creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

뇌인지과학석사학위논문

Awe is characterized as an ambivalent experience in the human behavior and cortex

행동 및 피질에서 양가적 경험으로 특정되는 경외감

2024 년 8 월

서울대학교 대학원

뇌인지과학과 뇌인지과학 전공

이 진 우

Awe is characterized as an ambivalent experience in the human behavior and cortex

지도교수 차 지 욱

이 논문을 뇌인지과학석사 학위논문으로 제출함

2024 년 8 월

서울대학교 대학원

뇌인지과학과 뇌인지과학 전공

이 진 우

이진우의 뇌인지과학석사 학위논문을 인준함

2024 년 8 월

위 원 장 Sang Ah Lee (인)

부위원장 차 지 욱 (인)

위 원 우 충 완 (인)

Abstract

The ambivalent nature has been emphasized as a unique quality of awe in qualitative descriptions and cited as a potential source of its various psychiatric and psychosocial benefits. However, due to the affective science's dichotomic positive/negative schema, this ambivalence has not been fully explored. This study aims to capture the valence dynamics inherent in awe by applying naturalistic VR-EEG paradigm and an extended valence measurement allowing ambivalent response. We investigated whether awe can be better characterized as ambivalent rather than simply positive or negative at both behavioral and cortical levels. Behaviorally, the awe intensity for each clip was precisely predicted by the duration and intensity of ambivalent feelings, but not by other valence metrics. In the cortical level, ambivalent feelings during awe showed unique neural representations in the latent cortical space, with significant individual variability in their distinctiveness from positive/negative representations. Nevertheless, the more distinctly ambivalent feelings were encoded in the cortex, the stronger individuals reported awe. Finally, frontal delta band power was mainly involved in distinguishing different valence representations in the cortices. This study not only explores the existence of unique neural representations of ambivalent feelings, a topic of debate in affective neuroscience, but also demonstrates that awe can be characterized as an ambivalent experience at both behavioral and cortical levels.

Keywords: Awe, Ambivalence, Latent cortical space, Electroencephalogram, Virtual Reality

Student Number: 2022-23358

Contents

List of Tables

List of Figures

Introduction

Awe is an intricate emotion evoked by facing something so enigmatic that individuals cannot get a sense. Such characteristics of awe make it appear similar to emotions such as fear; however, awe is distinguished by the fact that it is accompanied by the expansion of one's conceptual schemes in an attempt to comprehend something mysterious, incorporating not only overwhelming but also pleasant feelings. For these reasons, psychologists defined awe with its two key dimensions: 'perceived vastness' and 'a need for accommodation' (Keltner & Haidt, 2003). For instance, in the phenomenological study of awe (Yaden et al., 2016), astronauts described their awe experience during the space flight that they were bewildered by the vast scale of the universe in contrast to the smallness of Earth (i.e., perceived vastness), yet at the same time, they also felt ineffable beauty and fragility of Earth and realized that humankind's urgent task is to preserve this beauty (i.e., a need for accommodation). These multifaceted dimensions of awe shape its ambivalent nature, and the coexistence of opposing feelings in awe has been regarded as potential sources of its psychiatric, psychosocial, and intellectual benefits such as stress resilience, non-egocentric perspectives, and trait openness (Jiang et al., 2024). Thus, early awe studies asserted that "an adequate account of awe must explain how awe can be both profoundly positive and terrifyingly negative" (Keltner & Haidt, 2003).

Nevertheless, recent affective sciences have tried to characterize awe as a single-valence emotion. For instance, awe was split into two subtypes in terms of its dominant valence: positive awe and threat awe (Gordon et al., 2017; Piff et al., 2015). Based on this framework, neuroimaging studies reported distinguishable neural correlates between these two types of awe in terms of

structural (Guan et al., 2019) and functional patterns (Takano & Nomura, 2022). However, this approach does not fully address ambivalent nature of awe.

We diagnose that methodological issues regarding affective valence are limiting the research on the ambivalence of awe. Conventional measurement of valence such as the 1D bipolar continuum model (Russell, 2003) does not allow ambivalent responses. The unidimensional structure of this scale is highly problematic since it lacks behavioral and neurobiological plausibility of valence representation. Numerous psychometric studies observed that separate positivity and negativity dimensions displayed stronger predictive power than a unidimensional model of valence, and also positivity and negativity did not show negative correlation, challenging unidimensional assumptions of valence (An et al., 2017; Briesemeister et al., 2012; Cacioppo & Berntson, 1994; Moeller et al., 2018). From a neurobiological aspect, the neural circuits encoding positive and negative feelings share some common components but fundamentally operate in distinct ways, supporting multidimensional model of valence (Berridge, 2019; Lammel et al., 2012; Norman et al., 2011; Reynolds & Berridge, 2008). For example, while both circuits share the ventral tegmental area (VTA) as a common part, reward-VTA circuit receive inputs from the laterodorsal tegmentum but aversion-VTA circuit from the lateral habenula (Lammel et al., 2012). Regarding the awe, a recent study has found that threat-awe inducing images led higher co-occurrence between opposing valence compared to happy and fear images by using 2D measurement of valence (Chaudhury et al., 2022), implying that multidimensional valence scale can facilitate to investigate ambivalence of awe experience.

Additionally, we suggest that theoretical debates about the distinct neural representation for ambivalent feelings may act as another bottleneck for research on the ambivalent of awe. The constructive perspectives of emotion have asserted that ambivalent feelings just originated from

the rapid fluctuation between conflicting valence in the brain, arguing the absence of distinct neural pattern of ambivalent feelings (Barrett & Bliss-Moreau, 2009; Russell, 2017). Contrarily, recent studies support the uniqueness of ambivalent feelings in the cortex. Vaccaro et al. (2020) purposed that while in the subcortical regions centered around the brainstem and limbic system, opposing valences are rapidly co-regulated to preserve homeostasis, resulting coarse fluctuations, cortical areas such as the anterior insula cortex integrate these dynamics to produce a global 'mixed' affective representation. The constructivism has exerted a significant influence in this controversy, hindering systematic research on the ambivalence of awe. However, recent human fMRI studies support the latter. For example, the posterior-anterior axis gradient within the right temporoparietal cortex is associated with valence co-occurrence during movie watching (Lettieri et al., 2019). The ventromedial prefrontal cortex and the anterior cingulate cortex also exhibited consistent neural pattern for ambivalent feelings during movie watching (Vaccaro et al., 2024). These observations motivate the possibility that ambivalent feelings during awe experience are represented at the cortical level in a manner that is significantly segregated from the neural representation of simply positive or negative feelings.

Then, how can we identify neural representation of ambivalent feelings during awe experience? First, to induce more naturalistic awe experience in the laboratory, we designed 360° immersive clips in the virtual reality (VR). Some concerns about conventional image and movie stimuli for awe studies emerged due to their lack of ecological validity (Chirico et al., 2016; Silvia et al., 2015). Considering that 'perceived vastness' is one of the main key dimensions of awe, limited magnitude of these stimuli makes it elusive which emotion they trigger. To overcome this limitation, we developed VR videos based on the strengths demonstrated by VR protocols in eliciting awe (Chirico et al., 2024; Chirico et al., 2017; Chirico et al., 2018; Kahn & Cargile, 2021;

Quesnel & Riecke, 2018). Second, we recorded participants' electroencephalogram (EEG) signals while they watched VR clips. The insula synthesizes fluctuating bodily signals within a time window of approximately 125 ms to create a global affective representation (Picard & Craig, 2009; Vaccaro et al., 2020; Wittmann, 2013), implying that neuroimaging techniques with a sampling rate higher than 16 Hz can fully capture these dynamics. Hence, we chose EEG recordings instead of the other modalities such as fMRI. Last, we applied deep learning techniques to construct individualized latent valence-cortical space instead of conventional hand-crafted feature extraction approach. Previously, frontal alpha asymmetry (FAA) was understood as an electrophysiological index of valence (Berkman & Lieberman, 2010; Schmidt & Trainor, 2001). However, recent studies consistently reported that FAA correlates with motivational behavior rather than valence per se, implying the limited specificity of FAA as a valence representation (Gable & Harmon-Jones, 2010; Harmon-Jones & Gable, 2018; Honk & Schutter, 2006; Wacker et al., 2003). Leveraging notable representation learning ability of deep neural networks, we learned valence-specific latent neural space within individual-stimulus level, which contrasted EEG samples in terms of their valence states. Given that the architecture of valence representation in the brain displays large heterogeneity across individuals and sensory information (Čeko et al., 2022; Lee et al., 2024; Lettieri et al., 2024), our within individual-stimulus approach is expected to capture not only these variabilities but also commonality of the latent valence-cortical architecture.

In this study, we aimed to examine whether awe is more precisely characterized as an ambivalent experience than single-valence ones at the behavioral and cortical level. For this end, we formulated three research questions and corresponding hypotheses as follows: First, is awe intensity predicted by ambivalence-related behavior metrics more precisely than single valence features? We hypothesize that ambivalence-related features predict awe intensity score of each

clip more accurately than single valence ones. Second, does ambivalent feeling during awe experience have distinct neural representation in the latent cortical space? We expect that ambivalent feeling exhibits distinguishable cortical representation from single valence states. Third, does the distinctive neural representation predict the awe intensity score? We predicted that the more distinctively ambivalent feeling is represented in the latent cortical space, the more saliently individuals feel awe during VR watching.

Methods

Participants

We recruited 50 healthy young adult Koreans enrolled in psychology courses at Seoul National University for this study. Participants were excluded if they met any of the following criteria: (1) currently taking psychiatric medication, (2) history of psychiatric treatment, (3) left-handedness, (4) vestibular neuritis or balance disorder, (5) visual acuity before correction less than 0.2, (6) non-Korean native speakers, and (7) consumption of alcohol or use of hair rinse 24 hours before the experiment. Data from 43 participants were completely collected in the analyses (23 females; $M_{\text{age}} = 20.2$ years, $SD_{\text{age}} = 1.7$ years). Seven participants were excluded due to technical issues ($N = 3$), discontinuance due to motion sickness ($N = 2$), and lack of fidelity in VR watching task ($N = 2$). See **Figure 1a** for sampling procedures. Participants provided written informed consent before the experiment, and all procedures were approved by the Institutional Review Board of Seoul National University.

Experimental paradigm

VR clip design We collaborated with a professional filmmaker to design four audio-integrated 360° immersive videos using Unreal Engine (version 5.03). Each video lasted 120 seconds. Three of the videos were designed to evoke awe: *Space* (SP), *City* (CI), and *Mountain* (MO), while the other one, *Park* (PA), served as a control stimulus, designed not to elicit any specific emotional response. To investigate whether ambivalence is consistently observed in various awe experiences,

we varied (1) the clip themes, (2) the sub-components of awe, and (3) the perceptual features across the awe-inducing videos.

Firstly, considering that awe is most intensively triggered by massive landscapes (Chirico et al., 2018; Keltner & Haidt, 2003; Shiota et al., 2007; Yaden et al., 2019), we differentiated the semantic theme of the scenery: SP featured supernatural landscapes (i.e., black holes and planets in space), CI showcased urban landscapes (i.e., cityscape viewed from the top of skyscrapers), and MO depicted natural one (i.e., mountain scenery).

Secondly, following Chirico et al. (2018)'s qualitative framework to design videos that effectively elicit awe in VR, we aimed to represent the two key dimensions of awe, perceived vastness and a need for accommodation, through different cues in each video. Each video was designed so that perceivers would first experience vastness during the initial 60 seconds and then feel a need for accommodation in the latter 60 seconds. For example, in SP, participants watched a giant black hole approaching, consuming everything, and ultimately drawing them in, followed by the sudden appearance of Earth from space. Nevertheless, different sub-components were applied to realize each dimension across clips. Perceived vastness can be induced through perceptual (e.g., 'width' and 'height') and conceptual cues (e.g., 'complexity')(Chirico et al., 2017; Chirico et al., 2018). MO was designed to evoke vastness through perceptual width, CI through height, and SP through conceptual complexity. For the need for accommodation, we introduced surprise cues in each video around the 60-second mark as a trigger of accommodation (Chirico et al., 2017; Chirico et al., 2018), tailored to the context of each video to ensure that the cause of surprise did not overlap across videos. The design of three awe-inducing clips is summarized in **Table 1**.

Table 1. Design of awe-inducing VR clips for diverse awe experiences based on the framework of Chirico et al. (2018)

	0 – 60 secs			**60 - 120 secs**		
Dimensions	**Perceived vastness**			**A need for accommodation**		
Sub-components	perceptual		conceptual	surprise		
VR cues	width	height	complexity	Sudden transition from inside a cave to a mountain peak view	An elevator rapidly ascending	Abrupt adsorption into a black hole
Clip	**MO**	**CI**	**SP**	**MO**	**CI**	**SP**
Theme	natural panorama	urban cityscape	supernatural scenery	natural panorama	urban cityscape	supernatural scenery

Lastly, to prevent ambivalent feelings from being driven by specific perceptual factors, we intentionally composed the three awe videos with different audiovisual information. We synchronized visual content with ambient sounds using open-source audio samples from Freesound (https://freesound.org) and GarageBand (version 10.4.6). To verify our design, we calculated three perceptual features known to predict perceivers' emotional responses – brightness, hue, and loudness (Chua et al., 2022; Thao et al., 2019) – every second for each stimulus and visualized their time-course dynamics. We qualitatively confirmed that each video exhibited very different temporal dynamics for all features (see **Supplementary Figure 1**).

To validate the awe elicitation, we conducted a preliminary study with 28 independent young adult Koreans (five females; M_{age} = 20.2 years, SD_{age} = 1.9 years), who rated awe intensity using the Awe Experience Scale (Yaden et al., 2019) after watching each clip. Participants reported significantly higher awe scores for three awe clips compared to the control clip, with large effect sizes in two-sided paired t-tests (SP-PA: Cohen's d = 2.466, $P_{FDR} = 8\times10^{-13}$; CI-PA: Cohen's d = 2.193, $P_{FDR} = 6\times10^{-12}$; MO-PA; Cohen's d = 1.52, $P_{FDR} = 1\times10^{-6}$).

