UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Institut für systemische Neurowissenschaften

Prof. Dr. med. Christian Büchel

Exploring the Role of Contextual Influences on Cognition: Insights into Emotion Processing, Learning, and Crossmodal Correspondence

Dissertation

zur Erlangung des Doktorgrades PhD
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Carina Daniela Jaap

aus Augsburg

Hamburg, 2025

(wird von der Medizinischen Fakultät ausgefüllt)

Angenommen von der
Medizinischen Fakultät der Universität Hamburg am: 31.03.2025

Veröffentlicht mit Genehmigung der
Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende: Prof. Dr. Michael Rose

Prüfungsausschuss, zweite/r Gutachter/in: PD Dr. Patrick Bruns

Contents

1. Introduction

Context defines and shapes the perception of sensory cues, memories of past events and the meaning of words (Stark et al., 2018; Willems & Peelen, 2021). The classical view of context frames it as a rich set of information surrounding an event, covering a variety of different aspects. These aspects may be internal, such as physiological or emotional states, or external, including spatial and temporal cues (Maren et al., 2013). Context enables us to assess meaning and respond appropriately and flexible to events in real time. Think of a fire alarm: as the meaning of a fire alarm is ambiguous and depends on the situational context, your reaction will also adapt to the alarm depending on the given context. You might interpret the alarm as a threat if it is close to you, as part of a routine drill if one has been announced previously, or you may interpret it as an interesting change of events when it occurs in a movie (Rudy, 2009).

The following sections of this introduction first outline a field of research where operationalizing context is an established part of investigating the underlying mechanism. Next, the neural representations of context are elaborated. This is followed by a brief review of other fields in which the inclusion of contextual aspects has improved our understanding of specific cognitive processes. Finally, the research questions of this cumulative dissertation are addressed, which focus on the role of certain aspects of context in cognitive processes.

Operationalizing context plays a key role in studying the underlying mechanisms of learning and memory (Stark et al., 2018). A classic example of research which employs contextual factors is associative learning, particularly contextual fear conditioning (Fanselow, 2010; Maren et al., 2013; Stark et al., 2018). In contextual fear conditioning, the spatial aspect of context plays a central role in the experimental design. Typically, the process begins with introducing a neutral environment, such as a specific chamber or room, which serves as the contextual background. Encoding a specific context is relevant for the following conditioning of the context. The context conditioning consists of pairing the encoded context with an aversive stimulus, often an electric shock, to establish a learned association between the context and the threat. After a retention period, the participant is reintroduced to the same environment without the threat being present. The response to the context, such as freezing behavior in rodents or physiological indicators like skin conductance in humans, is measured to assess the strength of the memory formed (Curzon et al., 2009; Maren et al., 2013; Milad et al., 2009). These experiments have been crucial in understanding how spatial contexts are encoded, stored, and retrieved in the brain, overall contributing to a better understanding of various disorders related to fear conditioning and extinction processes such as post-traumatic stress disorder (PTSD; Maren et al., 2013; Milad et al., 2009). Furthermore, acknowledging the importance of context allowed for a deeper examination of how the brain handles environmental changes, detects similarities across events, and supports spatial navigation, to name just a few (Lisman et al., 2017; Schwabe & Wolf, 2009; Stark et al., 2018).

Studies that operationalize context in learning and memory have made significant contributions on understanding of how contextual processing is represented in the brain. The hippocampus, located at the top of a hierarchically large cortical system, was found to play a key role in the contextual aspects of memory (Eichenbaum et al., 2007). The hippocampus is well known for its role in memory and spatial navigation. For the contextual representation, the hippocampus was proposed to be involved for processing for example object, spatial and temporal contextual associations, as well as monitoring of environmental changes (Stark et al., 2018). However, the stand-alone role of the hippocampus has been questioned and it has been proposed that a whole network consisting of different regions inside and outside the Medial temporal lobe (MTL), is involved in processing contextual representations (Eichenbaum et al., 2007; Maren et al., 2013). It is proposed that other regions within the MTL form an interconnected system via structural projections from and feedback to hippocampal regions (Eichenbaum et al., 2007). The parahippocampal cortex (or postrhinal cortex in rodents) was found to be involved in object-location (where) associations, while the perirhinal cortex contributes to the identification of objects (what) (Eichenbaum et al., 2007; Sommer et al., 2005; Staresina et al., 2011). Additionally, the entorhinal cortex, which receives input from both the parahippocampal and the perirhinal cortices and is closely connected to the cornu ammonis (CA)1 hippocampal subfield, appears to play a crucial role in detecting changes and retrieving contextual information (Eichenbaum et al., 2007; Knierim et al., 2014; Stark et al., 2018). When emotional contextual information is processed, the amygdala comes into play, again in close interaction with the hippocampus (Kim & Cho, 2020; Zheng et al., 2022). Furthermore, Place together with Eichenbaum and colleagues (2016) proposed a bidirectional interaction between the hippocampus and the prefrontal cortex that supports successful memory retrieval, suggesting a coordinated system for managing contextual information (Lisman et al., 2017; Place et al., 2016). From studies focusing on neural oscillatory activity, cross-frequency coupling between theta and gamma oscillations has been closely linked with contextual representations, particularly in context-dependent episodic memory effects (Staudigl & Hanslmayr, 2013) and context-dependent facial processing (Zheng et al., 2022). This coupling is hypothesized to be linked with hippocampal activity (Staudigl & Hanslmayr, 2013; Zheng et al., 2022), again highlighting its role in complex contextual information processing.

From areas of research where contextual aspects have been extensively utilized to study specific cognitive processes, the focus in the next section changes to fields where context was initially not in the center of attention of research. However, recognizing the relevance of contextual aspects in these fields has substantially improved the understanding of certain mechanisms. For example, it has been shown that perceptual processing of objects (Oliva & Torralba, 2007) or facial expressions (Mobbs et al., 2006) is tremendously influenced by the situational context the stimuli are presented in. Situational context even has an effect on how painful stimuli are perceived, which has been observed in studies on the placebo effect (Wager & Atlas, 2015; Wittkamp et al., 2024). These findings further emphasize the relevance of contextual factors in cognitive processes.

A field of research that has emphasized the crucial role of context in the investigated processes is drug addiction. Studying the mechanisms of drug self-administration and reinforcement consisted for a long time of studying the intake of different substances in socially isolated animals, such as rats (Fredriksson et al., 2017). However, whether these methods sufficiently reflect addictive behavior in humans is questionable (Gage & Sumnall, 2019; Heilig & Epstein, 2016). One of the first scientist changing the social and environmental context in such studies was Bruce K. Alexander (Gage & Sumnall, 2019). When Alexander and colleagues changed the environmental context of the experiment in their 1978 study, they found a striking change of the behavioral outcome. The rats housed in a socially enriched environment preferred water over morphine in comparison with the rats kept in isolation (Alexander et al., 1978; Gage & Sumnall, 2019). Current research acknowledges the social component as one of the critical factors influencing the development, treatment, and relapse of addiction capable of exacerbating as well as mitigating its occurrence in humans (see for example: Nasir et al., 2011; Smith, 2012; Strickland & Smith, 2014; Venniro et al., 2018). Importantly, extensive and innovative research on the role of contextual aspects in addiction has shifted the understanding of addictive behavior from being viewed solely as a biological reaction to substance use to recognizing it as a multifaceted and complex process (Gage & Sumnall, 2019).

After an introduction to the basic definition of context, an overview of the common operationalization of context in research was provided. This was followed by a brief introduction of the neural representations of context. Finally, areas in which the consideration of context has provided deeper insights into the the underlying processes, such as drug addiction, were elaborated. In the next section, the motivation and aim of this cumulative dissertation are addressed.

Given that context is inherently broad, encompassing a wide range of different aspects, it seems fair to propose that there are certainly cognitive processes that are influenced by contextual factors that have been overlooked or under-researched to date. However, exploring the effects of under-researched contextual factors on cognitive processes could help to fill gaps in the literature or enhance our understanding of these processes, as seen in studies on drug addiction.

