kahraman, H. M., Struiksma, M. E., and Poeppel, D. (2021). Crowdsourcing neuroscience: Inter-brain coupling during face-to-face interactions outside the laboratory. *NeuroImage*, 227:117436.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., Ding, M., and Poeppel, D. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology*, 27(9):1375–1380.

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *NeuroImage*, 207:116117.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., and Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, 140(4):552–572.

Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., and Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nature Communications*, 5(1).

Dmochowski, J. P., Sajda, P., Dias, J., and Parra, L. C. (2012). Correlated Components of Ongoing EEG Point to Emotionally Laden Attention A Possible Marker of Engagement? *Frontiers in Human Neuroscience*, 6.

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., and Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, 25:69–91.

Ehinger, B. V. and Dimigen, O. (2019). Unfold: an integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *PeerJ*, 7:e7838.

Engbert, R. and Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences*, 103(18):7192–7197.

Faul, F., Erdfelder, E., Lang, A.-G., and 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.

Fink, L. K., Lange, E. B., and Groner, R. (2019). The application of eye-tracking in music research. *Journal of Eye Movement Research*, 11(2):1.

Fox, K. C., Spreng, R. N., Ellamil, M., Andrews-Hanna, J. R., and Christoff, K. (2015). The wandering brain: Meta-analysis of functional neuroimaging studies of mind-wandering and related spontaneous thought processes. *NeuroImage*, 111:611–621.

Galambos, R., Makeig, S., and Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proceedings of the National Academy of Sciences*, 78(4):2643–2647.

Goldstein, R. B., Woods, R. L., and Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine*, 37(7):957–964.

Gramann, K., Ferris, D. P., Gwin, J., and Makeig, S. (2014). Imaging natural cognition in action. *International Journal of Psychophysiology*, 91(1):22–29.

Gramann, K., Hohlefeld, F. U., Gehrke, L., and Klug, M. (2021). Human cortical dynamics during full-body heading changes. *Scientific Reports*, 11(1):18186.

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7.

Greeno, J. G. (1994). Gibson's affordances. *Psychological Review*, 101(2):336–342.

Gwin, J. T., Gramann, K., Makeig, S., and Ferris, D. P. (2011). Electrocortical activity is coupled to gait cycle phase during treadmill walking. *NeuroImage*, 54(2):1289–1296.

Hasson, U. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science*, 303(5664):1634–1640.

Hasson, U. and Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693):20150366.

Hasson, U. and Honey, C. J. (2012). Future trends in Neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, 62(2):1272–1278.

Heimann, K., Umiltà, M. A., Guerra, M., and Gallese, V. (2014). Moving Mirrors: A High-density EEG Study Investigating the Effect of Camera Movements on Motor Cortex Activation during Action Observation. *Journal of Cognitive Neuroscience*, 26(9):2087–2101.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Isik, A. I. and Vessel, E. A. (2019). Continuous ratings of movie watching reveal idiosyncratic dynamics of aesthetic enjoyment. *PLOS ONE*, 14(10):e0223896.

Isik, A. I. and Vessel, E. A. (2021). From visual perception to aesthetic appeal: Brain responses to aesthetically appealing natural landscape movies. *Frontiers in Human Neuroscience*.

Iwasaki, M., Kellinghaus, C., Alexopoulos, A. V., Burgess, R. C., Kumar, A. N., Han, Y. H., Lüders, H. O., and Leigh, R. J. (2005). Effects of eyelid closure, blinks, and eye movements on the electroencephalogram. *Clinical Neurophysiology*, 116(4):878–885.

Johnston, P., Molyneux, R., and Young, A. W. (2015). The N170 observed in the wild: robust event-related potentials to faces in cluttered dynamic visual scenes. *Social Cognitive and Affective Neuroscience*, 10(7):938–944.

Jongkees, B. J. and Colzato, L. S. (2016). Spontaneous eye blink rate as predictor of dopamine-related cognitive functionA review. *Neuroscience & Biobehavioral Reviews*, 71:58–82.

Kam, J. W. Y., Dao, E., Farley, J., Fitzpatrick, K., Smallwood, J., Schooler, J. W., and Handy, T. C. (2011). Slow Fluctuations in Attentional Control of Sensory Cortex. *Journal of Cognitive Neuroscience*, 23(2):460–470.