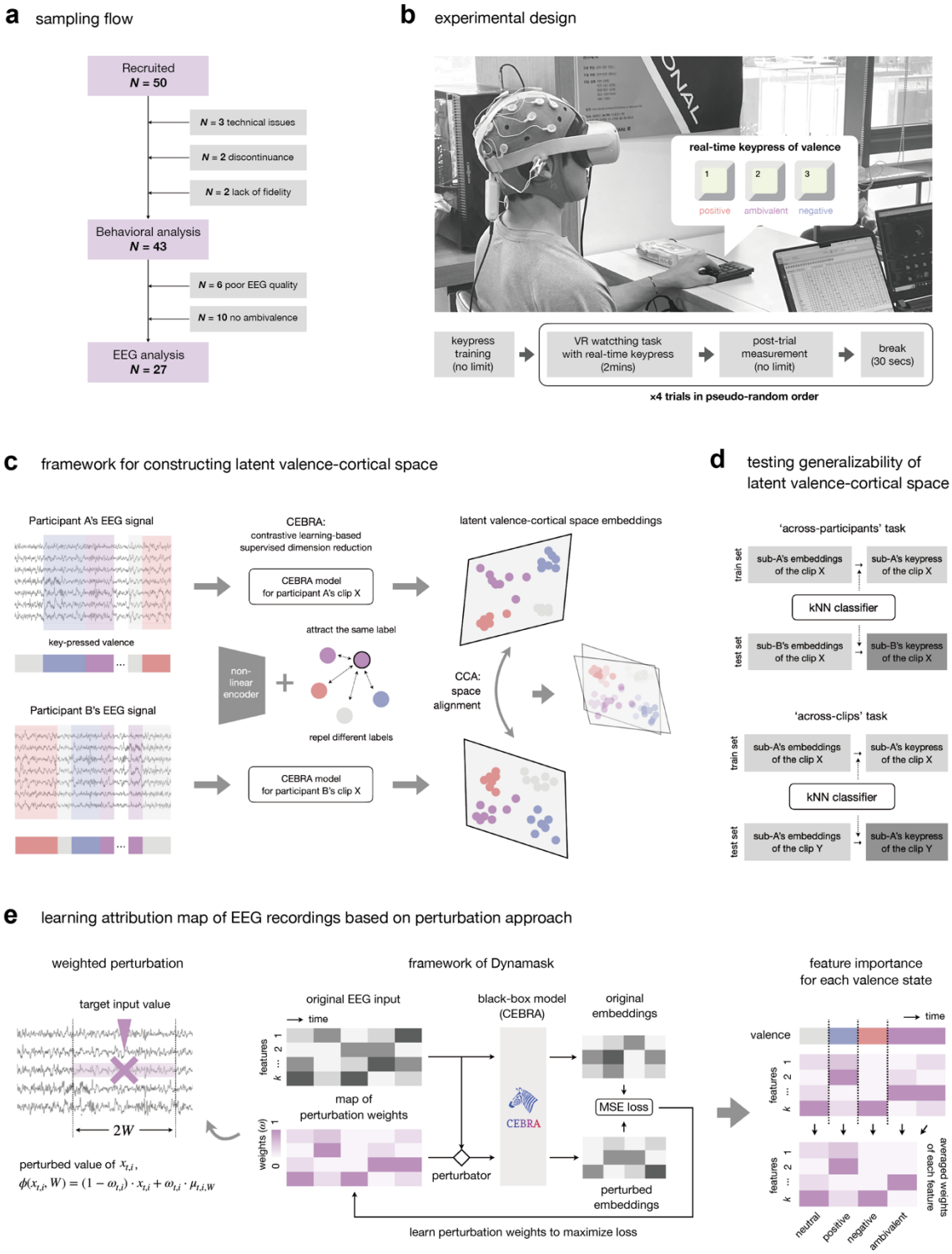


Figure 1. Experimental and methodological frameworks a, sampling procedures. **b,** integrated VR-EEG protocol and experimental procedures. **c,** construction of individualized latent valence-cortical spaces

using CEBRA and neural alignment. **d,** assessing generalizability of the individualized latent valence-cortical spaces across individuals and stimuli through two types of predictive tasks. **e,** inference of attribution map to construct latent valence-cortical space using Dynamask.

Baseline self-report Before the experiment, participants provided information on their sex, age, baseline mood states, and dispositional trait of experiencing awe in daily life. Baseline mood states were assessed using the Korean version of the Positive and Negative Affect Schedule (PANAS) validated by Lim et al. (2010). No participants showed exceptional mood states beyond 1.5 × interquartile range ($M_{\text{positive}} = 33.581$, $SD_{\text{positive}} = 6.284$; $M_{\text{negative}} = 23.140$; $SD_{\text{negative}} = 6.331$). Awe trait was measured using the awe-related items from the Dispositional Positive Emotions Scale (Shiota et al., 2006) translated into Korean ($M = 29.628$, $SD = 6.626$). The Korean-translated DPES items demonstrated acceptable reliability (Cronbach's $\alpha = .827$) but mixed results in internal validity (model fit of the original factor structure: CFI = .946, RMSEA = .127).

Procedures Participants sat on a sofa in a noise-isolated room and wore an Enobio 20 EEG device (Neuroelectrics) and a Quest 2 VR headset (Oculus). After checking EEG signal quality, the experiment proceeded as follows: baseline EEG recording, keypress training, VR watching task, post-trial measurement, and a break (see **Figure 1b**). Firstly, participants' resting EEG signals were recorded for 120 seconds with their eyes closed (baseline recording). These resting signals were used to normalize signals recorded during the movie-watching trials. Secondly, they practiced real-time valence keypress reporting (keypress training). Participants were explicitly asked to report their valence state in real-time manner with the following auditory instruction:

> *"While watching the video, please report your affective state in real-time by pressing a number pad: '1' for positive, '2' for ambivalent (feeling positive and negative feelings at the same time), and '3' for negative feelings. If you do not feel any affective feelings, please do not press anything. If specific state persists, continue to press and hold the corresponding key. It is important to report your subjective reactions, rather than which emotions the video intends to elicit".*

Participants practiced this for 60 seconds using sentences describing affective responses to prevent confusions which key they should press. Then, participants watched one VR clip in a pseudorandom order per trial, reporting their valence states in real time using keypress (VR watching task). After each trial, they reported awe intensity, overall valence, arousal, and motion sickness using controllers (post-trial measurement). Awe intensity was measured by the Korean-translated AWES (Yaden et al., 2019), valence by Evaluative Space Grid (Larsen et al., 2009), arousal by conventional 9-point Likert scale (Bradley & Lang, 1994), and motion sickness by a single 7-point Likert scale item. The Korean-translated AWES demonstrated acceptable psychometric properties (Cronbach's α = .928; model fit of the original factor structure: CFI = .881, RMSEA = .079). Participants took a 30-second break with eyes closed after each trial (break).

EEG recording and preprocessing We recorded EEG signals using 19 dry electrodes: AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC6, C3, Cz, C4, P7, P3, Pz, P4, P8, O1 and O2 with Neuroelectrics Enobio 20. Ground and reference electrodes were attached to the right earlobe. The embedded

software in the Enobio system assessed signal quality and visualized it using three channel colors: green (good), yellow (medium), and red (bad). We ensured no electrodes displayed red signals before starting the signal acquisition. We adopted the automated preprocessing pipeline validated by Delorme (2023). EEG signals for each trial were time-locked to the initiation of the video, excluding the last three seconds to avoid end-of-task effects (e.g., loss of attention or emotional confounding). High-pass filtering above 0.5 Hz and Artifact Subspace Reconstruction were performed. Unlike the original pipeline, we used interpolation to maintain consistent recording lengths across participants and trials instead of exclusion of time window with poor signal quality. We conducted independent component analysis-based artifact rejection to remove noise components, such as eye movement, muscle noise, or skin potentials, with over 90% probability. The preprocessed signal for each trial was normalized by subtracting the average resting signal value for each channel. All preprocessing was performed using the "EEGLAB" plugin (Delorme & Makeig, 2004) in MATLAB (version 2021a).

Short time/Fast Fourier transform With preprocessed and normalized EEG signals, we performed short time Fourier transform (STFT) and fast Fourier transform (FFT) to calculate the spectral power of five frequency bands for each channel: delta (1-4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-31 Hz), and gamma (31-49 Hz). For STFT, a Hanning window with a 500-sample window size (i.e., 1 sec) and a 250-sample hop size was applied. Participants' valence keypresses were embedded as event markers in EEG signals, categorizing EEG samples into one of four valence categories. The valence label of each 500-sample window after STFT was defined as the mode of the corresponding samples. FFT was also performed to calculate overall spectral powers marginalized across the whole time-series. Using the Welch method, we calculated the

power spectral density for each EEG channel, and then integrated it over specific frequency range described above to determine the band power. For relative band power, we normalized the power within each band by the total power across all frequencies. The "scipy" package (Virtanen et al., 2020) in Python (version 3.8) was used for STFT and FFT.

Behavioral analysis

Univariate statistical analysis Using data from 43 participants, we firstly assessed the univariate association between AWES ratings and 14 behavioral features measured before, during, and after each trial: sex, age, PANAS positive score, PANAS negative score, DPES awe score (before trial), duration of positive, ambivalent, negative, and neutral feelings (during trial), arousal, motion sickness score, and intensity of positive, ambivalent, and negative feelings (after trial). The duration of each valence type was calculated as the ratio of keypresses for that valence type to the total running time of each clip. Intensity was calculated based on the Evaluative Space Grid responses: positivity (x-axis value), negativity (y-axis value), and ambivalence (minimum value between positivity and negativity, following previous literature - e.g., (Berrios et al., 2015; Chaudhury et al., 2022; Ersner-Hershfield et al., 2008)). Firstly, we performed two-sided paired t-tests to examine statistical differences in AWES scores, duration, and intensity of each valence type between the three awe clips and the control clip at $P_{\mathrm{FDR}} < .05$. Next, to evaluate the explanatory power of the 14 metrics, we fit linear mixed effect models with each regressor and two random intercepts for participants and clips using the "lmerTest" package (Kuznetsova et al., 2015). Assumptions of normality were examined using the "DHARMa" package (Hartig, 2018). We confirmed that the distribution of residuals did not significantly deviate from a normal

distribution using the Kolmogorov-Smirnov test (all Ps $> .200$). All statistical analyses were conducted in R studio (version 2023.03.1+446).

Multivariate machine learning analysis Next, we conducted machine learning-based predictive modeling with 14 behavioral variables for AWES scores as multivariate analysis, considering potential non-linear interactions among features. Using "h2o" package (LeDell & Poirier, 2020), we split the dataset into training and test sets with a 4:1 ratio and conducted 5-fold cross-validation. A total of 22 models were constructed, and we selected the best model based on the lowest fold-averaged RMSE from the test set. Models that did not provide feature importance information (e.g., stacked ensemble models) were excluded for interpretability. As a result, gradient boost model (GBM) was chosen as the best model. The predictive performance was compared to a baseline ridge linear regression model without any interaction terms using for metrics: RMSE, MAE, MSE, and R^2. To identify the most influential features, we calculated feature importance and shapley values for each variable. All machine learning analyses were performed in R studio (version 2023.03.1+446).

Electrophysiological analysis

Construction of latent valence-cortical space Among 43 participants in the behavioral analysis, we excluded 16 individuals due to poor-quality preprocessed EEG signals ($N = 6$) and lack of ambivalent keypresses in at least one trial ($N = 10$; see **Figure 1a**). The quality of preprocessed signals was visually inspected. The primary objective of the electrophysiological analysis was to investigate whether ambivalent feelings during awe have distinct cortical

representation and to evaluate their relationship with the intensity of awe. Therefore, participants who responded ambivalent feelings for less than 5% of the total duration across all videos were excluded. Given that our behavioral analysis identified the duration of ambivalent feelings as the most salient positive predictor of awe intensity (see 'Results'), such exclusion is not expected to introduce sampling bias regarding the measurement of awe intensity.

Using 27 participants' EEG signals and valence keypress in the three awe clip trials, we constructed a latent valence-cortical space for each individual-trial design (i.e., total 81 latent spaces), using the "CEBRA" package (Schneider et al., 2023). CEBRA employs supervised contrastive learning to extract latent embeddings from the input data (i.e., STFT-processed EEG signals here), maximizing the attraction of EEG samples with the same valence labels and repelling those with different labels (see **Figure 1c**). As the dimensionality of latent spaces was elusive, we fitted CEBRA models for each participant-clip pair across dimensions ranging from one to nine. The following hyperparameters were applied: batch_size = {length of STFT EEG signals}, model_architecture = 'offset-10 model', number_of_hidden_units = 38, learning_rate = .001, the number_of_iterations = 500, and hybrid = False.

Validation of latent valence-cortical space with predictive tasks To test whether the individualized latent spaces hold significant information generalizable across different individuals and clips, we conducted a pairwise prediction task using a 2 tasks × 3 conditions design (see **Figure 1d**). For the 'across participants' task, a classifier trained on participant A's latent neural space embeddings and valence labels was used to predict the valence keypress of participant B's embeddings for the same clip. For the 'across clip' task, a classifier trained on clip X's embeddings

and labeled valence types was used to predict the valence of clip Y's embeddings within the same participant. Predictions were evaluated under three conditions: (1) a baseline null test with shuffled training valence keypress labels ('random'), (2) prediction using personalized latent neural embeddings without any alignment ('not aligned'), and (3) prediction using aligned embeddings between train and test embeddings ('aligned'). Neural alignment motivates to explore commonality among individual latent neural spaces (Gallego et al., 2020; Safaie et al., 2023). For the alignment, canonical correlation anlaysis (CCA) between train and test embeddings was employed using the "sklearn" package (Pedregosa et al., 2011).

In these prediction tasks, six participants who pressed all valence labels for at least 5% of the total duration across three clips were selected to facilitate the multi-label classification. A k-nearest neighbors (kNN) classifier with a neighborhood parameter of 15 was used without tuning (i.e., the nearest odd number to the square root of the input embedding length following the conventional heuristic). Considering imbalance in valence keypress labels, prediction performance was evaluated using the weighted F1 scores. Pairwise post-hoc comparisons at $P_{\mathrm{FDR}} < .05$ were conducted to compared predictive performances across the three conditions. Construction of latent neural spaces and prediction was performed in Python (version 3.8).

Dimensionality selection for the latent valence-cortical space We selected the optimal dimensionality of the latent valence-cortical space based on two assumptions: (1) The space shares the same dimensionality across individuals and clips. (2) The space displayed as good as predictive performances across individuals and clips with other neural spaces based on higher dimensionality even with fewer dimensions. The second assumption was based on the idea that generalizability, as reflected in predictive performance, should be the criterion for determining the canonical

dimensions in dimensionality reduction techniques (Cunningham & Yu, 2014). However, it also acknowledges that such predictive performance may be overestimated as the number of dimensions increases (Cunningham & Yu, 2014; Diaconis & Freedman, 1984). We calculated the mean weighted F1 score for each dimension in both predictive tasks and performed hierarchical clustering analysis to group dimensions with similar prediction performance. The cluster with the highest silhouette coefficient was chosen, and the lowest dimension in the highest-performing cluster was selected. Dimensions 6, 7, 8, and 9 formed the high-performance cluster for the across-participant task, and dimensions 7, 8, and 9 for the across-clip task. Thus, 7 dimensions were chosen as the canonical dimension. Hierarchical clustering analysis was conducted using the "cluster" package (Maechler et al., 2013) in R studio (version 2023.03.1+446).