Therefore, the aim of this dissertation was to examine cognitive processes and the potential influence of contextual factors on these processes. To achieve this, a variety of methodologies were employed, including virtual reality (VR), electroencephalography (EEG) and functional magnetic resonance imaging (fMRI).

In Study 1, EEG was used to examine the role of oscillatory activity (internal contextual factor) in influencing subsequent facial emotion processing and ngeativity bias.

In study 2, VR was used to investigate the probable influence of environmental contextual factors on the transition from implicit to explicit memory during learning.

In study 3, fMRI was utilized to study the neural mechanisms of distinct crossmodal correspondence (CMC) effects. The role of cultural and environmental contextual factors as drivers for the effect was explored.

In the following sections, the content of each study is presented. This is followed by a brief general discussion of the wider implications of context as well as future directions in terms of the findings of the conducted research.

2. Study overview & Findings

2.1. Study 1: Processing of emotional facial expressions is affected by preceding oscillatory activity

Jaap C, Rose M. (2024). Relevance of pre-stimulus oscillatory activity for the perceived valence of emotional facial expressions. Scientific Reports. 14(1):19263.

2.1.1. Introduction

Faces contain highly relevant social information such as a person’s affective state (Kappas et al., 2013). Here, stimulus independent contextual factors such as mood (Manierka et al., 2021), stress (Harp et al., 2022) or social connectedness (Neta & Brock, 2021) strongly contribute to the perception of emotional facial expressions. The modulatory effect of these stimulus-independent factors on the processing of facial expressions are likely reflected in pre-stimulus oscillatory activity. However, research directly investigating the relevance of pre-stimulus neural activity for emotion processing has not been performed yet.

Nevertheless, findings of studies in which the anticipation of an emotional or facial stimuli was altered, i.e. manipulated, suggested that pre-stimulus oscillations vary depending on the anticipated emotional content (Roehe et al., 2021; Sussman et al., 2016). That internal contextual factors such as pre-stimulus oscillations significantly modulate the processing of forthcoming stimuli has been already emphasized by findings in other domains such as pain perception. Here it has been demonstrated that oscillatory activity preceding a painful stimulus has a predictive function (Taesler et al., 2019).

Despite the human proficiency in identifying facial expressions fast and precise, an interpretation bias arises under certain conditions, for example when an expression is ambiguous (Neta & Whalen, 2010). This interpretation bias is reflected as a shift in valence ratings of, for example, a facial expression in a more positive or more negative direction than intended. The often observed interpretation bias in facial expressions often leans in a negative direction and is therefore typically referred to as negativity bias (Neta & Whalen, 2010; Petro et al., 2018). In mental disorders, such as major depression, this negativity bias is commonly observed and is proposed to be a hallmark in these disorders (Beck, 2008; Joormann & Gotlib, 2006; Penton-Voak et al., 2017). However, this bias is not limited to clinical populations. In healthy individuals, a negativity bias is observed when ambiguous expressions, such as faces expressing surprise, are presented without a specific context (Neta et al., 2021; Neta & Whalen, 2010).

In both healthy participants and patients, it was observed that the negativity bias can be modified by stimulus independent factors. In patients a negativity bias was observed to decrease under medication (Harmer et al., 2009) as well as after cognitive bias modification (Li et al., 2023). In healthy participants, mindfulness based stress reduction (Harp et al., 2022), as well as expectation modulation utilizing the placebo effect shifted a negativity bias in a positive direction (Baker et al., 2022). A study by Baker, Brassen, and Gamer (2022) utilized the placebo effect to manipulate healthy participants' anticipation of facial expressions, which led to a positive bias correlated with significant mood enhancement (Baker et al., 2022).

In major depression, treatment expectations contribute significantly to treatment efficacy, with evidence suggesting that up to 80 % of the effects of medication may be attributable to the placebo effect (Kirsch, 2019; Kirsch et al., 2002). Gaining a better understanding about the neural underpinnings of negativity bias is crucial for developing more effective treatments for depression and other mood disorders, as this bias plays a central role in these conditions (Joormann & Gotlib, 2006; Penton-Voak et al., 2017).

Given that stimulus-independent factors have been observed to influence negativity bias in processing of emotional facial expressions, we propose that oscillatory brain activity preceding stimulus presentation may already reflect the negativity bias observed in ambiguous facial expressions.
Using EEG, we address the questions of whether pre-stimulus oscillatory activity is relevant for the perceived valence of subsequent ambiguous facial expressions and whether this pre-stimulus oscillatory activity correlates with the commonly found negativity bias. To test this, all participants were presented with facial stimuli that expressed surprise. We utilized surprise as facial expression of choice since surprise can evoke both positive and negative interpretation. Furthermore, surprise has been implemented successfully in previous studies on biased emotion processing (Neta et al., 2021; Neta & Brock, 2021; Park et al., 2016). Subsequent to the facial expression, the participants were asked to classify the perceived valence of the faces. In order to reach a more general conclusion about the relevance of pre-stimulus neural activity in the latter processing of emotions, we used multivariate pattern analyses (MVPA) in addition to a classical univariate analyses. Using MVPA as method to analyze data have gained significant popularity in neuroscience over the past decades, and studies utilizing this method demonstrate that MVPA is sufficient for decoding complex neural data (Peelen & Downing, 2023). In recent studies focusing on the neural processing of emotions, MVPA was already successfully utilized to decode emotion categories in the brain from fMRI data (Liang & Liu, 2020; Saarimäki et al., 2022). The multivariate approach may also reflect the multiple contributors that can be assumed to constitute emotional processes and the corresponding modulatory factors. An elaborated introduction to this study can be found in Jaap & Rose, 2024b.

Study aims: The aim of this EEG study was to address the questions of whether pre-stimulus oscillations are relevant for the perceived valence of subsequent facial expressions and whether these pre-stimulus oscillations correlate with the commonly found negativity bias for ambiguous emotional facial expressions.

2.1.2. Methods

Participants, study design and behavioral data

Fifty-eight young and healthy participants took part in this study. Our main interest was to investigate the influence of pre-stimulus oscillatory activity on the valence processing of facial emotions. Therefore, EEG was chosen as the appropriate method for this purpose. In two separate within subject experiments, participants were presented with faces expressing surprise. Surprised faces were chosen as they can be interpreted as either negative or positive in valence, which was a crucial aspect that the facial stimuli had to fulfill. The use of surprised facial expressions is already established when studying valence processing of human faces (Neta et al., 2016). Our study consisted of two experiments. In Experiment 1, different artificial faces created with Faceprime were presented. In Experiment 2, faces of the Karolinska database were presented. These two distinct sets of facial stimuli were chosen to replicate the study within participants. According to the existing literature, emotional facial expressions are well recognized on both artificial and natural faces (Dyck et al., 2008). The objective of implementing two different sets of facial expressions was not to investigate the differences between them but rather to assess the probable involvement of pre-stimulus neural activity on emotion processing independent of the quality of the presented face, i.e. natural or artificial. In both experiments, the task of the participants was to rate the valence of a presented face as either positive, negative or neutral (see Fig. 1). To minimize forced choice valence ratings, the option for neutral was offered, however, the participants were informed that no actual neutral faces were presented during the experiments. The valence ratings were later used to divide the data into two groups (positive and negative ratings) as well as to perform analyses concerning a negativity bias.

As part of the study, participants filled out the BDI-2 and the STAI. The scores of these questionnaires were used as an exclusion criterion. To analyze the behavioral data, the percentage of negative ratings in relation to all ratings was computed for each participant in each experiment. Then, the means and standard deviations of the percentage of negative ratings relative to all ratings separately for each experiment was calculated. To test whether there is a correlation between the ratings and the scores of the STAI and BDI-2, a Pearson correlation coefficient between the percentage of negative ratings (in relation to all ratings) and the BDI-2 as well as the STAI scores within each experiment was computed. The behavioral data (ratings & questionnaire scores) was analyzed using R.