Kam, J. W. Y., Irving, Z. C., Mills, C., Patel, S., Gopnik, A., and Knight, R. T. (2021). Distinct electrophysiological signatures of task-unrelated and dynamic thoughts. *Proceedings of the National Academy of Sciences*, 118(4):e2011796118.

Kamienkowski, J. E., Ison, M. J., Quiroga, R. Q., and Sigman, M. (2012). Fixation-related potentials in visual search: A combined EEG and eye tracking study. *Journal of Vision*, 12(7):4–4.

Kaunitz, L. N., Kamienkowski, J. E., Varatharajah, A., Sigman, M., Quiroga, R. Q., and Ison, M. J. (2014). Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task. *NeuroImage*, 89:297–305.

Keitel, C., Maess, B., Schröger, E., and Müller, M. M. (2013). Early visual and auditory processing rely on modality-specific attentional resources. *NeuroImage*, 70:240–249.

Kriegeskorte, N. (2008). Representational similarity analysis connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*.

Lacourse, K., Yetton, B., Mednick, S., and Warby, S. C. (2020). Massive online data annotation, crowdsourcing to generate high quality sleep spindle annotations from EEG data. *Scientific Data*, 7(1):190.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4.

Linden, R. D., Picton, T. W., Hamel, G., and Campbell, K. B. (1987). Human auditory steady-state evoked potentials during selective attention. *Electroencephalography and Clinical Neurophysiology*, 66(2):145–159.

Liu, C., Herrup, K., Goto, S., and Shi, B. (2020). Viewing garden scenes: Interaction between gaze behavior and physiological responses. *Journal of Eye Movement Research*, 13(1).

Lopez-Persem, A., Bastin, J., Petton, M., Abitbol, R., Lehongre, K., Adam, C., Navarro, V., Rheims, S., Kahane, P., Domenech, P., and Pessiglione, M. (2020). Four core properties of the human brain valuation system demonstrated in intracranial signals. *Nature Neuroscience*.

Lu, Z. and Ku, Y. (2020). NeuroRA: A Python Toolbox of Representational Analysis From Multi-Modal Neural Data. *Frontiers in Neuroinformatics*, 14:563669.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. A Bradford book. The MIT Press, Cambridge, Massachusetts London, England, second edition edition. OCLC: 896211347.

Luck, S. J., Stewart, A. X., Simmons, A. M., and Rhemtulla, M. (2021). Standardized measurement error: A universal metric of data quality for averaged eventrelated potentials. *Psychophysiology*.

Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T. J., and Poizner, H. (2009). Linking brain, mind and behavior. *International Journal of Psychophysiology*, 73(2):95–100.

McVay, J. C. and Kane, M. J. (2012). Drifting from slow to d'oh!: Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3):525–549.

Meigen, T. and Bach, M. (1999). On the statistical significance of electrophysiological steady-state responses. *Documenta Ophthalmologica*, 98(3):207–232.

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., and Jacobsen, T. (2015). Towards a psychological construct of being moved. *PLOS ONE*, 10(6):e0128451.

Merrifield, C. and Danckert, J. (2014). Characterizing the psychophysiological signature of boredom. *Experimental Brain Research*, 232(2):481–491.

Mitrovic, A., Hegelmaier, L. M., Leder, H., and Pelowski, M. (2020). Does beauty capture the eye, even if it's not (overtly) adaptive? A comparative eye-tracking study of spontaneous attention and visual preference with VAST abstract art. *Acta Psychologica*, 209:103133.

Nikolaev, A. R., Meghanathan, R. N., and van Leeuwen, C. (2016). Combining EEG and eye movement recording in free viewing: Pitfalls and possibilities. *Brain and Cognition*, 107:55–83.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., and Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, 15(6):4.

Oostenveld, R. and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112(4):713–719.

Ossandon, J. P., Helo, A. V., Montefusco-Siegmund, R., and Maldonado, P. E. (2010). Superposition Model Predicts EEG Occipital Activity during Free Viewing of Natural Scenes. *Journal of Neuroscience*, 30(13):4787–4795.

Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., and Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. *Journal of Vision*, 8(14):21–21.

Palomba, D., Angrilli, A., and Mini, A. (1997). Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli. *International Journal of Psychophysiology*, 27(1):55–67.

Patrick, C. J., Bradley, M. M., and Lang, P. J. (1993). Emotion in the criminal psychopath: Startle reflex modulation. *Journal of Abnormal Psychology*, 102(1):82–92.

Pedroni, A., Bahreini, A., and Langer, N. (2019). Automagic: Standardized preprocessing of big EEG data. *NeuroImage*, 200:460–473.

Peirce, J. W. (2007). PsychoPyPsychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2):8–13.

Pernet, C., Garrido, M. I., Gramfort, A., Maurits, N., Michel, C. M., Pang, E., Salmelin, R., Schoffelen, J. M., Valdes-Sosa, P. A., and Puce, A. (2020). Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. *Nature Neuroscience*.

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., and Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, 6(1).

Pernet, C. R., Garrido, M., Gramfort, A., Maurits, N., Michel, C., Pang, E., Salmelin, R., Schoffelen, J. M., Valdes-Sosa, P. A., and Puce, A. (2018). Best Practices in Data Analysis and Sharing in Neuroimaging using MEEG. preprint, Open Science Framework.

Picton, T. W. (2011). *Human auditory evoked potentials*. Plural Publishing Inc., San Diego. OCLC: 1097163932.

Picton, T. W., John, M. S., Dimitrijevic, A., and Purcell, D. (2003). Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. *International Journal of Audiology*, 42(4):177–219.

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., and Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Scientific Reports*, 7(1).

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

Raffaelli, Q., Mills, C., and Christoff, K. (2018). The knowns and unknowns of boredom: a review of the literature. *Experimental Brain Research*, 236(9):2451–2462.

Rammstedt, B., Danner, D., Soto, C. J., and 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.

Ramos Gameiro, R., Kaspar, K., König, S. U., Nordholt, S., and König, P. (2017). Exploration and Exploitation in Natural Viewing Behavior. *Scientific Reports*, 7(1):2311.

Robbins, K. A., Touryan, J., Mullen, T., Kothe, C., and 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.

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. Appleton-Century-Crofts, New York, NY.

Roth, H. L. (2002). Effects of monocular viewing and eye dominance on spatial attention. *Brain*, 125(9):2023–2035.

Saupe, K., Widmann, A., Bendixen, A., Müller, M. M., and Schröger, E. (2009). Effects of intermodal attention on the auditory steady-state response and the event-related potential. *Psychophysiology*, 46(2):321–327.

Schandry, R., Sparrer, B., and Weitkunat, R. (1986). From the heart to the brain: A study of heartbeat contingent scalp potentials. *International Journal of Neuroscience*, 30(4):261–275.

Seabold, S. and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, 3(1):48–51.

Skosnik, P. D., Krishnan, G. P., and ODonnell, B. F. (2007). The effect of selective attention on the gamma-band auditory steady-state response. *Neuroscience Letters*, 420(3):223–228.

Smallwood, J., Beach, E., Schooler, J. W., and Handy, T. C. (2008). Going AWOL in the Brain: Mind Wandering Reduces Cortical Analysis of External Events. *Journal of Cognitive Neuroscience*, 20(3):458–469.

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., and Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, 13(4):657–690.

Smith, T. and Henderson, J. (2010). Attentional synchrony in static and dynamic scenes. *Journal of Vision*, 8(6):773–773.

Smith, T. J. (2013). Watching You Watch Movies: Using Eye Tracking to Inform Cognitive Film Theory. In Shimamura, A. P., editor, *Psychocinematics: exploring cognition at the movies*, page 54. Oxford Univ. Press, Oxford.

Smith, T. J. and Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, 13(8):16–16.

Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., and Trigwell, P. (1995). A scale for the assessment of hedonic tone the SnaithHamilton Pleasure Scale. *British Journal of Psychiatry*, 167(1):99–103.

Soto, C. J. and 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.