Comparing CEBRA-, PCA-, and FAA-driven embeddings For fair comparison of the predictive power of our CEBRA-based latent valence-cortical embeddings, we compared its performance with principle component analysis (PCA) and FAA-driven embeddings (i.e., PCA: conventional linear and unsupervised dimensionality reduction approach; FAA; hand-crafted valence-related EEG metrics). First, to extract PCA embeddings, we input only the STFT-processed EEG data, excluding the valence keypress, and computed latent embeddings with dimensions ranging from one to nine. We identified the 6-dimensional embedding as the optimal latent space due to its modest predictive performance with the fewest dimensions (see **Supplementary Figure 2**). Second, FAA embeddings were defined as the difference in alpha band power between the F4 and F3 channels for each timepoint in the STFT-featured EEG sequence. These two channels were selected based on previous studies (Brzezicka et al., 2017;

Quaedflieg et al., 2016; Van Der Vinne et al., 2017). The prediction tests with a 2×3 design were conducted using these three types of embeddings: CEBRA, PCA, and FAA. Test performances under the three conditions – random, not aligned, and aligned – were compared across CEBRA, PCA, and FAA-based embeddings. Additionally, within the CEBRA embeddings, performances were compared across the different conditions. To test the statistical differences in weighted F1 scores, two-sided paired t-tests were conducted at $P_{\mathrm{FDR}} < .05$. PCA was performed using the "sklearn" (Pedregosa et al., 2011) package in Python (version 3.8).

Assessing significance of cortical valence representation Using the chosen 7D latent CEBRA valence-cortical embeddings, we measured the segregation of ambivalent EEG samples from other valence samples using silhouette coefficients. While InfoNCE loss value could quantify contrast performance too, it lacks scaling and valence type-specific calculations, so we used silhouette coefficients instead. We computed the average silhouette coefficient for ambivalent-labeled EEG samples for each participant-clip latent space. Since silhouette coefficients can be overestimated under the latent space constructed in supervised manner, we assessed its statistical significance through permutation test. We randomly shuffled valence keypresses and trained CEBRA model with the original STFT-featured EEG signals and permuted valence sequence based on the identical hyperparameter set. Average silhouette coefficients of ambivalence-labeled EEG samples were extracted from the trained pseudo-embeddings, and we obtained its null distribution by repeating this 1,000 times. P-value calculated from the permutation test, P_{perm}, is formulated as follows:

$$P_{perm} = \frac{N_{s^* > s}}{N_{s^*}} \tag{1}$$

where s^* and s are average silhouette coefficients calculated from permuted and original latent spaces, and N_{s^*} is the number of s^* in the null distribution. All P_{perm} values were FDR corrected. We performed the same analysis with positive and negative valence samples.

Quantifying 'cortical distinctiveness' of each valence type We developed a metric called 'cortical distinctiveness', ϕ_k, indicating how distinguishable a reference valence cluster k is from the other valence clusters in the latent valence-cortical space. ϕ_k is defined as:

$$\phi_k = \frac{1}{N}\sum_{i=1}^{N} d(k, c_i) \tag{2}$$

where N is the number of other clusters, c_i is the i-th valence cluster, and $d(k,\ c_i)$ is the cosine distance between the cluster k and c_i. We applied cosine distance as a metric of cluster distance instead of other conventional metrics (e.g., Euclidean distance), considering that latent CEBRA embeddings are distributed on the hypersphere space. We initially measured $d(k,\ c_i)$ based on the average cluster distance. Average distance between the cluster k and c_i is calculated as the mean cosine distance between each point in k to every point in c_i as follows:

$$d(k, c_i) = \frac{1}{n_k n_{c_i}}\sum_{p=1}^{n_k}\sum_{q=1}^{n_{c_i}}\left(1 - \frac{\overrightarrow{x_p}\cdot\overrightarrow{y_q}}{\left\|\overrightarrow{x_p}\right\|\left\|\overrightarrow{y_q}\right\|}\right) \tag{3}$$

where n is the sample size of the corresponding cluster, $\vec{x}$ and $\vec{y}$ are the vector samples in each cluster. As a sensitivity check for cluster distance metric, we also calculated ϕ_k based on the medoid cluster distance. Medoid distance measures the distance between clusters by calculating the cosine distance between 'medoid samples' of each cluster that show the closest average cosine distance with samples within each cluster. Lastly, to examine the predictive power of cortical

distinctiveness metric of each valence type for AWES scores, we applied univariate and multivariate analysis framework described in the same manner with "Behavioral analysis" section.

Approximation of feature importance with perturbation-based XAI Due to the black-box nature of CEBRA, it was challenging to directly evaluate which STFT-processed EEG features were crucial for contrasting valence states in the latent space. To address this issue, we applied "Dynamask" (Crabbé & Van Der Schaar, 2021), a perturbation-based XAI techniques, to infer attribution maps from the trained CEBRA models. Dynamask learns perturbtation weights, ω, for each feature at every time point to generate pseudo-embeddings with maximal MSE compared to the original embeddings with the least perturbation (see **Figure 1e**). Here, i-th input feature at the timepoint t, $x_{i,t}$ is perturbed to $\pi(x_{i,t})$ as weighted sum of its own value and the average value in the time window it belongs to, formulated by the following equation:

$$\pi\big(x_{i,t}\big) = \big(1 - \omega_{i,t}\big) \cdot x_{i,t} + \omega_{i,t} \cdot \frac{1}{2W+1} \sum_{t'=t-W}^{t+W} x_{i,t'} \tag{4}$$

where W is a time window size. $\omega_{i,t} = 1$ indicates significant alteration of the CEBRA embeddings upon replacement, implying importance in contrasting valence labels. With a receptive field of 10 samples, $W = 5$ was set to align the working behavior of Dynamask with CEBRA's. We obtained ω matrices for each EEG feature from participant-clip paired data showing significant silhouette scores for ambivalence clusters ($P_{\text{perm}} < .05$), using the following parameters: keep_ratio = 0.1, n_epoch = 2,500, initial_mask_coef = 0.5, size_reg_factor_init = 0.5, size_reg_factor_dilation = 100, time_reg_factor = 0, learning_rate = 0.03, and momentum = 0.9. We confirmed that Dynamask reached plateau in training performance to learn attribution map (see **Supplementary**

Figure 3). By aligning STFT EEG time-series and valence keypress sequence, we calculated $\omega_{\text{ambivalent}}$ of each feature by averaging ω values of each feature for time points labeled as ambivalent states in the valence keypress. Consequently, $\omega_{\text{ambivalent}}$ value is indirect measure of feature importance to contrast ambivalence-labeled EEG samples from other valence-labeled EEG samples for the construction of latent valence-cortical space. The same analyses were conducted for positive and negative valence states with data displaying significant silhouette scores for each state, respectively.

Post-hoc analysis of feature importance with hidden Markov model We conducted a post-hoc analysis to confirm Dynamask-driven attribution weights of each feature in distinguishing different valence states within the latent valence-cortical space using hidden Markov model (HMM). Particularly, Dynamask revealed that the power of the delta band consistently held greater importance than the power of other frequency bands (see 'Results'). Based on this, we hypothesized that combinations of delta features would be temporally aligned with individual valence dynamics and tested this hypothesis using an HMM. For all HMM analysis, we used "brainiak" package (Kumar et al., 2021).

We divided the 95 STFT features into frequency bands to generate five input groups – delta, theta, alpha, beta, and gamma – for each participant and clip. These inputs were independently fitted to an HMM, estimating the time points of neural boundaries corresponding to the number of valence transitions reported via keypress. Thus, the event number, a hyperparameter of the HMM, was informed by the participants' reported valence keypress for the clip. Boundaries within ± 3 seconds (i.e., six STFT samples) of the actual valence transition time

points were considered a 'match', and the 'match rate' was calculated as the ratio of matched boundaries to the total number of boundaries.

Following the framework of Vaccaro et al. (2020), we assessed the statistical significance of the match rates for the five frequency bands across all participants and clips. This framework's advantage is that it accounts for variability in the number of valence transitions reported by each participant for each clip. The approach involves the following steps: First, for each participant and frequency band input, maintain the estimated number of intervals of neural boundaries but shuffle them randomly, comparing these to the actual valence transition time points to compute a pseudo-match rate. This process is repeated 1,000 times to generate a null distribution of match rates. Second, calculate the difference between actual match rate and the mean of the null distribution for each frequency band group for each participant-clip data. Averaging these differences across participants and clips yields the average mean difference for feature group k, denoted as M_k. Third, repeat this process 1,000 times using the null distribution of each participant-clip data to derive a null distribution of 1,000 mean differences between permuted match rates and the mean of the null distribution. Denote the i-th permuted mean difference for feature group k as $Q_k^{(i)}$. Last, calculate the P value, P_{perm}, for M_k using the following equation:

$$P_{\text{perm}} = \frac{1}{1000} \sum_{i=1}^{1000} \mathbf{1}\left(Q_k^{(i)} \geq M_k \right) \tag{5}$$

Results

Awe experience is consistently associated with longer and stronger ambivalent feelings, but not with single-valence feelings

We firstly investigated participants' valence dynamics based on their keypress reports. The valence dynamics displayed large variability across participants (see **Figure 2a**). Despite the individual difference, the valence dynamics was significantly intertwined with clips' visual and acoustic features predicting perceivers' affective response – color hue, brightness, and loudness (Chua et al., 2022; Thao et al., 2019) at the individual level (see **Table 2**). These results support the validity of our continuous valence response paradigm by demonstrating that individual's responses are systematically linked to affect-related sensory inputs, while also underlining the diversity inherent in the temporal patterns of individuals' valence dynamics.

Next, we examined whether our awe VR clips were associated with more salient awe experience and ambivalent responses (see **Figure 2b** and **2c**). We found that participants reported significantly higher AWES scores for all awe clips than the control one with large effect sizes (SP-PA: Cohen's $d = 1.837$, $P_{\mathrm{FDR}} = 1{\times}10^{-14}$; CI-PA: Cohen's $d = 1.493$, $P_{\mathrm{FDR}} = 3{\times}10^{-12}$; MO-PA; Cohen's $d = 1.373$, $P_{\mathrm{FDR}} = 2{\times}10^{-11}$). Three awe clips were also associated to significantly longer and stronger ambivalent responses than the control one (SP: Cohen's $d_{\mathrm{duration}} = .562$, $P_{\mathrm{duration/FDR}} = .001$, Cohen's $d_{\mathrm{intensity}} = .451$, $P_{\mathrm{intensity/FDR}} = .005$; CI: Cohen's $d_{\mathrm{duration}} = .514$, $P_{\mathrm{duration/FDR}} = .002$, Cohen's $d_{\mathrm{intensity}} = .449$, $P_{\mathrm{intensity/FDR}} = .005$; MO: Cohen's $d_{\mathrm{duration}} = .790$, $P_{\mathrm{duration/FDR}} = 2{\times}10^{-5}$, Cohen's $d_{\mathrm{intensity}} = .780$, $P_{\mathrm{intensity/FDR}} = 2{\times}10^{-5}$). In contrast, except for the duration of

Table 2. Explanatory power of perceptual features on valence keypress in multinomial mixed logistic regression

	positive vs. neutral			ambivalent vs. neutral			negative vs. neutral		
	β	SE	P	β	SE	P	β	SE	P
Time	.158	.063	.013***	.052	.065	.424	.063	.079	.428
Brightness	3.758	.122	$< 2\times10^{-16***}$	1.361	.115	$< 2\times10^{-16***}$	-1.191	.136	$< 2\times10^{-16***}$
Hue	1.581	.119	$< 2\times10^{-16***}$	.256	.091	.005**	-.543	.111	$1\times10^{-8***}$
Loudness	1.594	.093	$< 2\times10^{-16***}$	.559	.093	$< 2\times10^{-9***}$	1.637	.107	$< 2\times10^{-16***}$

Note. tested model: $\text{valence}_t = t + \text{brightness}_t + \text{hue}_t + \text{loudness}_t + (1 \mid \text{sub}) + (1 \mid \text{clip})$; $^*P < .05$; $^{**}P < .01$; $^{***}P < .001$.

Table 3. Statistical differences in valence and arousal ratings across clips

	SP - PA				CI - PA				MO - PA			
	t	df	d	P_{FDR}	t	df	d	P_{FDR}	t	df	d	P_{FDR}
Valence												
ambivalent(int)	2.957	42	.552	.005**	2.946	42	.576	.005**	5.118	42	.961	$2\times10^{-5***}$
ambivalent(dur)	3.685	42	.775	.001***	3.372	42	.770	.002**	5.179	42	1.218	$2\times10^{-5***}$
positive(int)	1.427	42	.266	.352	-.443	42	.092	.660	1.206	42	.225	.352
positive(dur)	-3.316	42	.667	.003**	-3.618	42	.795	.002**	-3.178	42	.634	.003**
negative(int)	-.295	42	.059	.769	2.562	42	.431	.021*	3.024	42	.536	.013*
negative(dur)	.421	42	.082	.676	4.329	42	.839	$1\times10^{-4***}$	5.243	42	1.085	$1\times10^{-5***}$
Arousal	5.810	42	1.151	$7\times10^{-7***}$	7.404	42	1.433	$6\times10^{-9***}$	9.301	42	1.777	$3\times10^{-11***}$

Note. $^*P_{\text{FDR}} < .05$; $^{**}P_{\text{FDR}} < .01$; $^{***}P_{\text{FDR}} < .001$.

positive feelings, behavioral metrics of positive and negative feelings did not show significant differences between conditions consistently. Awe clips were linked to significantly higher arousal compared to the control one (see **Table 3**).