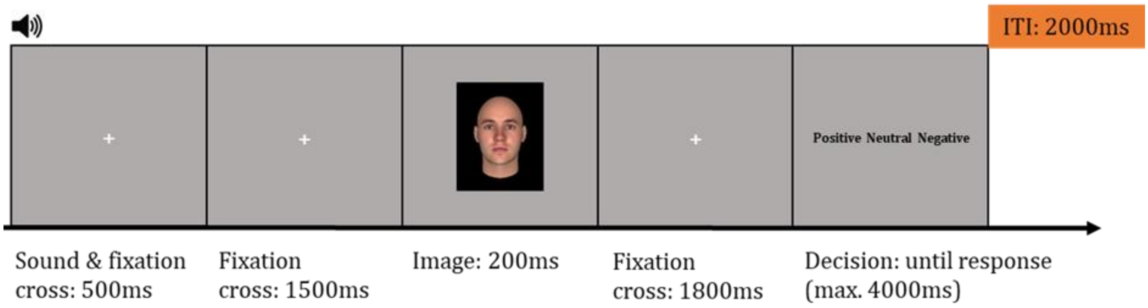


Figure 1. Timeline of a trial in the experiment. The image shown here is an example used in Experiment 1. This timeline was identical in both experiments. This image is reused from Jaap & Rose, 2024 licensed under CC by 4.0. Deed - Attribution 4.0 International - Creative Commons

Technical implementation

The EEG signal was recorded from 64 active electrodes (actiCap slim, Brain Products, Gilching, Germany) which were placed in accordance with the 10–20 system. The ground electrode was placed at the inion. The online reference was placed at FCz. Re-referencing to a common average reference was performed offline. Three EOGs were placed on the face to monitor eye movements. Two EOGs were placed on the canthi of the eyes and one under the left eye. Additionally, Fp1 was used as EOG. High-viscosity electrolyte-gel for active electrodes (SuperVisc, EasyCap GmbH) was used to fill the electrodes with. Electrode impedances were kept below 10 kΩ. The EEG signal was recorded with a sampling rate of 500 Hz and amplified with a low cutoff frequency of 0.5 Hz. The recording took place in an electrically shielded and sound-attenuated room.

The EEG signal was analyzed offline using the FieldTrip toolbox58 (Version 20220632; https://www.fieldtriptoolbox.org/) implemented in MATLAB (Version 2020a, The MathWorks Inc, Natick, Massachusetts, USA). Manual artifact rejection was applied, which included elimination of trials that contained muscle noise, eye blinks in close proximity to the stimulus onset. Also, trials in which the behavioral responses were given too early, i.e. responses given more than 200 ms before the decision window began, were excluded from further analyses. Following manual artifact rejection, a manual independent component analysis was performed to remove eye movement, blinks, cardiac, and muscle artifacts. The re-referencing was then performed followed by another manual artifact detection to remove any remaining artifacts. After the pre-processing was finished, the remaining trials were separated based on the valence response of the participant, i.e. negative and positive ratings of the faces.

Time frequency analysis

The EEG data of each participant where decomposed. The encompassed frequencies ranged from 2 to 120 Hz with a frequency resolution of 1 Hz. A wide range of frequencies was included because there was no previous study on which assumptions could have been based on. Epochs ranged from -2000 to 2000 ms, which ensured that pre- and post-stimulus effects were addressed.

Group-level analysis of time–frequency decomposed data

To test for differences between negative and positive rated facial expressions of the time-frequency spectra preceding and subsequent to stimulus presentation, a classical univariate analysis was performed. Paired samples t-tests were performed on the power spectrum of the temporally and spectrally decomposed data for the two emotional valences, i.e. negative and positive, on the group level. Then, the calculated t-values were corrected using non-parametric cluster-based permutation tests, which are available in the fieldtrip toolbox (Maris & Oostenveld, 2007). The analysis was conducted over a time window of 500 to 1500 ms with reference to the onset of the stimulus which was at 0 ms.

Correlation between oscillatory data and interpretation bias

We aimed to determine whether negativity bias correlates with the differences of the time–frequency spectra of the two perceived valences of facial expressions, independent of the different facial stimuli used in the two experiments. For this purpose the time–frequency spectra of both experiments were combined, followed by testing for a relation of the negativity bias and the difference in oscillatory power in perceived valence. For this purpose we used the implemented correlation function in FieldTrip (Oostenveld et al., 2011; https:// www. fieldtriptoolbox.org).

Multivariate Pattern Analysis of time–frequency decomposed data

To achieve a more general insight about the relevance of pre-stimulus neural activity in emotion processing, we used two different multivariate classification approaches. One approach was performed at subject- and the other at group-level. The latter was additionally conducted in order test whether a participant and stimulus overarching pattern in the pre-stimulus period between differently perceived valences of subsequent presented emotional facial expressions exists in our data. The multivariate approach also addresses the multiple contributors that can be assumed to constitute emotional processes and corresponding modulatory factors including stress (Harp et al., 2022), anxiety (Park et al., 2016) and mood (Manierka et al., 2021).

In both of the MVPA approaches, the classifier, a support vector machine, was trained on data containing only the pre-stimulus oscillatory activity (time-window: -1000 to -200 ms) from Experiment 1 and was then tested on data from Experiment 2 containing again only the pre-stimulus oscillatory activity (time-window: -1000 to -200 ms). This cross-decoding approach was chosen to determine whether a general pattern could be identified that did not depend on the type of face on which the emotional expression was shown, i.e. natural or artificial. The idea for this approach stems from the 2023 Paper of Jin and colleagues on study independent mind wandering (Jin et al., 2023).

2.1.3. Results

In both experiments, an alpha-band effect (~8-12 Hz) within -500 to 0 ms relative to stimulus onset was observed, primarily in the left centro-parietal regions. In Experiment 1, the alpha-band effect was accompanied by an effect in the theta frequency band (~3-7 Hz), predominantly in the left centro-parietal and mid fronto-central regions (see Fig. 2). In Experiment 2, this alpha-band effect was accompanied by increases in beta (~22-24 Hz) and gamma (~30-35 Hz) frequency bands, particularly in the left centro-parietal, left central and right frontal regions (see Fig. 3).

The results of the MVPAs performed exclusively on pre-stimulus data (-1000 to -200 ms in relation to the stimulus onset) show a successful classification of later rated negative and positive expressions already from the pre-stimulus data (see Fig. 4). This is indicated by an accuracy above chance level (.5) for each participant in the MVPA performed on subject level, as well as an accuracy above chance level (.5) for each iteration in the MVPA performed at group-level (see Fig. 4).

Across all participants ($t_{35} = 26.856$, $p < .001$) and across all iterations ($t_{35} = 21.4$, $p < .001$) these accuracies also significantly differed from chance level.