Stapells, D. R., Linden, D., Suffield, J. B., Hamel, G., and Picton, T. W. (1984). Human Auditory Steady State Potentials:. *Ear and Hearing*, 5(2):105–113.

Stern, J. A., Boyer, D., and Schroeder, D. (1994). Blink Rate: A Possible Measure of Fatigue. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36(2):285–297.

Struk, A. A., Carriere, J. S. A., Cheyne, J. A., and Danckert, J. (2017). A Short Boredom Proneness Scale: Development and Psychometric Properties. *Assessment*, 24(3):346–359.

Team, R. C. (2018). R: A language and environment for statistical computing.

Thielen, J., Bosch, S. E., van Leeuwen, T. M., van Gerven, M. A. J., and van Lier, R. (2019). Evidence for confounding eye movements under attempted fixation and active viewing in cognitive neuroscience. *Scientific Reports*, 9(1).

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.

Tinio, P. P. L., Smith, J. K., and Smith, L. F. (2013). The walls do speak: psychological aesthetics and the museum experience. In Tinio, P. P. L. and Smith, J. K., editors, *The Cambridge Handbook of the Psychology of Aesthetics and the Arts*, pages 195–218. Cambridge University Press, Cambridge.

Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*, 3(31):1026.

Van Eeckhoutte, M., Luke, R., Wouters, J., and Francart, T. (2018). Stability of Auditory Steady State Responses Over Time:. *Ear and Hearing*, 39(2):260–268.

van Gent, P., Farah, H., van Nes, N., and van Arem, B. (2019). HeartPy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation Research Part F: Traffic Psychology and Behaviour*, 66:368–378.

Vessel, E. A. (2020). Neuroaesthetics. In *Reference Module in Neuroscience and Biobehavioral Psychology*. Elsevier.

Vessel, E. A., Isik, A. I., Belfi, A. M., Stahl, J. L., and Starr, G. G. (2019). The default-mode network represents aesthetic appeal that generalizes across visual domains. *Proceedings of the National Academy of Sciences*, 116(38):19155–19164.

Vessel, E. A., Maurer, N., Denker, A. H., and Starr, G. G. (2018). Stronger shared taste for natural aesthetic domains than for artifacts of human culture. *Cognition*, 179:121–131.

Vessel, E. A. and Rubin, N. (2010). Beauty and the beholder: Highly individual taste for abstract, but not real-world images. *Journal of Vision*, 10(2):1–14.

Vessel, E. A., Starr, G. G., and Rubin, N. (2012). The brain on art: intense aesthetic experience activates the default mode network. *Frontiers in Human Neuroscience*, 6.

Vodrahalli, K., Chen, P.-H., Liang, Y., Baldassano, C., Chen, J., Yong, E., Honey, C., Hasson, U., Ramadge, P., Norman, K. A., and Arora, S. (2018). Mapping between fMRI responses to movies and their natural language annotations. *NeuroImage*, 180:223–231.

Vrana, S. R., Spence, E. L., and Lang, P. J. (1988). The startle probe response: A new measure of emotion? *Journal of Abnormal Psychology*, 97(4):487–491.

Waskom, M. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021.

Winkler, I., Haufe, S., and Tangermann, M. (2011). Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. *Behavioral and Brain Functions*, 7(1):30.

Winton, W. M., Putnam, L. E., and Krauss, R. M. (1984). Facial and autonomic manifestations of the dimensional structure of emotion. *Journal of Experimental Social Psychology*, 20(3):195–216.

Wong, P. K. and Bickford, R. G. (1980). Brain stem auditory evoked potentials: The use of noise estimate. *Electroencephalography and Clinical Neurophysiology*, 50(1-2):25–34.

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., and Deouell, L. Y. (2008). Transient Induced Gamma-Band Response in EEG as a Manifestation of Miniature Saccades. *Neuron*, 58(3):429–441.

Supplementary Material

Participant demographics

See Tabs. 2 and 3 for full sample demographics.

Log transformation of ASSR SNR

SNR values were log transformed to shift them from a skewed gamma to a more gaussian distribution (see Fig. 5). We applied the natural logarithm to average SNR values over all EEG channels for each participant and each trial.