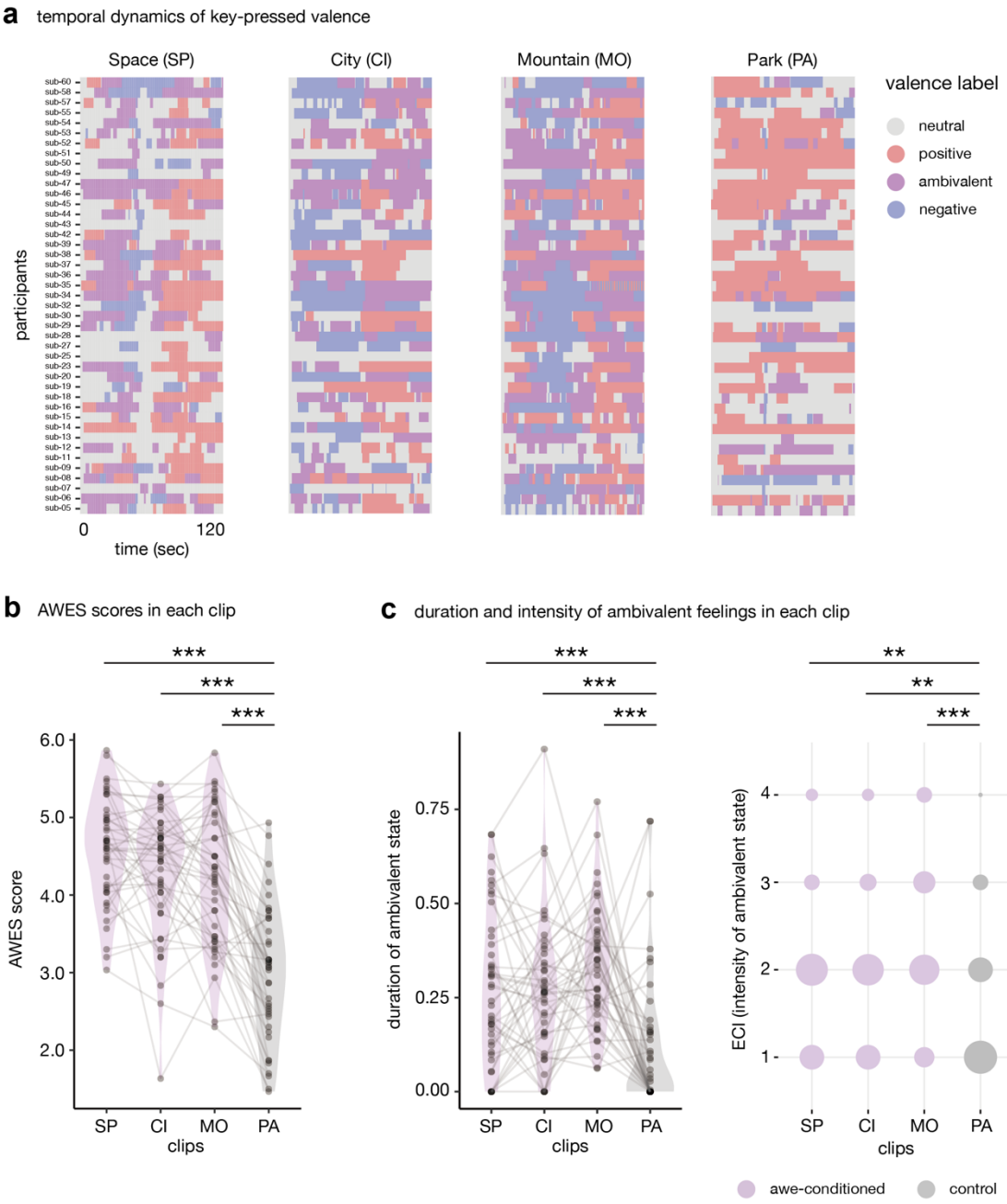


Figure 2. Valence dynamics, awe intensity, and ambivalent feelings across clips a, participants' valence dynamics reported by keypress for each clip. **b,** AWES scores reported for each clip after every trial. **c,** duration (left) and intensity (right) of ambivalent feelings for each clip. This figure is based on data of $N = 43$ included in the behavioral analysis and four clips. *$P_{\text{FDR}} < .05$; **$P_{\text{FDR}} < .01$; ***$P_{\text{FDR}} < .001$.

Ambivalent feelings predict the awe intensity more precisely than other single valence metrics

We tested whether metrics of ambivalence (i.e., its intensity and duration) had more predictive power for awe intensity than other single valence-related features. In the univariate analysis based on linear mixed model, only duration and intensity of ambivalent feelings and intensity of positive ones showed significant fixed effects (duration of ambivalent states: $\beta = .565$, 95% CI = [.033, 1.097], $P = .039$; intensity of ambivalent states: $\beta = .220$, 95% CI = [.090, .351], $P = .001$; intensity of positive states: $\beta = .094$, 95% CI = [.007, .180], $P = .035$; see **Figure 3a**).

In the multivariate analysis, our machine learning-based model exhibited better predictive performance than the linear regression model (see **Figure 3b**). In this model, duration and intensity of ambivalent feelings showed higher feature importance than other single valence metrics (see **Figure 3c**). Computing shapley values, we found that duration and intensity of ambivalent feelings were positively associated with the awe intensity ratings (see **Figure 3d**).

Results of both univariate and multivariate analyses imply that the awe intensity rating is more precisely predicted by ambivalence-related behaviors compared to other valence feelings.

Aligned latent cortical spaces share valence representation architecture across individuals and stimuli

We constructed a latent valence-cortical space for each participant for every clip trial. To evaluate the generalizability of the personalized latent space and determine the optimal dimensionality, predictive analysis of 2 tasks ('across participants', 'across clips') × 3 conditions

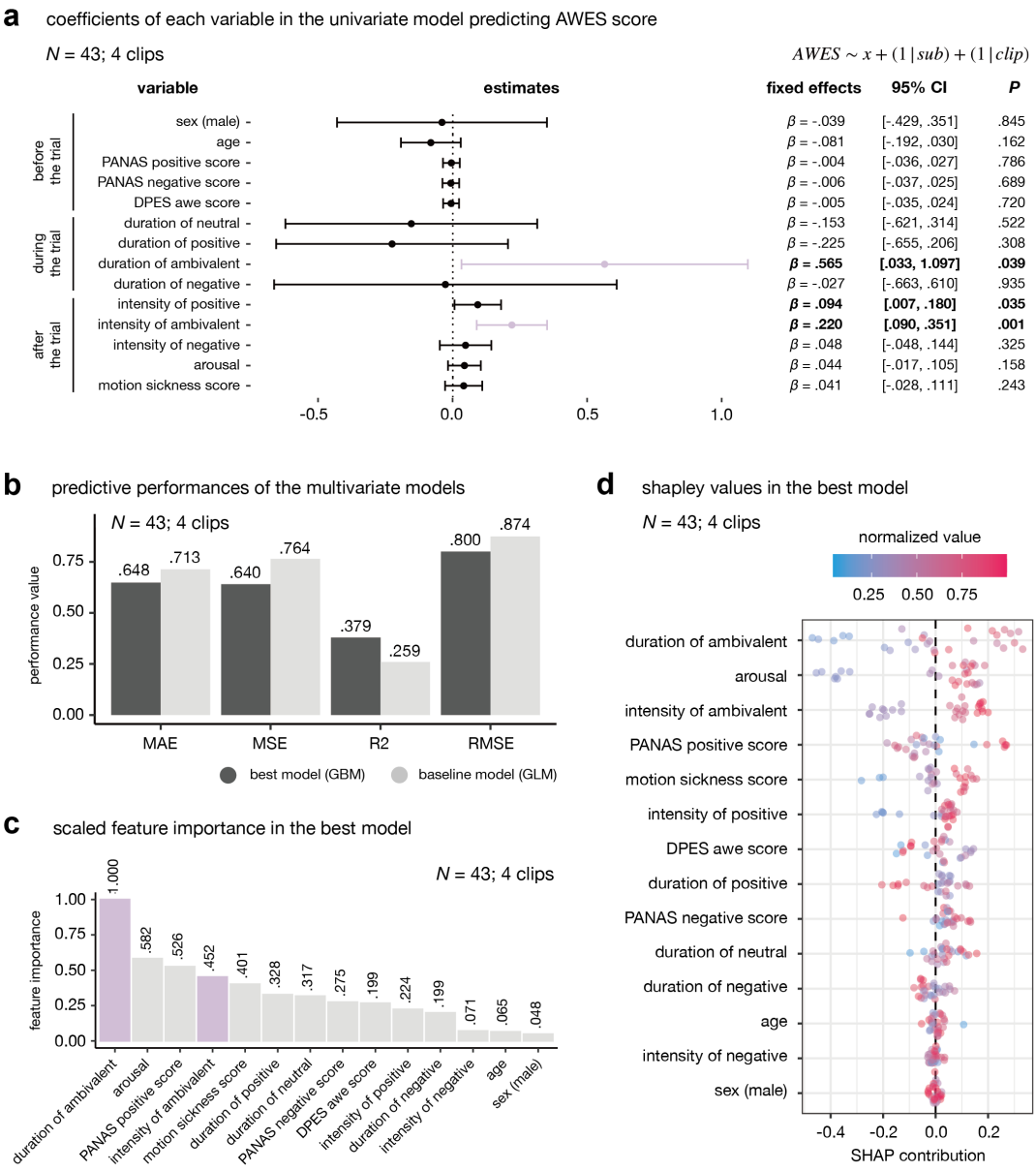


Figure 3. Association between affective components and awe intensity a, beta coefficients of each behavioral metric in the univariate linear mixed effect models. Error bars denote 95% confidence intervals of fixed effects. Purple bars show the estimates of ambivalence-related features. Bolded statistics indicate statistically significant results at $P < .05$. **b,** predictive performance of multivariate machine learning-based models. **c,** scaled feature importance of all behavioral features calculated from the best model. Purple bars show the importance of ambivalence-related features. **d,** shapley value of all metrics computed from the best model. This figure is based on data of $N = 43$ included in the behavioral analysis and four VR clips.

('random', 'not aligned', 'aligned') design was conducted. First, we identified the canonical dimension that demonstrated high predictive performance in the 'aligned' condition with the lowest dimensionality across tasks. Consequently, a 7D space was selected for the CEBRA-based latent embeddings (see **Figure 4a**), while a 6D space was chosen for the PCA-based ones (see **Supplementary Figure 2**).

Next, we compared the generalizability of neural embeddings across conditions and across analytic approaches – CEBRA, PCA, and FAA (see **Figure 4b**). In the 'across-participants' task, a kNN classifier trained with each participant's latent neural embeddings and valence dynamics predicted other participants' valence dynamics above the random chance (aligned – random: Cohen's d = 1.122, P_{FDR} = 5×10^{-17}; not aligned – random: Cohen's d = .249, P_{FDR} = .020). Additionally, aligned embeddings displayed higher predictive performances compared to not-aligned embeddings (aligned – not aligned: Cohen's d = .548, P_{FDR} = 2×10^{-6}). Our aligned CEBRA embeddings achieved significantly more precise prediction compared to aligned PCA-based (Cohen's d = 1.093, P_{FDR} = 2×10^{-16}) and FAA-based embeddings (Cohen's d = 1.115, P_{FDR} = 6×10^{-17}), but not in the random and not-aligned conditions (see **Table 4**).

In the 'across-clips' task, only aligned CEBRA embeddings showed significant predictive power for valence dynamics in the other clip (aligned – random: Cohen's d = .501, P_{FDR} = .015; not aligned – random: Cohen's d = .103, P_{FDR} = .540). Aligned embeddings exhibited higher performance than not-aligned embeddings, but its significance did not reach a threshold (aligned – not aligned: Cohen's d = .283, P_{FDR} = .147). Our aligned CEBRA embeddings predicted valence dynamics in other trials more precisely compared to aligned PCA-based (Cohen's d = .553, P_{FDR}

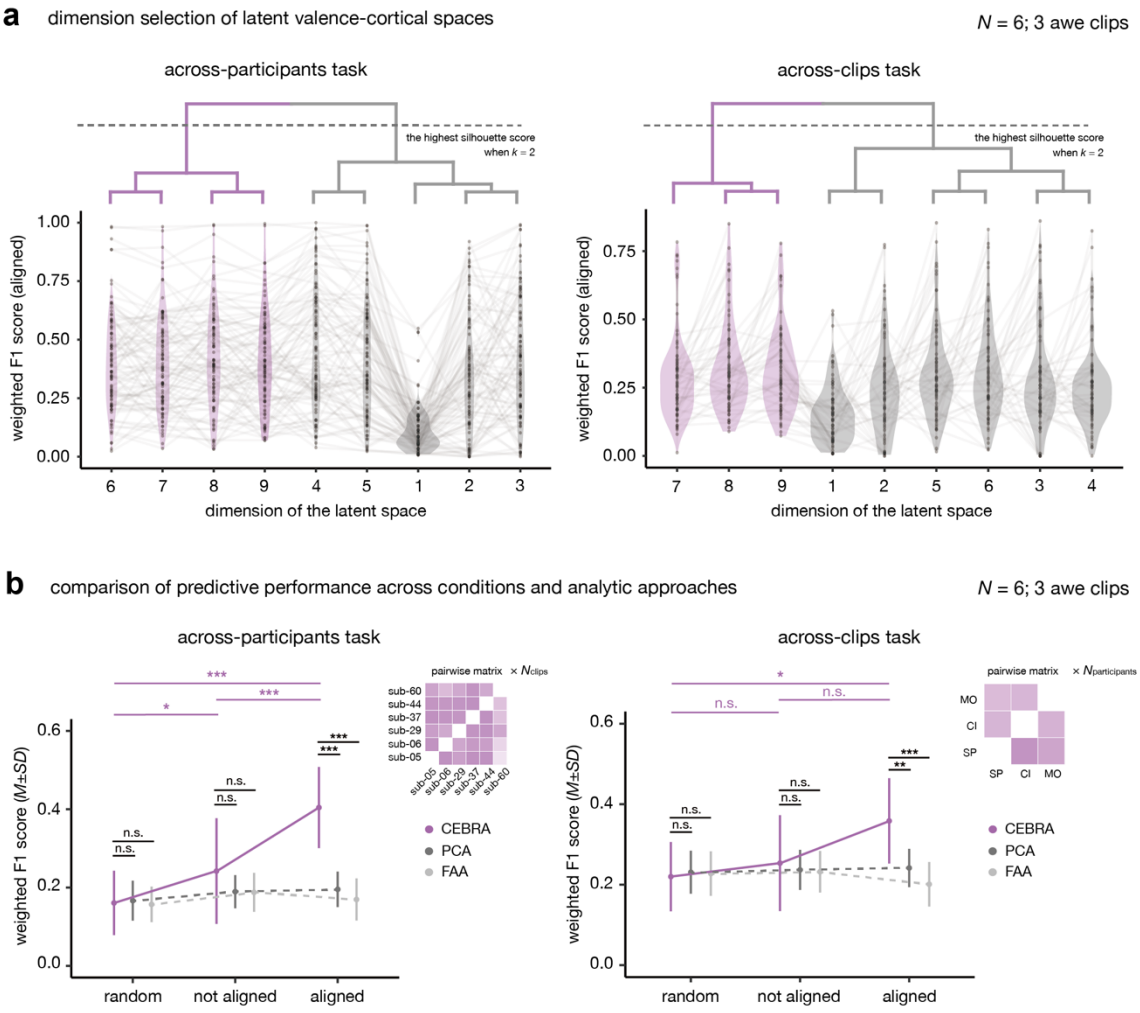


Figure 4. Generalizability of individualized latent valence-cortical spaces across participants and stimuli in predictive tasks a, selected dimensionality of CEBRA-based latent valence-cortical spaces. Results from across-participants (left) and across-clips (right) tasks. Purple violin plots denote the high-performing clusters identified through hierarchical clustering analysis. **b,** predictive performance of CEBRA-, PCA-, and FAA-driven embeddings in three conditions. Purple asterisks indicate the statistical differences in test performances within CEBRA-based prediction. Black asterisks denote the statistical differences in test performances in CEBRA versus PCA and CEBRA versus FAA prediction tasks. Purple heatmaps visualize conceptual scheme of results from pairwise predictive tasks. This figure is based on data of $N = 6$ reporting all valence types in three awe clips and three awe-inducing clips. *$P_{\mathrm{FDR}} < .05$; **$P_{\mathrm{FDR}} < .01$; ***$P_{\mathrm{FDR}} < .001$.