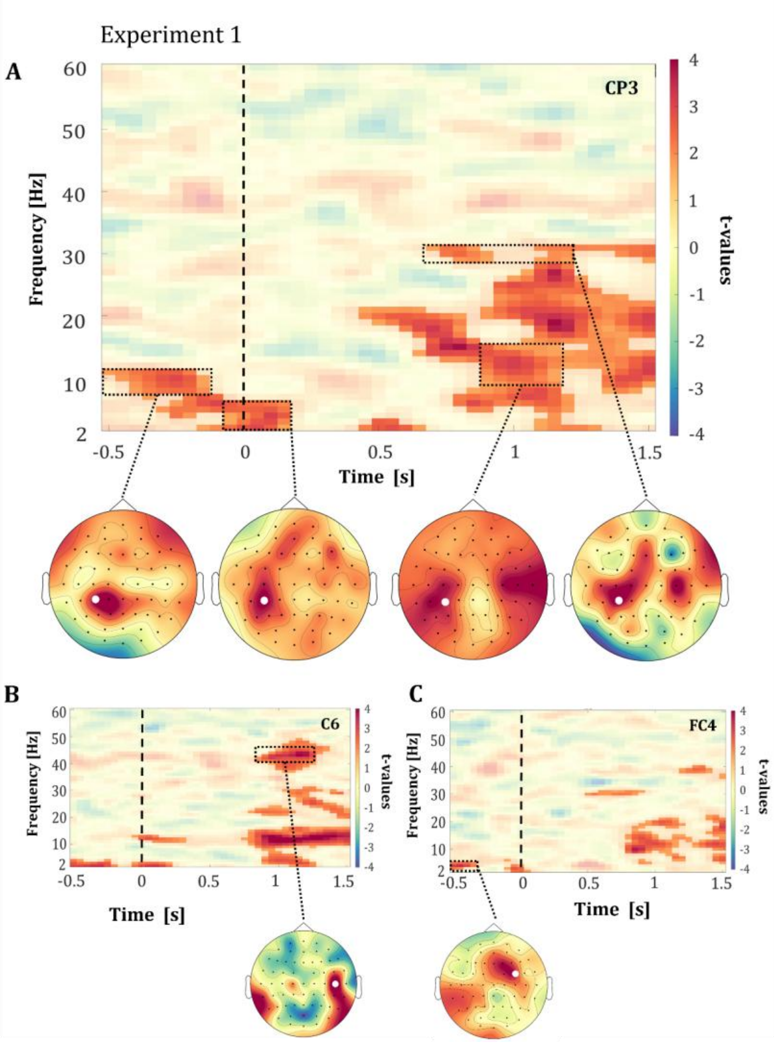


Figure 2. Time-frequency representations (TFRs) showing the t-values of difference in power between negatively and positively rated facial expressions across all participants for Experiment 1 in A) channel CP3, B) channel C6, and C) channel FC4. The natural facial expression was presented for 200 ms and the dashed line marks its onset at 0 ms. The significant t-values ($p < .05$, corrected) are represented as highlighted areas. The spatial distribution of the t-values in the respective time-frequency windows, marked by a dotted rectangle, is shown in the topographic plots. The white dot in the topographic plots marks the position of the channel that is shown in the respective TFR. This image is reused from Jaap & Rose, 2024b licensed under CC by 4.0. Deed - Attribution 4.0 International - Creative Commons

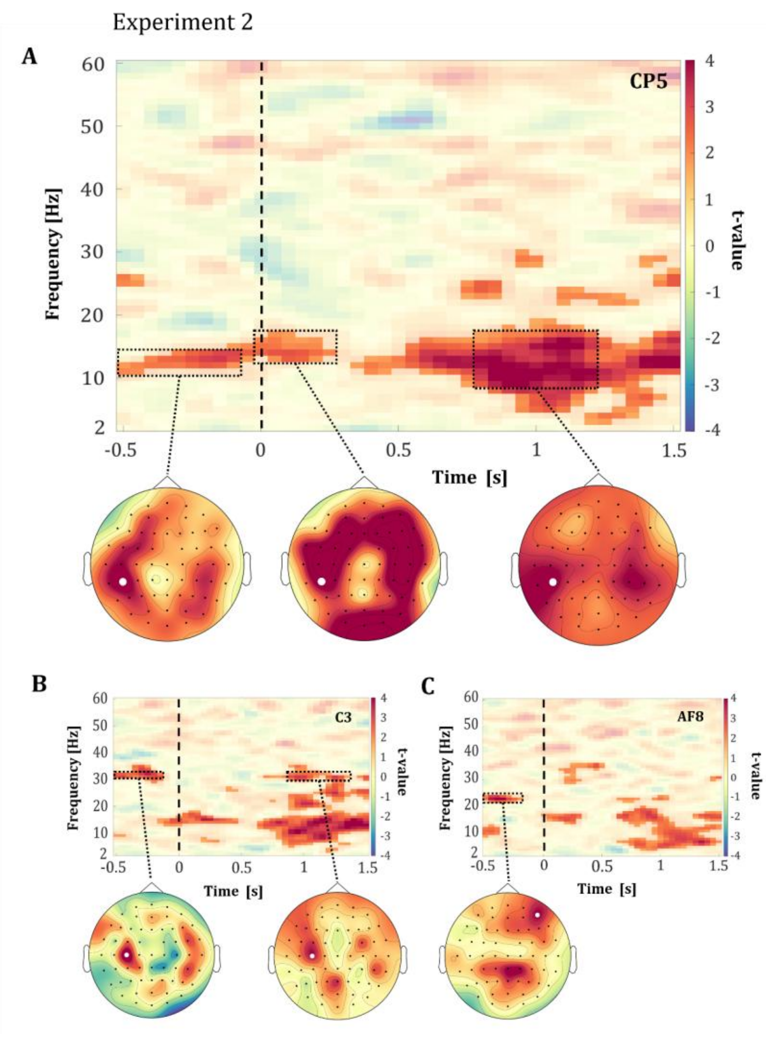


Figure 3. TFRs showing the t-values of the difference in power between negatively and positively rated facial expressions across all participants for Experiment 2 in A) channel CP5, B) channel C3 and C) channel AF8. The natural facial expression was presented for 200 ms and the dashed line marks its onset at 0 ms. The significant t-values ($p < .05$, corrected) are represented as highlighted areas. The spatial distribution of the t-values in the respective time-frequency windows, marked by a dotted rectangle, is shown in the topographic plots. The white dot in the topographic plots marks the position of the channel that is shown in the respective TFR. This image is reused from Jaap & Rose, 2024b licensed under CC by 4.0. Deed - Attribution 4.0 International - Creative Commons

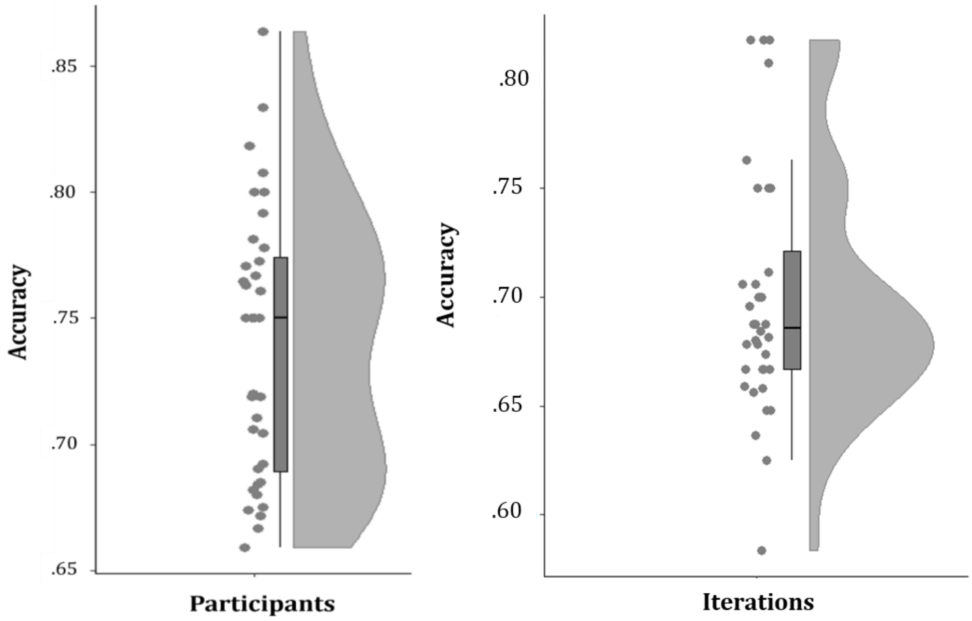


Figure 4. The distribution of peak accuracies per participant in the binomial statistic obtained from the within subjects MVPA performed across experiments (left). The distribution of peak accuracies per iteration in the binomial statistic obtained from the between subjects MVPA performed across experiments (right). The dots represent the peak accuracies of the participants (left) or iterations (right). The results indicate that both MVPAs performed well, indicated by accuracies above chance level (.5). The image on the left is reused from Jaap & Rose, 2024b licensed under CC by 4.0. Deed - Attribution 4.0 International - Creative Commons

A clear behavioral negativity bias was observed for both artificial and natural face presentations expressing surprise (see Fig. 5). No significant correlation between the ratings and the BDI-2 or STAI-S/STAI-T scores were observed in neither of the two experiments.

A positive relationship between neural data and negativity bias was found throughout the pre-stimulus and post-stimulus periods over theta to gamma frequency bands. The effects in the pre-stimulus period were observed in the 500 ms that preceded stimulus onset (see Fig. 6).