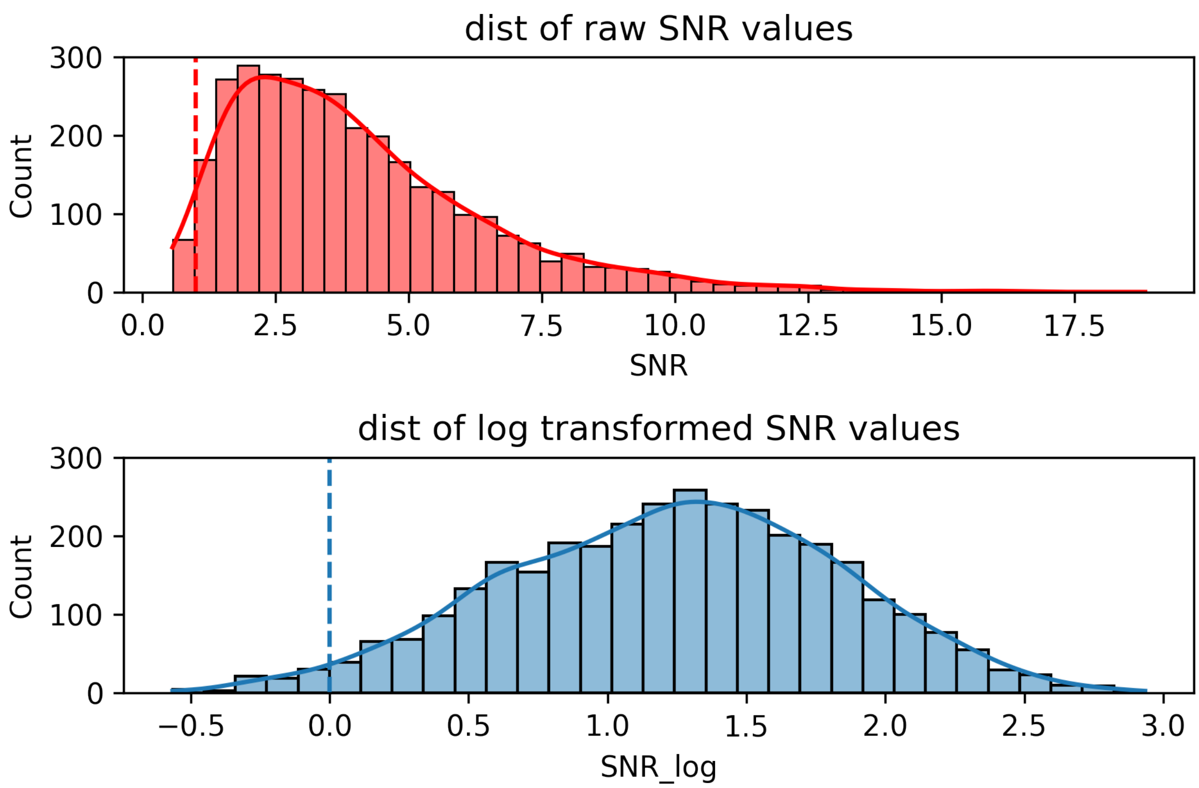

Figure 5: Distribution of raw and log transformed SNR values across all participants (N=43)

Full ANOVA tables

Results of all repeated measures ANOVA models in the study are compiled in Tab. 4.

Trial wise correlation of all dependent measures

See Tab. 5

Table 2: Categorical demographic factors

Variable	n
Sex	
Female	26
Male	17
Other	0
Handedness	
Right	39
Left	2
Ambidextrous	2
Eye dominance	
Right	33
Left	6
No dominance	4
Highest degree of education	
Mittlere Reife	2
Ausbildung	1
Abitur	24
Studium	16
Mental disorder	
No	34
Yes	9
Neurological disorder	
No	42
Yes	1