Table 4. Statistical differences in predictive performances of latent cortical embeddings across conditions and analytic approaches

	across participants				across clips			
	t	*df*	*d*	P_{FDR}	*t*	*df*	*d*	P_{FDR}
within CEBRA embeddings								
aligned – random	10.641	89	1.122	$5\times10^{-17***}$	3.005	35	.501	.015*
aligned – not aligned	5.200	89	.548	$2\times10^{-6***}$	1.699	35	.283	.147
not aligned – random	2.366	89	.249	.020*	.620	35	.103	.540
between embeddings								
aligned								
CEBRA - PCA	10.371	89	1.093	$2\times10^{-16***}$	3.320	35	.553	.006**
CEBRA - FAA	10.576	89	1.115	$6\times10^{-17***}$	4.840	35	.807	$8\times10^{-5***}$
not aligned								
CEBRA - PCA	1.830	89	.193	.106	.360	35	.060	.721
CEBRA - FAA	1.898	89	.200	.091	.452	35	.075	.782
random								
CEBRA - PCA	-.379	89	.040	.706	-.410	35	.068	.721
CEBRA - FAA	.269	89	.028	.789	-.278	35	.046	.782

Note. *$P_{FDR} < .05$; **$P_{FDR} < .01$; ***$P_{FDR} < .001$.

= .006) and FAA-based embeddings (Cohen's d = .807, $P_{FDR} = 8\times10^{-5}$), but not in the random and not aligned conditions (see **Table 4**).

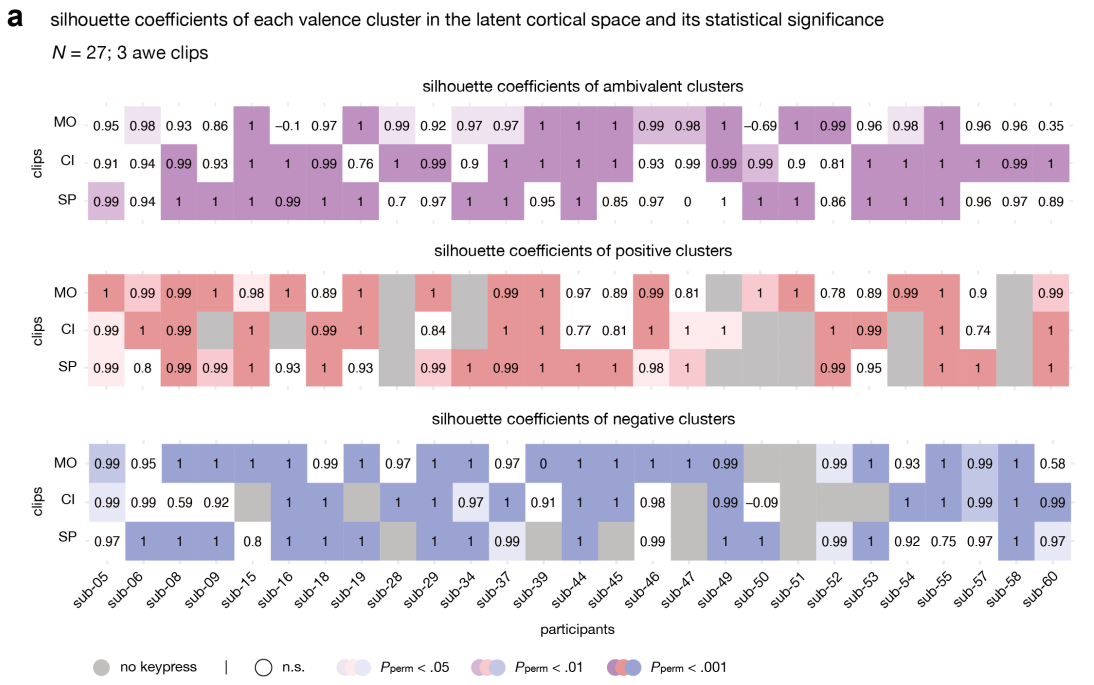
These results imply that our aligned latent spaces across participants and clips share general representation of different valence states in the cortex along with idiosyncratic structures, which could not be captured by conventional linear and unsupervised approaches.

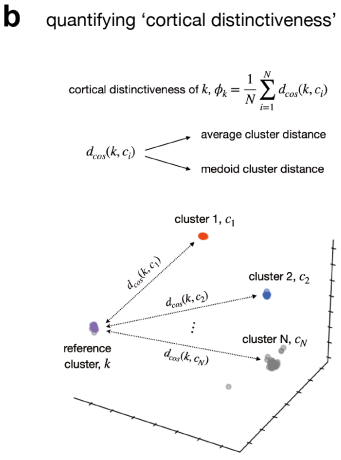
The more distinctively ambivalent feelings are represented in the cortices, the more saliently individuals experience awe

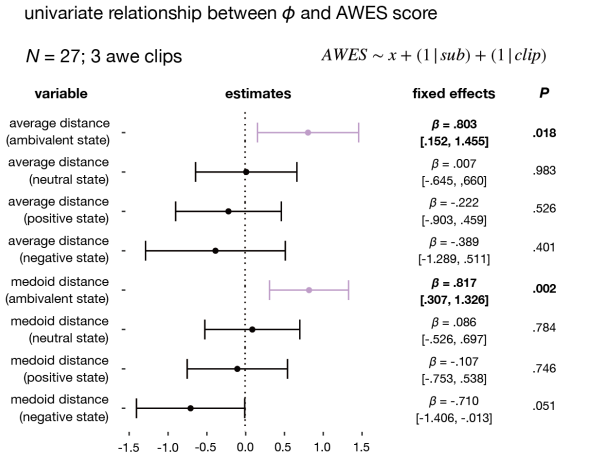
With the latent valence-cortical spaces, we investigated whether ambivalent states exhibited distinct representation distinguishable from neural patterns of the other single valence

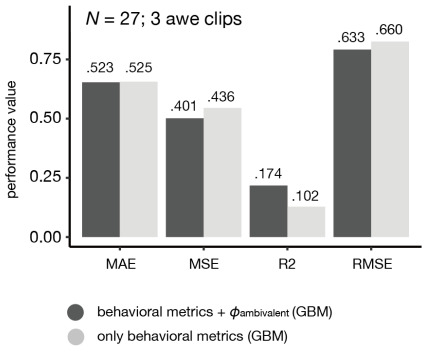
states. The silhouette coefficients of ambivalent clusters in the latent space showed large variability across participants (see **Figure 5a**). For instance, five participants' ambivalent states displayed significant silhouette coefficients in all awe clips while ten individuals showed it only one clip. Contrarily, most participants displayed significant silhouette scores of single valence states more consistently than ambivalent states (see **Figure 5a**). Given the concern about possible confounding effects of cluster size on the statistical significance of silhouette scores, we further tested the association between the cluster size of ambivalent cluster (i.e., the duration of ambivalent feelings) and the P_{perm} values of silhouette coefficients for every clip. No significant correlation between them was detected in any of the videos (SP: R = .221, 95% CI = [-.174, .554], P = .268; CI: R = .198, 95% CI = [-.197, .538], P = .323; MO: R = .126, 95% CI = [-.267, .483], P = .531), indicating that the individual differences in distinct cortical cluster of ambivalent feelings were not driven by its cluster size.

Next, we examined our hypothesis that the more distinctively ambivalent states are represented in the cortical regions, the more saliently individuals experience awe. For this end, we developed new metric called 'cortical distinctiveness' of each valence cluster, ϕ (see **Figure 5b**). In the linear mixed model including four cortical distinctiveness metrics - ϕ_{neutral}, ϕ_{positive}, $\phi_{\text{ambivalent}}$, and ϕ_{negative} and two random intercepts of participant and clip as regressors and awe intensity ratings as outcome variable, only $\phi_{\text{ambivalent}}$ significantly predicted the awe intensity score (β = .817, 95% CI = [.307, 1.326], P = .003; see **Figure 5c**). The predictive power of $\phi_{\text{ambivalent}}$ was not killed when the other cluster distance metric – medoid cluster distance was used to compute $\phi_{\text{ambivalent}}$ values (β = .803, 95% CI = [.152, 1.455], P = .018; see **Figure 5c**).


a
silhouette coefficients of each valence cluster in the latent cortical space and its statistical significance
N = 27; 3 awe clips
silhouette coefficients of ambivalent clusters
silhouette coefficients of positive clusters
silhouette coefficients of negative clusters
clips
MO
CI
SP
participants
sub-05 sub-06 sub-08 sub-09 sub-15 sub-16 sub-18 sub-19 sub-28 sub-29 sub-34 sub-37 sub-39 sub-44 sub-45 sub-46 sub-47 sub-49 sub-50 sub-51 sub-52 sub-53 sub-54 sub-55 sub-57 sub-58 sub-60
no keypress
n.s.
Pperm < .05
Pperm < .01
Pperm < .001


b
quantifying 'cortical distinctiveness'
cortical distinctiveness of k, $\phi_k = \frac{1}{N}\sum_{i=1}^{N} d_{cos}(k, c_i)$
$d_{cos}(k, c_i)$
average cluster distance
medoid cluster distance
cluster 1, c_1
cluster 2, c_2
cluster N, c_N
reference cluster, k


c
univariate relationship between ϕ and AWES score
N = 27; 3 awe clips
$AWES \sim x + (1 \mid sub) + (1 \mid clip)$
variable
estimates
fixed effects
P
average distance (ambivalent state)
β = .803 [.152, 1.455]
.018
average distance (neutral state)
β = .007 [-.645, .660]
.983
average distance (positive state)
β = -.222 [-.903, .459]
.526
average distance (negative state)
β = -.389 [-1.289, .511]
.401
medoid distance (ambivalent state)
β = .817 [.307, 1.326]
.002
medoid distance (neutral state)
β = .086 [-.526, .697]
.784
medoid distance (positive state)
β = -.107 [-.753, .538]
.746
medoid distance (negative state)
β = -.710 [-1.406, -.013]
.051
-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5


d
$\phi_{ambivalent}$-driven predictive performance gain
N = 27; 3 awe clips
performance value
0.00 0.25 0.50 0.75
.523 .525
.401 .436
.174 .102
.633 .660
MAE MSE R2 RMSE
behavioral metrics + $\phi_{ambivalent}$ (GBM)
only behavioral metrics (GBM)

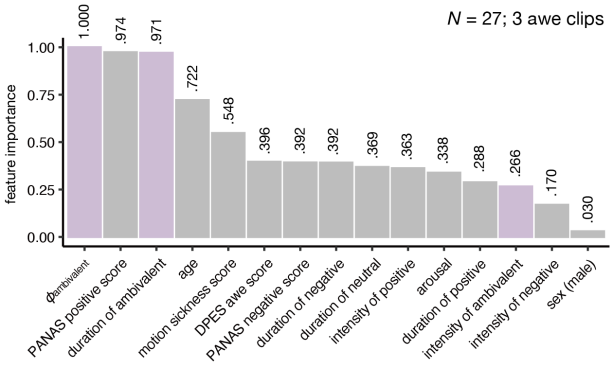

e
scaled feature importance when including $\phi_{ambivalent}$
N = 27; 3 awe clips
feature importance
0.00 0.25 0.50 0.75 1.00
1.000 .974 .971 .722 .548 .396 .392 .392 .369 .363 .338 .288 .266 .170 .030
$\phi_{ambivalent}$
PANAS positive score
duration of ambivalent
age
motion sickness score
DPES awe score
PANAS negative score
duration of negative
duration of neutral
intensity of positive
arousal
duration of positive
intensity of ambivalent
intensity of negative
sex (male)

Figure 5. Individual variability of latent valence representation and its predictive power on awe intensity a, silhouette coefficients of ambivalent (top), positive (middle), and negative (bottom) valence clusters in the latent valence-cortical space. Color of each cell shows statistical significance of the silhouette coefficients assessed through 1,000 times permutation test. **b,** conceptual framework to calculate cortical distinctiveness value of each valence cluster - ϕ. **c,** explanatory power of each valence cluster's ϕ in the univariate linear mixed effect models. Error bar denotes 95% confidence interval of fixed effects. Purple bars show the estimates of ambivalence-related features. Bolded statistics indicate statistically significant results at $P < .05$. **d,** performance gain when $\phi_{\text{ambivalent}}$ was added in the behavioral predictive model. **e,** scaled feature importance of $\phi_{\text{ambivalent}}$ and behavioral features calculated from the best model. Purple bars show the importance of ambivalence-related variables. This figure is based on data of $N = 27$ included in the electrophysiological analysis and three awe-inducing clips.

Contrarily, ϕ values of other valence clusters did not show significant predictive power. As a control analysis, we performed the same univariate analysis with FAA metrics. To extract a single FAA value from each participant-clip data, we calculated FAA values using FFT-driven band power features instead of STFT one, which was marginalized across the whole time-series. We found that FAA did not show any significant predictive power for AWES score ($\beta = .000$, 95% CI = [-.004, .003], $P = .789$).

In addition, when $\phi_{\text{ambivalent}}$ metrics were added to the machine learning model predicting the awe intensity with 14 behavioral variables, its R^2 value was improved about 7.2% (see **Figure 5d**). Furthermore, in this predictive model, $\phi_{\text{ambivalent}}$ displayed higher predictive power than any other behavioral metrics (see **Figure 5e**).

These results suggest individual differences in the distinctiveness of cortical representation related to ambivalent feelings and elucidate that such individual variability can specifically account for the awe experience.

The delta oscillation in the frontal channels mainly engages in distinguishing different valence representation

Lastly, using Dynamask, we investigated which EEG features were importantly used to contrast different valence states to construct the latent valence-cortical spaces. We observed that delta band power features exhibited higher mean perturbation weights for ambivalent states than other band power features (see **Figure 6a**). Within the delta band power features, frontal channels showed larger weights than channels in the other areas (see **Figure 6b**). We performed the same analysis for positive and negative states and found that the delta band power in the frontal channels exhibited consistently higher mean perturbation weights for both states (see **Supplementary Figure 4**).

To confirm the importance of the delta oscillation in valence representation, we performed additional post-hoc analysis using HMM. We hypothesized that the time-series of delta band power in all channels may be temporally aligned with valence dynamics and found that neural boundaries extracted from the combination of delta-related features exhibited significant match rates with participants' valence transition above the random chance (match rate = 53.6%; P_{perm} = .047). We conducted the same analysis with the other four band power features and observed that only beta oscillation displayed significant match rate (see **Figure 6c**).

These results imply that delta oscillation in the frontal channels crucially participate in distinguishing ambivalent feelings to other valence states in the cortices, but their importance is not limited to ambivalent state. They also engage in distinguishing other valence states.