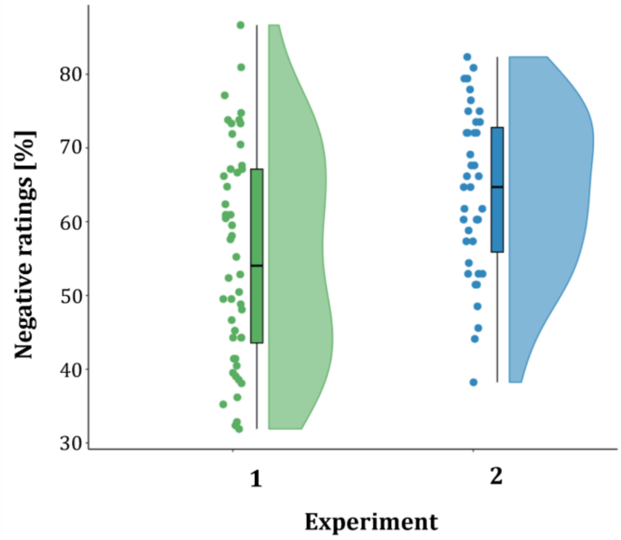


Figure 5. The distribution of negative ratings as a percentage of all ratings for faces expressing ambiguous emotions in Experiment 1 (N = 48; data range = 32 - 87%; left) and Experiment 2 (N = 39; data range = 38 - 82%; right). Each point represents the rating of a participant. This is image is reused from Jaap & Rose, 2024b licensed under CC by 4.0. Deed - Attribution 4.0 International - Creative Commons

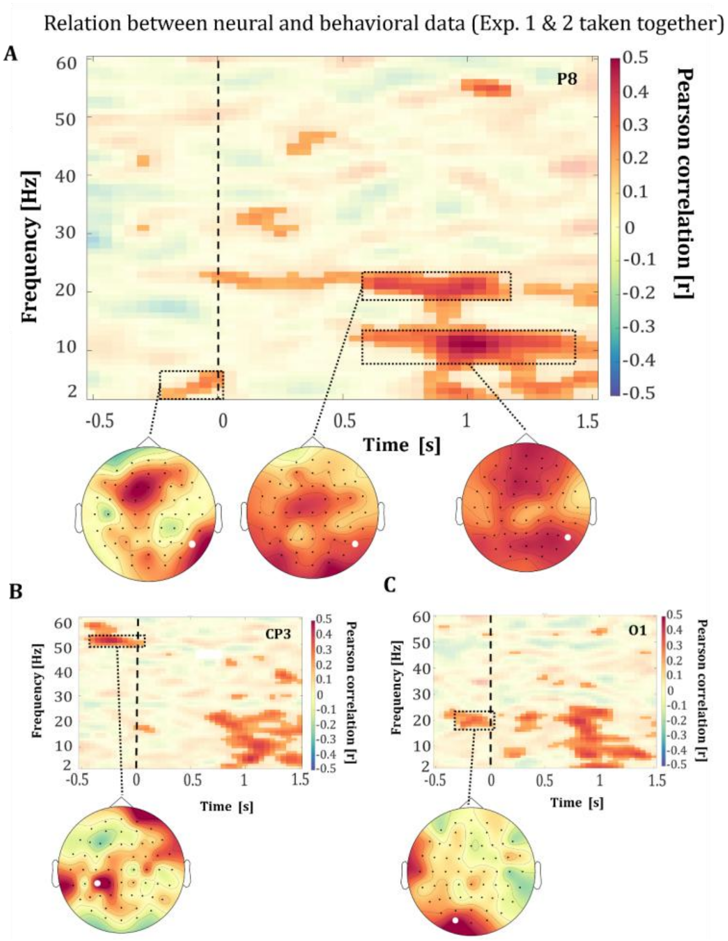


Figure 6. TFRs illustrating the representation between the negative ratings of facial expressions in percent in relation to all ratings and the difference (between negatively and positively rated facial expressions) in oscillatory power of the face ratings in A) in channel P8, B) channel CP3, and C) channel O1. The natural facial expression was presented for 200 ms and the dashed line marks its onset at 0 ms. The spatial distribution of the t-values in the respective time-frequency windows, marked by a dotted rectangle, is shown in the topographic plots. The white dot in the topographic plots marks the position of the channel that is shown in the respective TFR. This is reused from Jaap & Rose, 2024b licensed under CC by 4.0. Deed - Attribution 4.0 International - Creative Commons

2.1.4. Discussion

Surprisingly strong effects preceding stimulus presentation were observed in both eperiments. These findings were confirmed by both the univariate and the multivariate analysis approaches. This is a novel finding in regard to facial expression processing. Furthermore, the current study revealed differences in valence processing related to negativity bias during the pre-stimulus phase in healthy participants. The results of this study allow the assumption that overarching patterns exist in oscillations during the pre-stimulus period between differently perceived valences of emotional facial expressions.

The results of our study emphasize the role of the alpha band in the processing of differently perceived valences of emotional facial expressions. Previous research on affective stimuli (Onoda et al., 2007; Strube et al., 2021) and face processing (Roehe et al., 2021) has elucidated the involvement of alpha oscillations in predictive processing of emotional or face stimuli. However, in the present study, the expectation of the valence of the upcoming facial expressions was not manipulated in either of the experiments. While our study does not provide definitive evidence, it can be assumed that the differentiation observed in the alpha band prior to the stimulus is related to individual, fluctuating and valence dependent expectations. Conversely, fluctuations in emotional states could also explain the effects observed during the pre-stimulus period. The latter assumption is supported by a former study in which spontaneous emotional states occurring during resting-state were observed (Kragel et al., 2016). On this basis, we assume that spontaneous fluctuations in emotional states may influence the subsequent processing of facial expressions in our study. These fluctuations may also be related to the observed effects in the oscillatory activity preceding the stimulus.

The expected negativity bias was found for both artificial and natural facial expressions. Our study revealed differences in valence processing related to negativity bias during the pre-stimulus phase. Interestingly, the results indicate that the observed positive association of neural activity with negativity bias was independent of mood, as evidenced by the non-significant associations of negativity bias with STAI-S/-T and BDI-2 scores. This suggests that the observed effects may arise from trial-by-trial fluctuations in oscillatory states that directly influence the biased processing of ambiguous facial expressions. However, it remains unclear what factors apart from mood might influence these fluctuations. Recent studies suggest a significant influence of various factors on the negativity bias in healthy participants, including social connectedness, stress and emotional regulation skills (Harp et al., 2022; Neta & Brock, 2021). Given these observations, other factors such as stress, which were not monitored in our study, may have contributed to the neural effects of biased valence processing observed in this study.

The findings of this study underline the importance of addressing pre-stimulus activity as modulatory aspect of emotion processing in future studies. For mental disorders such as depression, in which negativity bias plays a crucial role, understanding the role of pre-stimulus processes in negativity bias could improve our knowledge of the mechanisms underlying these disorders.

2.2. Study 2: On the transition from implicit to explicit knowledge in enriched versus non-enriched environments

Jaap, C., Maack, M. C., Taesler, P., Steinicke, F., & Rose, M. (2022). Enriched environments enhance the development of explicit memory in an incidental learning task. Scientific Reports, 12(1).

2.2.1. Introduction

Learning takes place in an implicit (unconscious) or explicit (conscious) way and is influenced by several factors such as attention and motivation (Anderson, 2016). There is evidence that implicit acquired knowledge, such as incidentally learned regularities in our environment, can become explicit (Clos et al., 2018; Wessel et al., 2012). This enables the conscious use of this knowledge, which is crucial for higher cognitive functions such as logical thinking and language (Rose et al., 2010).

In our study, two different hypothesis and their contribution to the transition from implicit to explicit knowledge where compared. The first hypothesis is the fluency hypothesis. The underlying assumption for why implicit knowledge becomes explicit is based on the continuous evaluation of predictions. A sequential task in which the sequence is predictable to a certain degree is usually used to study this effect (Rose et al., 2010). At a certain point in such tasks, a discrepancy between the predicted speed and the actual speed of responses to the sequential stimuli can be detected by the participants. This difference in predicted in actual speed of responses may trigger a search for the cause of this discrepancy. This so called fluency hypothesis states that the transition from incidental learning to explicit memory is based on the fluency of ongoing evaluation of predictions (Rose et al., 2010; Ru & Frensch, 2008).