Table 3: Continuous demographic factors

Variable	n	M	SD	min	max	25%	50%	75%
Age	43	27.1	7.06	19	52	22.5	25	30
Years of education	43	17.8	3.55	9	25	15	18	20
Caffeine intake [mg/kg]	43	0.84	1.041	0.00	4.28	0.00	0.52	1.43
BFI (range 3-15)								
extraversion	43	9.30	1.897	5	13	8	9	10.5
open mindedness	43	12.14	1.684	8	15	11	13	13
agreeableness	43	11.28	1.894	7	15	10	12	12.5
conscientiousness	43	10.65	2.516	5	15	9	11	13
negative emotionality	43	8.23	2.983	3	15	6	8	10
PANAS (range 6-30)								
positive	43	17.60	4.588	9	25	14	18	21
negative	43	7.47	2.364	6	15	6	6	7.5
SHAPS (range 0-14)	43	1.02	1.318	0	5	0	1	2
BPS (range 8-56)	43	19.07	8.213	8	49	15.5	17	21.5

Table 4: full ANOVA results

Measure	n	M	SD	$F(1,42)$	η_p^2
ASSR SNR	344	1.430	0.499		
fixation task				6.24*	0.13
stimulus dynamics				0.81	0.02
stimulus content				2.47	0.06
dynamics x task				1.83	0.04
dynamics x content				2.04	0.05
content x task				0.01	0.00
dynamics x content x task				0.94	0.02
AESTHETIC RATING	344	0.18	0.32		
fixation task				4.02	0.09
stimulus dynamics				29.72***	0.41
stimulus content				50.21***	0.54
dynamics x task				2.63	0.06
dynamics x content				18.50***	0.31
content x task				0.02	0.00
dynamics x content x task				3.29	0.07
BOREDOM RATING	344	-0.19	0.36		
fixation task				2.16	0.05
stimulus dynamics				41.18***	0.50
stimulus content				1.30	0.03
dynamics x task				0.15	0.00
dynamics x content				18.92***	0.31
content x task				0.97	0.02
dynamics x content x task				0.03	0.00
EYEBLINKS	344	0.51	0.78		
fixation task				27.46***	0.40
stimulus dynamics				4.96*	0.11
stimulus content				3.63	0.08
dynamics x task				0.90	0.02
dynamics x content				9.89**	0.19
content x task				2.49	0.06
dynamics x content x task				3.46	0.08
SACCADES	344	6.91	6.25		
fixation task				420.17***	0.91
stimulus dynamics				152.88***	0.78
stimulus content				135.02***	0.76
dynamics x task				125.92***	0.75
dynamics x content				58.36***	0.58
content x task				113.23***	0.73
dynamics x content x task				28.85***	0.41
MICROSACCADES	344	8.67	4.54		
fixation task				0.45	0.01
stimulus dynamics				79.49***	0.65
stimulus content				25.43*	0.11
dynamics x task				0.86	0.02
dynamics x content				24.13***	0.36
content x task				31.25***	0.43
dynamics x content x task				7.83**	0.16
HEART RATE DECELERATION	344	-3.55	2.00		
fixation task				3.90	0.08
stimulus dynamics				11.24**	0.21
stimulus content				0.13	0.00
dynamics x task				0.30	0.01
dynamics x content				0.34	0.01
content x task				1.07	0.02
dynamics x content x task				0.53	0.01

Table 5: full correlation table for trialwise measures

Measure	n	M	SD	1	2	3	4	5	6
ASSR SNR	3349	1.43	0.64	–					
Ratings									
aesthetic	3349	.19	.53	$-.046$	–				
boredom	3349	-.19	.58	.038	$-.580^{***}$	–			
Eye movements									
eye blinks	3189	0.54	1.06	$-.053^{*}$	$-.059^{*}$	$.078^{***}$	–		
saccades	3349	7.03	6.91	$-.041$	.018	.029	–	–	
microsaccades	3349	8.64	6.01	.000	$-.086^{***}$	$.126^{***}$	–	–	–
Heart rate change (bpm)	3336	-3.57	4.48	$-.036$	.005	$-.010$	.049	$.059^{*}$	$-.000$

Note: r values computed using repeated measures correlation (Bakdash and Marusich, 2017). $^{*}p < .05$, $^{***}p < .001$. p values corrected using Holm's method (Holm, 1979). n trials for eye blinks is lower because two participants exhibited zero blinks over all trials, which prevented the model to converge; n trials for HR deceleration is lower because outliers were rejected. Correlation between the different eye measures was not investigated.