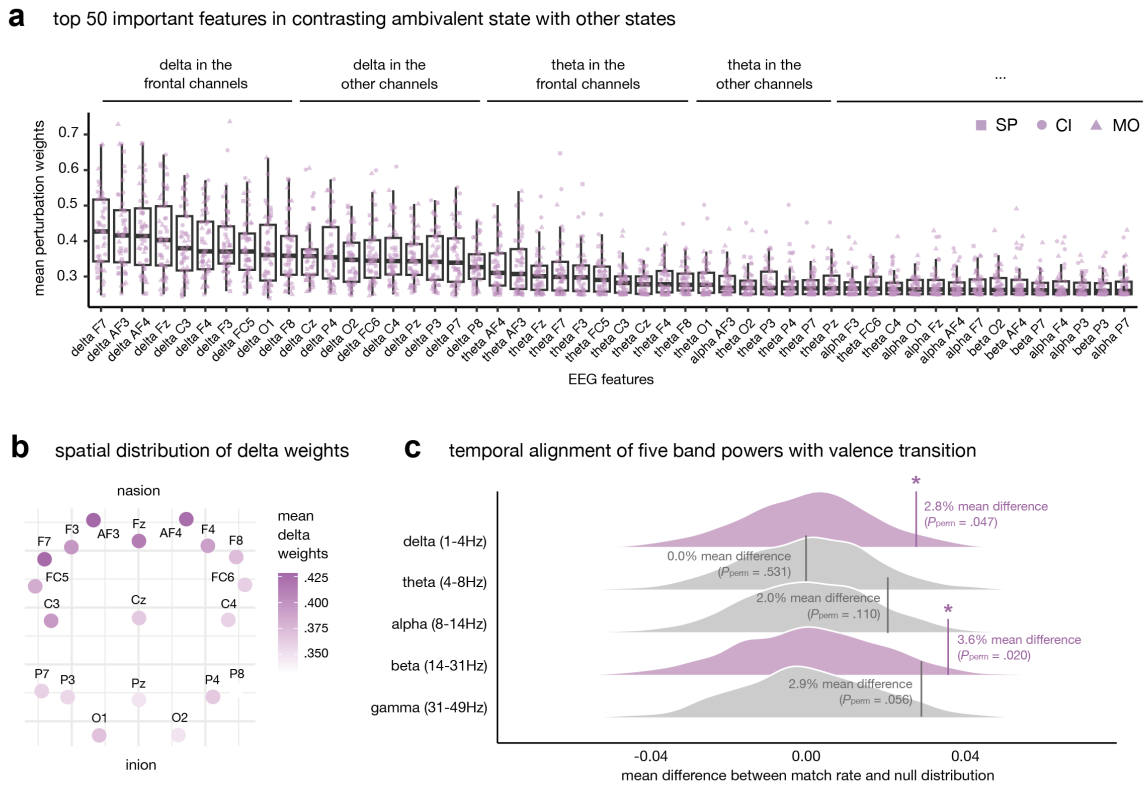


Figure 6. Attribution map of EEG features in constructing latent cortical space a, top 50 important features in contrasting ambivalent states to other valence types. **b,** spatial distribution of delta power-related features' importance. **c,** mean differences between HMM-based match rate and null distribution of each frequency group. The vertical line denotes the real mean differences value. Purple and gray plots show statistically significant and non-significant results, respectively. This figure is based on data of $N = 27$ included in the electrophysiological analysis and three awe-inducing clips. $^{*}P_{\text{perm}} < .05$.

Discussion

In this study, using an integrated VR-EEG protocol, we showed that awe is characterized as an ambivalent feeling rather than a simply positive or negative one at both behavioral and cortical levels. Behaviorally, the intensity of awe rated by participants for each VR clip could be significantly predicted by the duration and intensity of ambivalent feelings experienced while viewing the clip, whereas variables related to positive and negative feelings were less predictive. At the cortical level, we identified a latent valence-cortical space from each individual's EEG signals during watching each movie and aligned these spaces to extract common architecture of valence representation across individuals and stimuli. This revealed that ambivalent feelings have distinct neural representations shared across individuals and clips, though with significant individual variability. This individual difference in distinctiveness of ambivalence-related cortical representation predicted the awe intensity ratings, with more distinct representations of ambivalent feelings correlating with stronger awe. The frontal region's delta oscillation played a key role in distinguishing ambivalent feelings from other valence states and was consistently important in differencing other valence states too.

The high predictive power of the duration and intensity of ambivalent feelings on the level of awe supports our hypothesis that awe is characterized as an ambivalent experience at the behavioral level. This aligns with recent research on the ambivalent nature of awe. Chaudhury et al. (2022) reported that Western population rated threat-awe-inducing images (e.g., photo of Niagara Falls) as having stronger ambivalence compared to stimuli evoking happiness or fear. Here, by utilizing real-time valence ratings, we newly discovered that the duration of ambivalent feelings has a higher explanatory power for the awe ratings than its intensity. We speculate that

its high predictive power may be related to the emotion regulation involved in both ambivalent feelings and awe. Theoretical models of ambivalent feelings suggest they stem from reappraisal (Vaccaro et al., 2020; Van Tilburg et al., 2018). For example, to feel ambivalence toward a stimulus, one must retrieve memories or knowledge related to an opposite valence from the initial valence feelings evoked by the stimulus. In the case of awe, baseline liability of reappraisal for emotion regulation significantly predicts awe ratings for memory recall (Chirico et al., 2024; Chirico et al., 2021), indicating a close relationship between awe and reappraisal. Furthermore, prior search on affect dynamics reports that emotion regulation types show significant correlations with the duration of negative feelings triggered by stimuli (Van Mechelen et al., 2013; Verduyn et al., 2009; Verduyn et al., 2011), but not with their intensity per se (Brans & Verduyn, 2014). Synthesizing our behavioral results with this previous literature, we propose a new 'reappraisal hypothesis' as the cognitive process bridging ambivalent feeling duration and awe ratings, positing that the type of emotion regulation strategy employed by an individual while watching a video explains their awe rating and this relationship is mediated by duration of ambivalent feelings. Future studies tracking the cognitive dynamics of emotion regulation during VR watching are expected to validate this hypothesis.

In our electrophysiological analysis, we identified a latent cortical space that shared valence representations across individuals and sensory input. Specifically, in a pairwise prediction task, the aligned latent valence-cortical embeddings significantly predicted the valence dynamics obtained from other participants and different clips. Meanwhile, embeddings obtained using PCA and FAA did not capture the commonality of valence representations that could be generalized across individuals and clips even after alignment. In the case of FAA, we could confirm that FAA lacks generalizability and specificity as an electrophysiological index of valence as previous studies

have criticized (Gable & Harmon-Jones, 2010; Harmon-Jones & Gable, 2018; Honk & Schutter, 2006; Wacker et al., 2003). Contrarily, the results of PCA-based embeddings were somewhat novel, given that aligned PCA-driven embeddings captured common neural trajectories for various behaviors such as motor control (Gallego et al., 2020; Safaie et al., 2023). We guess that this finding may justify our neural network-based approaches for nonlinear dimensionality to explore shared representation across individuals for affective valence, considering nonlinear relationship between brain activities and valence (Aftanas et al., 1998; Berridge, 2019; Viinikainen et al., 2010). Additionally, the significant predictive performance in the across-clips prediction task indicates that our supervised learning-based dimensionality reduction technique can be effective in disentangling sensory information from valence-cortical embeddings.

The individual-specific latent neural spaces we derived revealed that ambivalent feelings have distinct cortical representations, while also showing significant individual differences in how these neural patterns are differentiated from those of other valence states. We allude that these findings may reconcile conflicting viewpoints about distinct neural system of ambivalent feelings. Our findings support recent studies indicating that ambivalent feelings exhibit unique neural patterns in the cortical regions (Lettieri et al., 2019; Man et al., 2017; Vaccaro et al., 2020; Vaccaro et al., 2024), challenging the constructivist view that ambivalent feelings would not have distinct neural representations since they are merely fluctuations between opposing valence states (Barrett & Bliss-Moreau, 2009; Russell, 2017). However, the individual differences in the cortical distinctiveness of ambivalent feelings can be also interpreted from a constructivist perspective. For instance, neuropsychological factors constructing affect into emotions, such as beliefs about emotions, emotional granularity, and the time window of event segmentation, could contribute to

this individual variability in experiencing mixed emotion and its neural representation (Hoemann et al., 2017).

The cortical distinctiveness of ambivalent feelings during awe experiences shows significant individual difference, but this variability can predict the awe intensity. Specifically, the more distinct the cortical representation of ambivalent feelings compared to positive, negative, and neutral feelings, the higher the reported intensity of awe. Given that the cortical distinctiveness of other feeling categories did not significantly predict awe ratings, this implies a specific relationship between the neural representation of ambivalent feelings and awe intensity. These results highlight a new aspect of awe that conventional approaches, which focus on activation levels of specific regions or networks - e.g., (Guan et al., 2019; Hu et al., 2017; Takano & Nomura, 2022), are insufficient to address by considering the geometrical characteristics of affect-related latent neural spaces.

We suggest that the predictive power of the cortical distinctiveness of ambivalent feelings for awe can be understood through the lens of 'holistic meaning-making', a key cognitive aspect observed in awe experiences (Bonner & Friedman, 2011; Dai et al., 2022; Ihm et al., 2019; Sawada et al., 2024; Yin et al., 2024). During awe, individuals face an extraordinary object that expands their belief or cognitive scheme, leading to the generation of new meanings. In terms of affect, this process integrates the initial negative feelings evoked by the object with the pleasure derived from epistemic transformation. For example, awe is often described as a 'self-transcendent' experience, where conflicting two feelings – 'self-diminishment' and 'connectedness' are harmonized (Yaden et al., 2019). Thus, awe is fundamentally based on the integration of these opposing feelings, resulting in an emotion that cannot be reduced to merely positive or negative information. We propose that the new meaning generated during an awe experience is perceived as an ambivalent

feeling, encompassing both positive and negative aspects. This integration process is reflected in cortical representations distinct from those of simply positive or negative feelings. While our findings do not elucidate the specific attributes of awe's high-level cognitive dynamics, they suggest the potential connection between these cognitive processes and cortical representation patterns through the spatial analysis of latent neural spaces.

Finally, we consistently observed that the delta oscillation of the frontal channels is a significant feature for distinguishing valence states in both Dynamask and HMM analyses. Firstly, the result that frontal channels have higher importance compared to other regions within the same frequency range aligns with several studies on human neuroimaging. The prefrontal cortex (PFC), particularly the orbitofrontal cortex, shows unique activation patterns for conflicting affective information (Levens & Phelps, 2010; Rolls & Grabenhorst, 2008; Simmons et al., 2006) and consistent activity patterns when individuals feel ambivalence during naturalistic movie watching (Vaccaro et al., 2024). We believe this result is related to the valence-related information stored in the PFC, which integrates conflicting bodily signals from interoceptive circuits to create a global mixed feeling.

In terms of the frequency band below 5-6 Hz, encompassing delta and low-theta ranges, has been reported to be closely associated with emotional processing and regulation. For emotional processing, it has been observed that this frequency range in the frontal areas increases when emotional memories, crucial for determining valence feelings to the stimuli, are encoded and retrieved (Brenner et al., 2014; Hutchison & Rathore, 2015; Nishida et al., 2009; Sopp et al., 2017). Unlike other frequency bands, the microstate features of the 1-3 Hz range successfully decode valence ratings to images (Shen et al., 2020). Additionally, intermittent theta burst stimulation, enhancing the 5 Hz band power in the left dorsolateral PFC, improved emotion recognition for

lexical and facial stimuli (Dumitru et al., 2020; Moulier et al., 2021), supporting the role of low-frequency bands in the frontal regions in emotional feeling and processing. In terms of emotion regulation, successful regulation of negative feelings induced by image stimuli has been correlated to increased 4 Hz band power in frontal channels, as reported in both experimental (Ertl et al., 2013) and meta-analytical studies (Cavanagh & Shackman, 2015). Heroin-addicted patients with disrupted regulatory abilities exhibited consistent decreases in the < 5 Hz frequency band in these regions when viewing affect-charged images (Jiang et al., 2022). Interestingly, the synchrony of delta and beta features, which displayed significant alignment with keypressed valence dynamics in the HMM analysis, have been also reported to correlate positively with the efficiency of emotion and stress regulation (Brooker et al., 2021; Myruski et al., 2022; Phelps et al., 2016; Putman et al., 2012). In this context, our results and prior studies may support the rationale of the 'reappraisal hypothesis' we suggested.

Our findings require consideration of three major limitations. First, our emphasis on ambivalent feelings in awe may be overestimated due to the inclusion of only Asian participants in this study. Western individuals report fewer ambivalent feelings in awe experiences compared to Asians (Nakayama et al., 2020). Nonetheless, even among Western populations, ambivalent feelings were rated higher for awe-inducing image stimuli than those inducing single-valence emotions such as happiness or fear (Chaudhury et al., 2022). Thus, it remains to be verified whether the cortical distinctiveness of ambivalent feelings can significantly predict awe intensity in Western group. Second, real-time valence ratings through keypresses might have unintentionally influenced the emotion generation process. Continuous introspection and reporting of feelings can interfere with natural emotional generation (Larsen & Fredrickson, 1999). Despite this, we validated participants' valence ratings by predicting this sequence from each video's

perceptual information that could predict viewers' emotional responses at both individual and stimulus levels. Nevertheless, using collaborative filtering (Jolly et al., 2022) for interpolation of dense reports could mitigate this potential limitation. Third, we assumed that the dimensions of the latent valence-cortical space are identical for all participants. It remains necessary to verify whether individuals' idiosyncratic valence representations are encoded in a space of the same dimensionality.

Despite these considerations, our study is significant in that it elucidates the importance of the ambivalent nature of awe at both behavioral and cortical levels, a topic that has not been sufficiently highlighted in quantitative research. Our approach, which explains awe through the hierarchical integration of negative and positive feelings, offers a new perspective on the origins of the psychiatric benefits of awe (e.g., stress resilience or non-egocentric schemes). Additionally, our approach provides insights not only into awe itself but also significant implications for the distinct neural representation of ambivalent feelings, a topic of debate in affective neuroscience. We anticipate that our study will stimulate further research on topics not directly addressed in this study, such as the relationship between the experience of ambivalent feelings in awe and improvement in mental health, and the cognitive dynamics of emotion regulation required to shape ambivalent feelings during awe experience.

References

Aftanas, L. I., Lotova, N. V., Koshkarov, V. I., Makhnev, V. P., Mordvintsev, Y. N., & Popov, S. A. (1998). Non-linear dynamic complexity of the human EEG during evoked emotions. I*nternational journal of psychophysiology, 28*(1), 63-76.

An, S., Ji, L.-J., Marks, M., & Zhang, Z. (2017). Two sides of emotion: Exploring positivity and negativity in six basic emotions across cultures. *Frontiers in psychology, 8*, 253368.

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Advances in experimental social psychology, 41*, 167-218.

Berkman, E. T., & Lieberman, M. D. (2010). Approaching the bad and avoiding the good: Lateral prefrontal cortical asymmetry distinguishes between action and valence. *Journal of cognitive neuroscience, 22*(9), 1970-1979.

Berridge, K. C. (2019). Affective valence in the brain: modules or modes? *Nature Reviews Neuroscience, 20*(4), 225-234.

Berrios, R., Totterdell, P., & Kellett, S. (2015). Eliciting mixed emotions: a meta-analysis comparing models, types, and measures. *Frontiers in psychology, 6*, 133792.

Bonner, E. T., & Friedman, H. L. (2011). A conceptual clarification of the experience of awe: An interpretative phenomenological analysis. *The humanistic psychologist, 39*(3), 222-235.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry, 25*(1), 49-59.

Brans, K., & Verduyn, P. (2014). Intensity and duration of negative emotions: Comparing the role of appraisals and regulation strategies. *PloS one, 9*(3), e92410.

Brenner, C. A., Rumak, S. P., Burns, A. M., & Kieffaber, P. D. (2014). The role of encoding and attention in facial emotion memory: an EEG investigation. *International journal of psychophysiology, 93*(3), 398-410.