The second hypothesis we tested as driving factor for the transition from implicit to explicit knowledge was the enriched environment hypothesis. In the last decades, there has been a shift towards implementing more complex and enriched settings in learning and memory research in order to get a better picture of learning mechanisms in real-life like scenarios (Hölle et al., 2021; Salvidegoitia et al., 2019; Schwabe & Wolf, 2009). The challenge in such task designs is the need for controllable environments in order to study learning mechanisms while implementing enrichment at the same time. A solution to this problem is using VR.

In humans, it has been shown that learning in near-natural settings, such as settings created with VR, leads to a better performance compared to classical test environments. This is explained by the immersive and engaging (Lüddecke & Felnhofer, 2022) learning environment created with VR, which likely elevates motivation in participants (Anderson, 2016; Chelazzi et al., 2013; Sanz et al., 2018). This motivational effect may also enhance the detection of altered predictions due to an increased level of attention instead of performance fluency. The enriched environmental hypothesis posits that the identification of unexpected events during ongoing prediction evaluations, which are essential for the transition from implicit to explicit knowledge, would be facilitated by the near-natural experimental application in VR, as evidenced by enhanced explicit memory performance. An elaborated introduction to this study can be found in Jaap et al., 2022.

Study aims: The aim of this study was a direct comparison of the fluency and the enriched environment hypothesis and their contribution to the transformation from implicit to explicit learning. It can be assumed, that the fluency of stimulus processing is improved by our implemented conventional 2D-computer based application. This assumption is mainly based on the smooth and fast transition between trials. In our VR application, on the other hand, the evaluation of the stimulus material as well as the response is slowed down. The latter one is for example due to the difference between the responses devices used in the applications (controller in VR and keyboard in the conventional application). We hypothesized therefore to observe a greater amount of explicit knowledge in the conventional 2D-computer based application, if fluency of stimulus processing is essential for the transition of implicit to explicit knowledge. On the other hand, we hypothesized to find an enhanced emergence of explicit knowledge in the VR application, if the context of the presented (realistic) stimuli directly facilitate the transfer process from implicit to explicit knowledge (S. M. Smith & Vela, 2001).

The contribution of both hypothesis on the transfer from implicit to explicit learning can be evaluated in the distinct experimental contexts by examining the extent of implicit and explicit memory formation following learning in two tasks that were identical across all experimental conditions. Specifically, the tasks were a completion task and a free recall task.

2.2.2. Methods

In this study, a sequential-association task was administered to healthy participants under three experimental conditions: A virtual reality (VR, 520 trials, N = 49) application, a short version (PC-short, 260 trials, N = 21), and a long version of a conventional 2D computer-screen based application (PC-long, 520 trials, N = 27). The eight digit sequence in each of these experimental conditions consisted of alternating trials of four different acoustic and four different visual stimuli. In 15 % of the trials, the stimulus following the last was a deviant and not part of the sequence. Deviants have been implemented to prevent explicit knowledge from being acquired too quickly. This between-subjects experiment aimed to explore how different contexts of presentation of stimuli and task lengths impact learning outcomes, particularly the transition from implicit to explicit learning. Given that the evaluation of stimulus and the responses are probably slowed down in VR compared to the conventional application, the longer version of the 2D computer-screen based application functioned as a control group for the VR application. In the PC-long group, participants practiced the same amount of trials as in the VR condition to eliminate the likelihood that experiment duration influences the amount of transferred knowledge. The short version, i.e. PC-short, with half of the trials compared to the long version of the computer-screen based application and the VR application was conducted because other studies, such as the study performed by Rose et al. (2010), have already demonstrated that a smaller number of trials is sufficient enough to develop explicit knowledge in an incidental sequential task.

After a test phase, performed in either the VR or in the conventional 2D computer-screen based application, the participants performed two post-experimental assessments of knowledge. A completion task and a free recall. The completion task (see Fig. 7) consisted of 64 trials. Half of them were crossmodal, the other half consisted of either visual or acoustic only trials. After the completion task, the participants were informed that the stimuli formed a sequence and they were asked to recreate this sequence four times (= four trial) in a free recall. After each trial of the completion task and the free recall, the participants had to rate their confidence on a scale from *unsure* (left) and *sure* (right). This confidence rating serves as the basis for distinguishing between implicit and explicit knowledge of correct responses. Correct and unsure responses were counted as implicit and correct and sure responses were counted as explicit knowledge about the sequence. The participants also performed an emotional salience evaluation in which the visual stimuli were rated on a continuous scale from *I dislike it* (negative rating), over *Neutral* to *I like it* (positive rating). We implemented a stimulus value rating to assess potential motivational significance between the stimuli used in the visually enriched VR experiments and the non-enriched screen-based experiments. This contextual influence is likely to correlate with learning effects driven by motivational and attentional factors (Chelazzi et al., 2013). The post-experimental assessments were identical in both the VR environment and the 2D computer-screen based applications, however the stimuli used in these tasks were different. For participants in the VR group, the visual and acoustic stimuli closely resembled those experienced during the sequential-association task they performed in the VR. In contrast, for participants in the PC group, the stimuli mirrored those presented during the sequential-association task they performed in the PC application.

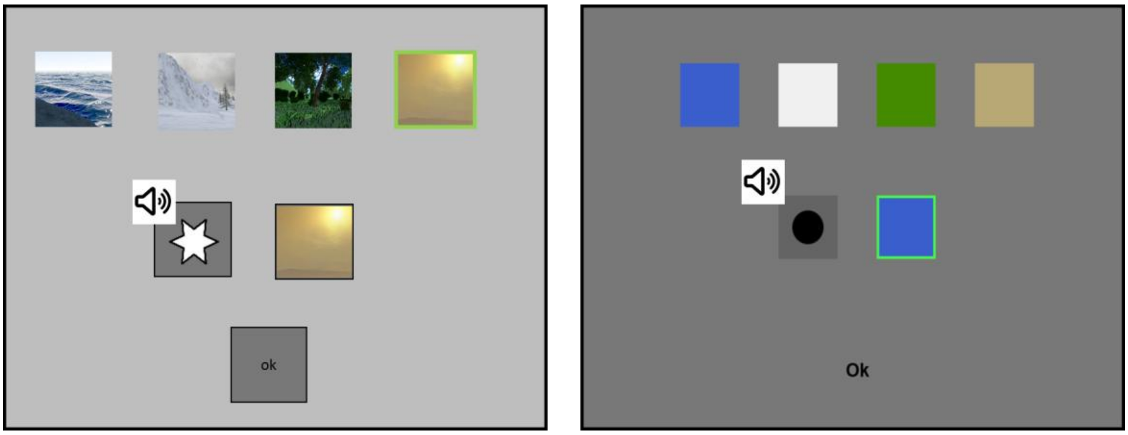


Figure 7. Examples of a crossmodal trial in the completion task, performed after the experiment in the VR environment (left) or the PC application (right). In these crossmodal trials, participants were presented with an acoustic stimulus and had to match it with one of the visual options they associated it with. This image is reused from Jaap et al., 2022 licensed under CC by 4.0. Deed - Attribution 4.0 International - Creative Commons

For the analysis of the completion task and the free recall we used the correct answers and performed an ANOVA with the factors Condition (VR; PC-short; PC-long) and Learning-Type (implicit; explicit) with performance in each learning type as the dependent variable. When analyzing the ratings of stimulus values, i.e. emotional salience, we performed a two-sided two-sample t-test over the mean of the stimulus ratings in VR versus both PC applications taken together. The statistical analysis was performed in R (4.0.5) and Matlab (2020b).

2.2.3. Results summary

The completion task revealed no significant differences in the overall number of learned sequence pairs across the different experimental conditions (VR, PC-short, and PC-long). However, the development of explicit memory varied significantly between these conditions, with higher level of explicit memory in the VR group compared to both PC groups. In general, participants acquired more implicit than explicit memory.

In the free recall, the overall number of learned sequence pairs did not differ between conditions. However, awareness and explicit memory were influenced by the different learning environments. Similar to the completion task, the participants in the VR group generated more explicit memory than the participants in the PC groups.