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2012). Emotional valence: A bipolar continuum or two independent dimensions? *Sage Open, 2*(4), 2158244012466558.

Brooker, R. J., Mistry-Patel, S., Kling, J. L., & Howe, H. A. (2021). Deriving within-person estimates of delta-beta coupling: A novel measure for identifying individual differences in emotion and neural function in childhood. *Developmental psychobiology, 63*(6), e22172.

Brzezicka, A., Kamiński, J., Kamińska, O. K., Wołyńczyk-Gmaj, D., & Sedek, G. (2017). Frontal EEG alpha band asymmetry as a predictor of reasoning deficiency in depressed people. *Cognition and emotion, 31*(5), 868-878.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological bulletin, 115*(3), 401.

Cavanagh, J. F., & Shackman, A. J. (2015). Frontal midline theta reflects anxiety and cognitive control: meta-analytic evidence. *Journal of physiology-Paris, 109*(1-3), 3-15.

Čeko, M., Kragel, P. A., Woo, C.-W., López-Solà, M., & Wager, T. D. (2022). Common and stimulus-type-specific brain representations of negative affect. *Nature neuroscience, 25*(6), 760-770.

Chaudhury, S. H., Garg, N., & Jiang, Z. (2022). The curious case of threat-awe: A theoretical and empirical reconceptualization. *Emotion, 22*(7), 1653.

Chirico, A., Borghesi, F., Yaden, D. B., Pizzolante, M., Sarcinella, E. D., Cipresso, P., & Gaggioli, A. (2024). Unveiling the underlying structure of awe in virtual reality and in autobiographical recall: an exploratory study. *Scientific reports, 14*(1), 12474.

Chirico, A., Cipresso, P., Yaden, D. B., Biassoni, F., Riva, G., & Gaggioli, A. (2017). Effectiveness of immersive videos in inducing awe: an experimental study. *Scientific reports, 7*(1), 1218.

Chirico, A., Ferrise, F., Cordella, L., & Gaggioli, A. (2018). Designing awe in virtual reality: An experimental study. *Frontiers in psychology, 8*, 293522.

Chirico, A., Shiota, M. N., & Gaggioli, A. (2021). Positive emotion dispositions and emotion regulation in the Italian population. *PloS one, 16*(3), e0245545.

Chirico, A., Yaden, D. B., Riva, G., & Gaggioli, A. (2016). The potential of virtual reality for the investigation of awe. *Frontiers in psychology, 7,* 223153.

Chua, P., Makris, D., Herremans, D., Roig, G., & Agres, K. (2022). Predicting emotion from music videos: exploring the relative contribution of visual and auditory information to affective responses. *arXiv preprint* arXiv:2202.10453.

Crabbé, J., & Van Der Schaar, M. (2021). Explaining time series predictions with dynamic masks. *International Conference on Machine Learning*.

Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature neuroscience, 17*(11), 1500-1509.

Dai, Y., Jiang, T., & Miao, M. (2022). Uncovering the effects of awe on meaning in life. *Journal of Happiness Studies, 23*(7), 3517-3529.

Delorme, A. (2023). EEG is better left alone. *Scientific reports, 13*(1), 2372.

Delorme, A., & 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.

Diaconis, P., & Freedman, D. (1984). Asymptotics of graphical projection pursuit. *The annals of statistics*, 793-815.

Dumitru, A., Rocchi, L., Saini, F., Rothwell, J. C., Roiser, J. P., David, A. S., Richieri, R. M., Lewis, G., & Lewis, G. (2020). Influence of theta-burst transcranial magnetic stimulation over the dorsolateral prefrontal cortex on emotion processing in healthy volunteers. *Cognitive, Affective, & Behavioral Neuroscience, 20*, 1278-1293.

Ersner-Hershfield, H., Mikels, J. A., Sullivan, S. J., & Carstensen, L. L. (2008). Poignancy: mixed emotional experience in the face of meaningful endings. *Journal of personality and social psychology, 94*(1), 158.

Ertl, M., Hildebrandt, M., Ourina, K., Leicht, G., & Mulert, C. (2013). Emotion regulation by cognitive reappraisal—the role of frontal theta oscillations. *NeuroImage, 81*, 412-421.

Gable, P., & Harmon-Jones, E. (2010). The motivational dimensional model of affect: Implications for breadth of attention, memory, and cognitive categorisation. *Cognition and emotion, 24*(2), 322-337.

Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A., & Miller, L. E. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. *Nature neuroscience, 23*(2), 260-270.

Gordon, A. M., Stellar, J. E., Anderson, C. L., McNeil, G. D., Loew, D., & Keltner, D. (2017). The dark side of the sublime: Distinguishing a threat-based variant of awe. *Journal of personality and social psychology, 113*(2), 310.

Guan, F., Zhao, S., Chen, S., Lu, S., Chen, J., & Xiang, Y. (2019). The neural correlate difference between positive and negative awe. *Frontiers in Human Neuroscience, 13*, 206.

Harmon-Jones, E., & Gable, P. A. (2018). On the role of asymmetric frontal cortical activity in approach and withdrawal motivation: An updated review of the evidence. *Psychophysiology, 55*(1), e12879.

Hartig, F. (2018). DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. R Package version 020.

Hoemann, K., Gendron, M., & Barrett, L. F. (2017). Mixed emotions in the predictive brain. *Current Opinion in Behavioral Sciences, 15*, 51-57.

Honk, J. v., & Schutter, D. J. (2006). From affective valence to motivational direction: the frontal asymmetry of emotion revised. *Psychological science, 17*(11), 963-965.

Hu, X., Yu, J., Song, M., Yu, C., Wang, F., Sun, P., Wang, D., & Zhang, D. (2017). EEG correlates of ten positive emotions. *Frontiers in Human Neuroscience, 11*, 26.

Hutchison, I. C., & Rathore, S. (2015). The role of REM sleep theta activity in emotional memory. *Frontiers in psychology, 6*, 1439.

Ihm, E. D., Paloutzian, R. F., van Elk, M., & Schooler, J. W. (2019). Awe as a meaning-making emotion: On the evolution of awe and the origin of religions. In *The evolution of religion, religiosity and theology* (pp. 138-153). Routledge.

Jiang, H., Ding, X., Zhao, S., Li, Y., Bai, H., Gao, H., & Gao, W. (2022). Abnormal brain oscillations and activation of patients with heroin use disorder during emotion regulation: The role of delta-and theta-band power. *Journal of Affective Disorders, 315*, 121-129.

Jiang, T., Hicks, J. A., Yuan, W., Yin, Y., Needy, L., & Vess, M. (2024). The unique nature and psychosocial implications of awe. *Nature Reviews Psychology*, 1-14.

Jolly, E., Farrens, M., Greenstein, N., Eisenbarth, H., Reddan, M. C., Andrews, E., Wager, T. D., & Chang, L. J. (2022). Recovering individual emotional states from sparse ratings using collaborative filtering. *Affective Science, 3*(4), 799-817.

Kahn, A. S., & Cargile, A. C. (2021). Immersive and interactive awe: Evoking awe via presence in virtual reality and online videos to prompt prosocial behavior. *Human Communication Research, 47*(4), 387-417.

Keltner, D., & Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition and emotion, 17*(2), 297-314.

Kumar, M., Anderson, M. J., Antony, J. W., Baldassano, C., Brooks, P. P., Cai, M. B., Chen, P.-H. C., Ellis, C. T., Henselman-Petrusek, G., & Huberdeau, D. (2021). BrainIAK: the brain imaging analysis kit. *Aperture neuro, 1*(4).

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest'. R package version, 2(0), 734.

Lammel, S., Lim, B. K., Ran, C., Huang, K. W., Betley, M. J., Tye, K. M., Deisseroth, K., & Malenka, R. C. (2012). Input-specific control of reward and aversion in the ventral tegmental area. *Nature, 491*(7423), 212-217.

Larsen, J. T., Norris, C. J., McGraw, A. P., Hawkley, L. C., & Cacioppo, J. T. (2009). The evaluative space grid: a single-item measure of positivity and negativity. *Cognition and emotion, 23*(3), 453-480.

Larsen, R. J., & Fredrickson, B. L. (1999). Measurement issues in emotion research. *Well-being: The foundations of hedonic psychology, 40*, 60.

LeDell, E., & Poirier, S. (2020). H2o automl: Scalable automatic machine learning. *Proceedings of the AutoML Workshop at ICML.*

Lee, S. A., Lee, J.-J., Han, J., Choi, M., Wager, T. D., & Woo, C.-W. (2024). Brain representations of affective valence and intensity in sustained pleasure and pain. *Proceedings of the National Academy of Sciences, 121*(25), e2310433121.

Lettieri, G., Handjaras, G., Cappello, E. M., Setti, F., Bottari, D., Bruno, V., Diano, M., Leo, A., Tinti, C., & Garbarini, F. (2024). Dissecting abstract, modality-specific and experience-dependent coding of affect in the human brain. *Science Advances, 10*(10), eadk6840.

Lettieri, G., Handjaras, G., Ricciardi, E., Leo, A., Papale, P., Betta, M., Pietrini, P., & Cecchetti, L. (2019). Emotionotopy in the human right temporo-parietal cortex. *Nature communications, 10*(1), 5568.

Levens, S. M., & Phelps, E. A. (2010). Insula and orbital frontal cortex activity underlying emotion interference resolution in working memory. *Journal of cognitive neuroscience, 22*(12), 2790-2803.

Lim, Y.-J., Yu, B.-H., Kim, D.-K., & Kim, J.-H. (2010). The positive and negative affect schedule: Psychometric properties of the Korean version. *Psychiatry investigation, 7*(3), 163.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., Studer, M., Roudier, P., & Gonzalez, J. (2013). Package 'cluster'. Dosegljivo na, 980.

Man, V., Nohlen, H. U., Melo, H., & Cunningham, W. A. (2017). Hierarchical brain systems support multiple representations of valence and mixed affect. *Emotion Review, 9*(2), 124-132.

Moeller, J., Ivcevic, Z., Brackett, M. A., & White, A. E. (2018). Mixed emotions: Network analyses of intra-individual co-occurrences within and across situations. *Emotion, 18*(8), 1106.

Moulier, V., Gaudeau-Bosma, C., Thomas, F., Isaac, C., Thomas, M., Durand, F., Schenin-King Andrianisaina, P., Valabregue, R., Laidi, C., & Benadhira, R. (2021). Effect of intermittent theta burst stimulation on the neural processing of emotional stimuli in healthy volunteers. *Journal of Clinical Medicine, 10*(11), 2449.

Myruski, S., Bagrodia, R., & Dennis-Tiwary, T. (2022). Delta-beta correlation predicts adaptive child emotion regulation concurrently and two years later. *Biological Psychology, 167*, 108225.

Nakayama, M., Nozaki, Y., Taylor, P. M., Keltner, D., & Uchida, Y. (2020). Individual and cultural differences in predispositions to feel positive and negative aspects of awe. *Journal of Cross-Cultural Psychology, 51*(10), 771-793.

Nishida, M., Pearsall, J., Buckner, R. L., & Walker, M. P. (2009). REM sleep, prefrontal theta, and the consolidation of human emotional memory. *Cerebral Cortex, 19*(5), 1158-1166.

Norman, G. J., Norris, C. J., Gollan, J., Ito, T. A., Hawkley, L. C., Larsen, J. T., Cacioppo, J. T., & Berntson, G. G. (2011). Current emotion research in psychophysiology: The neurobiology of evaluative bivalence. *Emotion Review, 3*(3), 349-359.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research, 12*, 2825-2830.

Phelps, R. A., Brooker, R. J., & Buss, K. A. (2016). Toddlers' dysregulated fear predicts delta–beta coupling during preschool. *Developmental Cognitive Neuroscience, 17*, 28-34.

Picard, F., & Craig, A. (2009). Ecstatic epileptic seizures: a potential window on the neural basis for human self-awareness. *Epilepsy & Behavior, 16*(3), 539-546.

Piff, P. K., Dietze, P., Feinberg, M., Stancato, D. M., & Keltner, D. (2015). Awe, the small self, and prosocial behavior. *Journal of personality and social psychology, 108*(6), 883.

Putman, P., Arias-Garcia, E., Pantazi, I., & van Schie, C. (2012). Emotional Stroop interference for threatening words is related to reduced EEG delta–beta coupling and low attentional control. *International journal of psychophysiology, 84*(2), 194-200.

Quaedflieg, C. W., Smulders, F. T., Meyer, T., Peeters, F., Merckelbach, H., & Smeets, T. (2016). The validity of individual frontal alpha asymmetry EEG neurofeedback. *Social cognitive and affective neuroscience, 11*(1), 33-43.

Quesnel, D., & Riecke, B. E. (2018). Are you awed yet? How virtual reality gives us awe and goose bumps. *Frontiers in psychology, 9*, 403078.

Reynolds, S. M., & Berridge, K. C. (2008). Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens. *Nature neuroscience, 11*(4), 423-425.

Rolls, E. T., & Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: from affect to decision-making. *Progress in neurobiology, 86*(3), 216-244.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review, 110*(1), 145.

Russell, J. A. (2017). Mixed emotions viewed from the psychological constructionist perspective. *Emotion Review, 9*(2), 111-117.

Safaie, M., Chang, J. C., Park, J., Miller, L. E., Dudman, J. T., Perich, M. G., & Gallego, J. A. (2023). Preserved neural dynamics across animals performing similar behaviour. *Nature, 623*(7988), 765-771.

Sawada, K., Koike, H., Murayama, A., Nishida, H., & Nomura, M. (2024). Appreciation processing evoking feelings of being moved and inspiration: Awe and meaning-making. *Journal of Creativity, 34*(1), 100076.

Schmidt, L. A., & Trainor, L. J. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition & Emotion, 15*(4), 487-500.

Schneider, S., Lee, J. H., & Mathis, M. W. (2023). Learnable latent embeddings for joint behavioural and neural analysis. *Nature, 617*(7960), 360-368.

Shen, X., Hu, X., Liu, S., Song, S., & Zhang, D. (2020). Exploring EEG microstates for affective computing: decoding valence and arousal experiences during video watching. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).*

Shiota, M. N., Keltner, D., & John, O. P. (2006). Positive emotion dispositions differentially associated with Big Five personality and attachment style. *The journal of positive psychology, 1*(2), 61-71.

Shiota, M. N., Keltner, D., & Mossman, A. (2007). The nature of awe: Elicitors, appraisals, and effects on self-concept. *Cognition and emotion, 21*(5), 944-963.

Silvia, P. J., Fayn, K., Nusbaum, E. C., & Beaty, R. E. (2015). Openness to experience and awe in response to nature and music: personality and profound aesthetic experiences. *Psychology of aesthetics, creativity, and the arts, 9*(4), 376.

Simmons, A., Stein, M. B., Matthews, S. C., Feinstein, J. S., & Paulus, M. P. (2006). Affective ambiguity for a group recruits ventromedial prefrontal cortex. *NeuroImage, 29*(2), 655-661.