The analysis of stimulus value, i.e. emotional salience ratings, indicated no significant differences in the mean ratings between the enriched VR environment and the non-enriched PC tasks. This suggests that the differences in explicit memory formation were not referable to the salience of the visual stimuli alone but rather to the enriched nature of the VR environment.

2.2.4. Discussion

The aim of this study was to compare the role of fluency and environmental enrichment on the transfer of implicit to explicit knowledge. The study found that an enriched environment, implemented using VR, had a greater impact on explicit memory than task fluency alone. While the amount of learned sequence pairs were not significantly different between the groups performing our sequential task in an enriched VR environment or in a non-enriched 2D-computer-screen environment, the group that performed the sequential task in the enriched environment showed a higher degree of awareness.

Given that we could not observe a significant difference between the visual stimulus value ratings of the VR compared to the conventional 2D-computer-screen based application the stimulus material alone cannot resolve the difference in the amount of generated explicit knowledge between the conditions. Instead, we assume that a major factor contributing to the transfer from implicit to explicit knowledge is motivation, which is enhanced in environments that are more enriched compared to conventional non-enriched 2D-computer-screen based applications. An advantage, which probably boosts motivation in tasks performed in a VR environment like the one used in this study, is the inclusion of the whole body. Natural movements, such as maintaining an upright body position and allowing freedom of head and arm movement, can make VR applications more motivating, enhancing the sense of embodiment and physical presence in the scenario, which in turn likely improves encoding efficiency and the transition to explicit memory formation (Kilteni et al., 2012; S. A. Smith et al., 2021). The role of motivation in learning is closely tied to the ventral striatum's involvement in the transition from implicit to explicit memory. The ventral striatum, associated with motivation and reward (Pool et al., 2022), releases dopamine in rewarding situations, which likely boosts explicit memory formation. A study by Clos et al. (2018) showed that increasing dopamine levels improved information transfer, enhancing explicit memory during tasks (Clos et al., 2018). Our findings suggest that rewarding environments encourage the recognition of patterns, leading to better explicit memory formation.

2.3. Study 3: Crossmodal correspondence effect

Jaap, C., & Rose, M. (2024). Dissociable neuronal mechanism for different crossmodal correspondence effects in humans. Acta Neurobiologiae Experimentalis, 136–152.

2.3.1. Introduction

When two seemingly arbitrary crossmodal stimuli are perceived as corresponding, this is referred to as crossmodal correspondence, i.e. CMC. Well-studied examples of CMCs are simple pitch-based CMCs, such as the perceived congruity of a small object and a high pitched tone and vice versa, pitch-size CMC or an object elevated in space with a high pitched tone and vice versa, i.e. pitch-elevation CMC (Evans & Treisman, 2010). The phenomenon of CMCs is not exclusive to the domains of acoustic and vision. CMCs have been documented in the domains of olfaction (Guetta & Loui, 2017; Wagner et al., 2021), haptic (Misselhorn et al., 2016; Wagner et al., 2021), and even gustation (Spence et al., 2013). Nevertheless, the field of audio-visual CMCs has been the subject of the most extensive research (Chiou & Rich, 2012; Evans & Treisman, 2010; Parise & Spence, 2012; Spence, 2011; Spence & Sathian, 2020).

Despite audio-visual correspondences were subject of extensive research and are described to be almost universally experienced, the understanding of their origins and underlying neural mechanisms of these CMCs remains sparse (McCormick et al., 2018; Spence & Parise, 2012; Spence & Sathian, 2020). There are several theories regarding the origin of these pitch-based CMCs. Two theories specifically relevant to our study are presented in the following section.

The first theory we are addressing posits that the association between pitch and spatial elevation has its origin in a cultural contextual aspect, i.e. language (Parise & Spence, 2013; Spence & Sathian, 2020). Apart from some exceptions, most cultures use the terms ‘high’ and ‘low’ to describe certain pitches. Although language may be a plausible explanation for the origin of some CMCs, contrasting findings (Korzeniowska et al., 2019; Parkinson et al., 2012) prompt the question of whether other variables than language are responsible for the formation of CMCs (McCormick et al., 2018; Parise & Spence, 2013; Parkinson et al., 2012). Furthermore, this linguistic link is not described pitch and size correspondences.

A second theory suggests that CMCs originate from the statistical likelihood of certain sensory combinations perceived in our environment that are stored in memory (Parise & Spence, 2013; Spence, 2011; Spence & Sathian, 2020). For example, children typically have a higher pitched voice than adults (Lee et al., 1999). In the animal kingdom, it has been observed that vocalization gets lower in frequency with an animal’s body size (Bowling et al., 2017). Also objects elevated in space usually resonate higher pitches than on the ground (Parise et al., 2014). Following this observation, CMCs may have their roots in statistical regularities, i.e. naturally learned rules and assumptions from our environment (Parise & Spence, 2013; Spence, 2011; Spence & Sathian, 2020).

So far there is only one study using fMRI to investigate the neural basis of simple pitch-based CMCs (McCormick et al., 2018). In this study, presumed pitch-elevation congruency were observed bilaterally in several regions, including the inferior frontal gyrus, the insula and the angular gyrus (McCormick et al., 2018). The findings of this first study remained inconclusive regarding the proposed hypotheses. However an addressed issue of the study was the implemented one-back task since the demand for working memory probably interfered with cortical activity during any trial and condition (McCormick et al., 2018). An elaborated introduction to this study can be found in Jaap & Rose, 2024a.

Study aims and hypotheses: In our study the aim was to investigate whether distinct CMCs have common or differentiating underlying neural mechanisms and whether the regions involved in the effect are in favor of a given hypotheses about the origin of the CMC effect. For this purpose two pitch-based CMCs were utilized: The CMC effect regarding the combination of pitches with object sizes and pitches with elevated objects in space. We hypothesize to observe an effect within the left inferior frontal gyrus (IFG), if pitch-elevation CMCs are language driven. We propose to most likely find an effect within areas known for attention and memory retrieval like the anterior cingulate cortex (ACC) or the AnG, if CMCs are based on statistical representations of our environment.

2.3.2. Methods

The study included 33 young, mentally and physically healthy participants. In each experiment, participants were simultaneously presented with a pair of visual and auditory stimuli that were either congruent or incongruent in accordance with the CMC effect (see Fig. 8 a). The visual stimuli were presented via a mirror mounted in front of the participants and the acoustic stimuli were presented via headphones. The task was to press a button which is in accordance with the visual stimulus as fast and precise as possible (see Fig. 8 b).

Behavioral data acquisition and analyses

A comparable investigation of pitch-based CMCs demonstrated that a conventional RT task exhibits suboptimal performance in the scanner (McCormick et al., 2018). We therefore implemented a congruence classification task, which was performed by each participant outside the scanner (see Fig. 8 c). This task was implemented to verify that participants showed the correct CMCs as intended by the in-scanner task. This task was central in our analysis of the behavioral data and served as a key measure for verifying the intended effects in the participants. All statistical tests on the behavioral data were performed in JASP (Version 0.16.1).