Sopp, M. R., Michael, T., Weeß, H.-G., & Mecklinger, A. (2017). Remembering specific features of emotional events across time: The role of REM sleep and prefrontal theta oscillations. *Cognitive, Affective, & Behavioral Neuroscience, 17*, 1186-1209.

Takano, R., & Nomura, M. (2022). Neural representations of awe: Distinguishing common and distinct neural mechanisms. *Emotion, 22*(4), 669.

Thao, H. T. P., Herremans, D., & Roig, G. (2019). Multimodal Deep Models for Predicting Affective Responses Evoked by Movies. *ICCV Workshops.*

Vaccaro, A. G., Kaplan, J. T., & Damasio, A. (2020). Bittersweet: the neuroscience of ambivalent affect. *Perspectives on Psychological Science, 15*(5), 1187-1199.

Vaccaro, A. G., Wu, H., Iyer, R., Shakthivel, S., Christie, N. C., Damasio, A., & Kaplan, J. (2024). Neural patterns associated with mixed valence feelings differ in consistency and predictability throughout the brain. *Cerebral Cortex, 34*(4), bhae122.

Van Der Vinne, N., Vollebregt, M. A., Van Putten, M. J., & Arns, M. (2017). Frontal alpha asymmetry as a diagnostic marker in depression: Fact or fiction? A meta-analysis. *Neuroimage: clinical, 16*, 79-87.

Van Mechelen, I., Verduyn, P., & Brans, K. (2013). The duration of emotional episodes. In *Changing emotions* (pp. 174-180). Psychology Press.

Van Tilburg, W. A., Wildschut, T., & Sedikides, C. (2018). Nostalgia's place among self-relevant emotions. *Cognition and emotion, 32*(4), 742-759.

Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Van Mechelen, I. (2009). Predicting the duration of emotional experience: two experience sampling studies. *Emotion, 9*(1), 83.

Verduyn, P., Van Mechelen, I., & Tuerlinckx, F. (2011). The relation between event processing and the duration of emotional experience. *Emotion, 11*(1), 20.

Viinikainen, M., Jääskeläinen, I. P., Alexandrov, Y., Balk, M. H., Autti, T., & Sams, M. (2010). Nonlinear relationship between emotional valence and brain activity: evidence of separate negative and positive valence dimensions. *Human brain mapping, 31*(7), 1030-1040.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods, 17*(3), 261-272.

Wacker, J., Heldmann, M., & Stemmler, G. (2003). Separating emotion and motivational direction in fear and anger: effects on frontal asymmetry. *Emotion, 3*(2), 167.

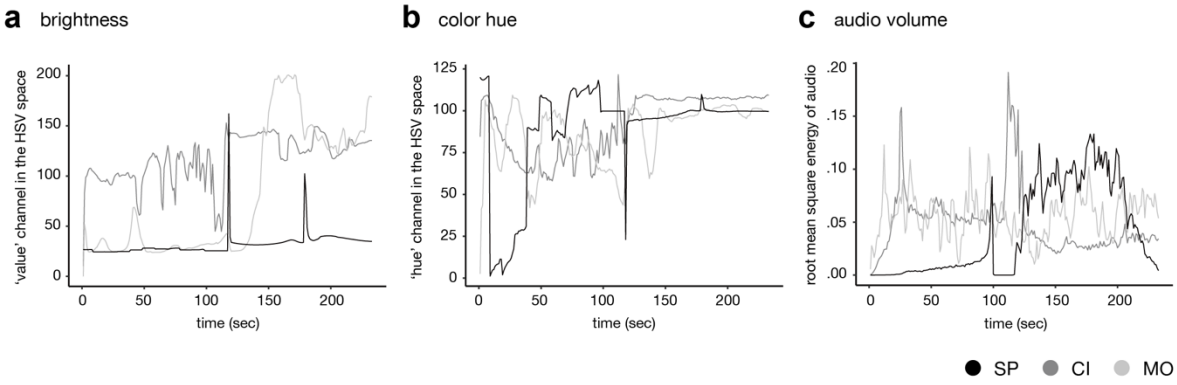
Wittmann, M. (2013). The inner sense of time: how the brain creates a representation of duration. *Nature Reviews Neuroscience, 14*(3), 217-223.

Yaden, D. B., Iwry, J., Slack, K. J., Eichstaedt, J. C., Zhao, Y., Vaillant, G. E., & Newberg, A. B. (2016). The overview effect: awe and self-transcendent experience in space flight. *Psychology of Consciousness: Theory, Research, and Practice, 3*(1), 1.

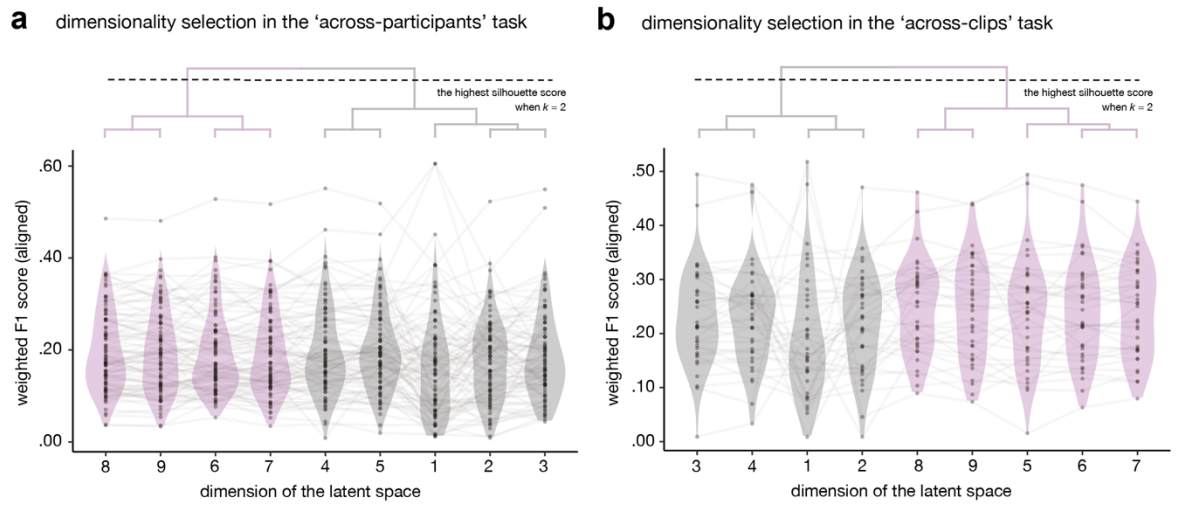
Yaden, D. B., Kaufman, S. B., Hyde, E., Chirico, A., Gaggioli, A., Zhang, J. W., & Keltner, D. (2019). The development of the Awe Experience Scale (AWE-S): A multifactorial measure for a complex emotion. *The journal of positive psychology, 14*(4), 474-488.

Yin, Y., Yuan, W., Hao, C., Du, Y., Xu, Z., Hicks, J. A., & Jiang, T. (2024). Awe fosters positive attitudes toward solitude. *Nature Mental Health*, 1-11.

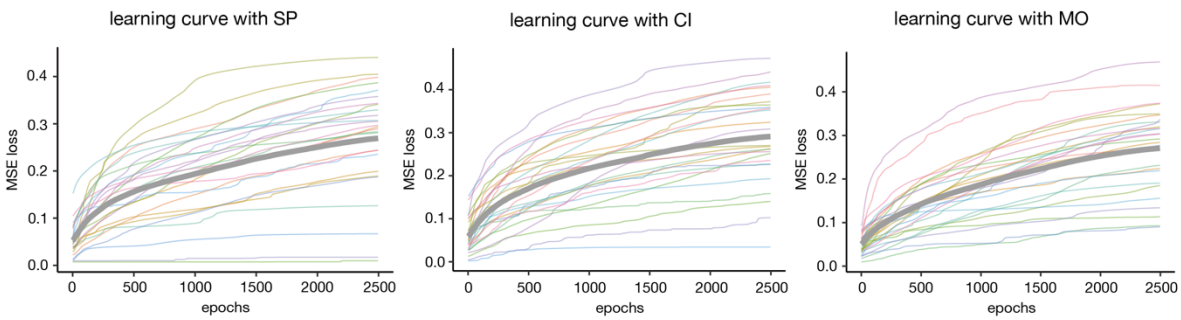
Supplementary Materials



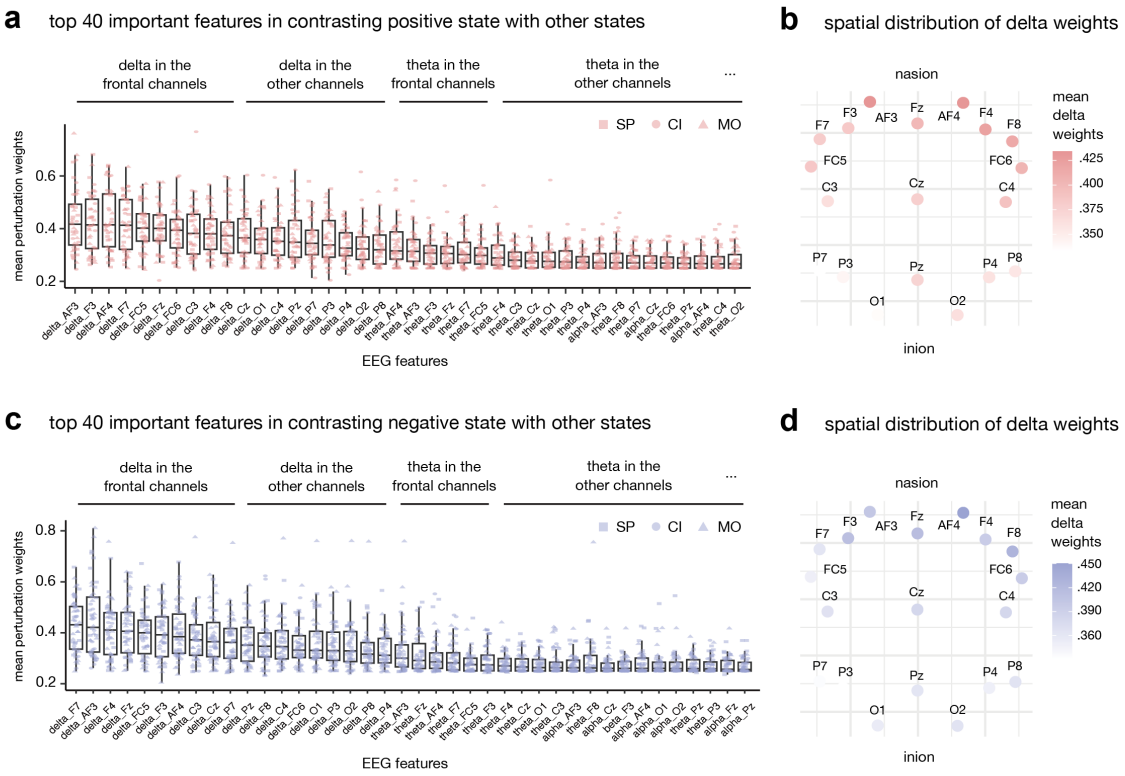
Supplementary Figure 1. Diverse sensory dynamics of three awe-inducing clips **a**, brightness of each clip. Brightness is measured as 'value' channel in the hue-saturation-value (HSV) color space. **b**, color hue of each clip. Hue is calculated as 'hue' channel in the HSV color space. **c**, audio volume of each clip. Volume is defined as root mean square energy of audio segments.



Supplementary Figure 2. Dimensionality selection for PCA-driven embeddings **a**, Test performances in the 'across-participants' task. Purple violin plots denote high-performing group identified by hierarchical clustering analysis. **b**, Test performances in the 'across-clips' task. Based on these two predictive tasks, 6D embeddings were chosen as the optimal dimensionality for PCA-based latent valence-cortical spaces.



Supplementary Figure 3. Learning curves of Dynamask Dynamask learns perturbation weights of EEG features for every time point so that it generates perturbed embeddings showing large mean square error (MSE) loss with the original CEBRA embeddings. Colored curves denote its learning curves based on each participant's data. Thick gray curves visualize the average loss values at every epoch.



Supplementary Figure 4. Dynamask weights of EEG features in positive and negative states **a**, top 40 EEG features with high perturbation weights for positive states. **b,** spatial distribution of delta features' perturbation weights in the positive states. **c**, top 40 EEG features with high perturbation weights for negative states. **d,** spatial distribution of delta features' perturbation weights in the negative states.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1C1C1006503, RS-2023-00266787, RS-2023-00265406, RS-2024-00421268), by Creative-Pioneering Researchers Program through Seoul National University (No. 200-20230058), by Semi-Supervised Learning Research Grant by SAMSUNG (No.A0426-20220118), by Identify the network of brain preparation steps for concentration Research Grant by LooxidLabs (No.339-20230001), by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)] and by the National Supercomputing Center with supercomputing resources including technical support (KSC-2023-CRE-0568) and by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A3A2A02090597).

Abstract in Korean

경외감에 동반하는 양가적 느낌은 경외 체험에 대한 질적 묘사에서 독자적인 성질로 강조되어 왔으며, 그것이 갖는 다양한 정신의학적 및 심리사회적 이점의 가능한 근원으로 언급되어 왔다. 그럼에도 불구하고 정서과학의 이분법적인 긍/부정 도식으로 인해 경외감의 양가적 성질은 온전히 연구되지 못했다. 본 연구에서는 경외에 내포된 유인성의 동역학을 풍부하게 포착하기 위한 가상현실-뇌전도 결합 프로토콜과 양가적 느낌의 응답을 허용하는 확장된 유인성 측정도구를 사용하여, 행동 및 피질 수준에서 경외감이 단순 긍/부정 느낌보다 양가적 느낌으로 더 정확히 특정될 수 있는지 살펴보았다. 행동 수준에서, 참여자들이 각 가상현실 영상에 대해 평정한 경외 수준은 양가적 느낌의 길이와 강도에 의해 정확히 예측될 수 있었던 반면, 다른 유인성에 관한 평정값으로는 유의하게 예측되지 못했다. 피질 수준에서, 경외 체험 동안의 양가적 느낌은 잠재 신경 공간에서 독자적인 신경 표상을 보였지만, 그 신경 표상이 긍/부정 느낌의 표상에 대해 갖는 구분가능성은 큰 개인차를 보였다. 그럼에도 불구하고, 양가적 느낌이 긍/부정 느낌의 표상과 질적으로 구분되어 피질에 부호화 될수록, 더 강한 경외감이 보고되었다. 마지막으로, 다른 종류의 유인성 표상을 구분하는데 있어 주로 전측 영역의 델타파의 파워가 관여하는 것으로 관찰되었다. 이 연구는 정서 신경과학에서 논쟁되어오던 양가적 느낌의 독자적인 신경표상을 탐색했을 뿐 아니라, 이를 기반으로 행동 및 피질 수준에서 경외감이 양가적 경험으로써 더욱 정확히 특정될 수 있음을 밝혔다는 점에서 의의를 갖는다.

주요어: 경외, 양가성, 잠재 신경 공간, 뇌전도, 가상 현실

학번: 2022-23358