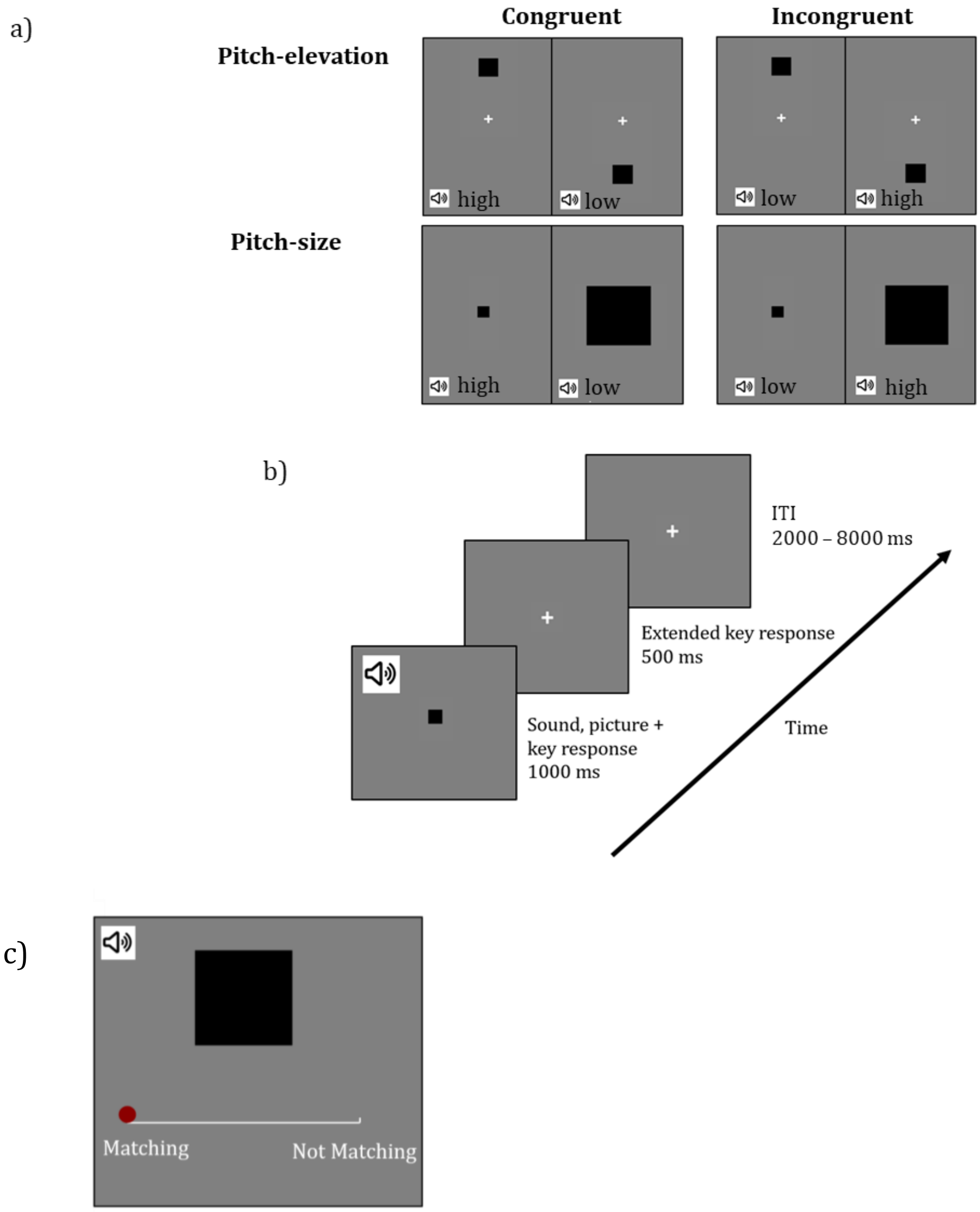


Figure 8. (a) Illustration of the congruent (left) and incongruent (right) stimulus pairs for the CMCs pitch-elevation (top) and pitch-size (bottom). (b) Sequence of events within a trial (here pitch-size). (c) Example of a trial in the matching task, which was performed outside the scanner. This image is reused from Jaap & Rose, 2024a licensed under CC by 4.0. Deed - Attribution 4.0 International - Creative Commons

fMRI data acquisition and analyses

For each condition (congruent; incongruent), the hemodynamic response was modelled as an event-related design. After calculating contrast images per participant from onsets of each condition (main effects), these images were entered into a flexible factorial group level analysis. All further statistical comparisons were estimated on the group level.

A global conjunction was estimated to test for common congruency effects between the distinct CMCs (Pitch-size (C > I) & pitch-elevation (C > I)). To test for differentiating effects between the distinct CMCs, an interaction analysis was conducted. To test selectively for enhanced neural effects of congruent trials in the pitch-size CMC we estimated an interaction contrast (Pitch-size (C > I) > pitch-elevation (C < I)) and vice versa to test for enhanced neural effects of congruent trials selectively in the pitch-elevation CMC (Pitch-size (C < I) < pitch-elevation (C > I)). All reported whole brain fMRI effects were family wise error corrected (FWE, *p* < .05).

There are hardly any studies on pitch-based CMCs performed with fMRI (McCormick et al., 2018; Spence & Parise, 2012). As a consequence, regions of interest (ROIs) were selected based on brain regions that, according to the proposed theories, are probably involved in the processing of the studied CMCs. To test for a statistically or environmentally driven CMC we included the ACC (Roelofs et al., 2006) and the AnG (Humphreys & Ralph, 2015). To test for language-related effects, we included the triangular part and the orbital part of the left IFG (Liuzzi et al., 2017) as our final ROIs. Significant fMRI effects within our ROIs were reported using a sphere with a radius of 10 mm, which was small volume corrected with FWE (p < .05).

2.3.3. Results summary

We were interested in whether participants show the expected classification of pairs in accordance with the CMC theory. Behavioral ratings of congruence indicated a substantial correspondence effect for pitch-size, as well as pitch-elevation pairs (see Table 1).

Table 1. Congruence classification of the distinct CMCs for congruent and incongruent stimulus presentations for pairs classified as matching. Displayed is the mean in percent for stimulus conditions (congruent, incongruent) within each distinct CMC (pitch-size, pitch-elevation) classified as matching, i.e. congruent. The maximum possible percentage for each presented condition (congruent; incongruent) in each distinct CMC was 100 % for

classifications of matching and not matching taken together, whereby only pairs classified as matching are presented.

Condition presented	Pitch-size classified as	Pitch-elevation classified as
	matching	matching
congruent	92	80.5
incongruent	23.6	31

When analyzing the neural data, we were interested in finding the neural components of different CMC types to evaluate different and common neural mechanisms underlying pitch-based CMC effects.

No common effects were found, however, clear differences between the distinct CMCs were observed when we tested for enhanced neural effects of congruent trials selectively in the pitch-size CM (see Fig. 9 & Fig. 10). The interaction analysis pitch-size (C > I) > pitch-elevation (C < I) yielded significant effects in the whole brain (see Fig. 9) as well as certain regions of interest (ROIs, see Fig. 10).

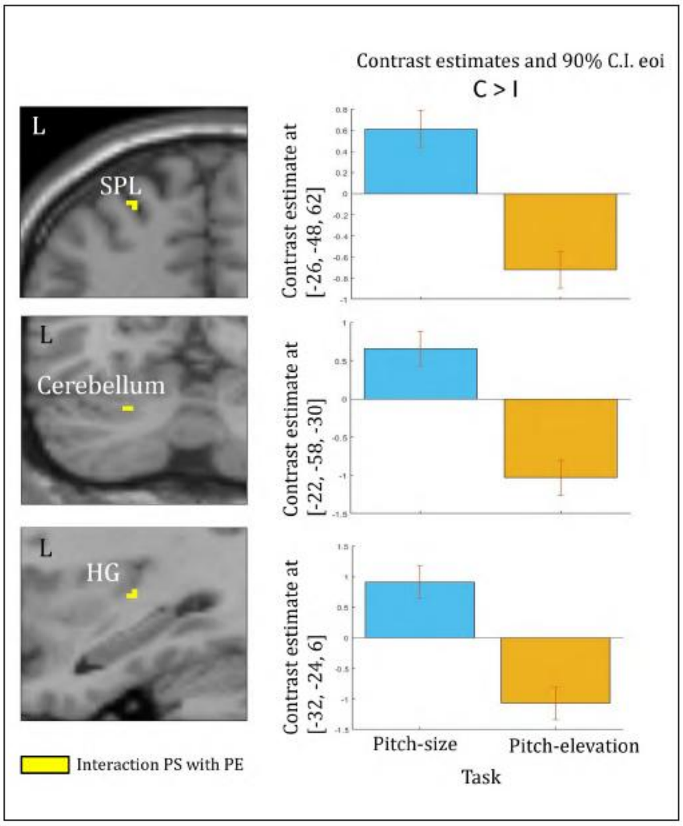


Figure 9. Effects of the interaction for pitch-size with pitch-elevation (left), which tested for enhanced neural effects of congruent trials selectively in the pitch-size CMC. The contrast estimates for the corresponding region are shown on the right (blue = Pitch-size; orange = Pitch-elevation). Each bar reflects to the activation difference (C > I) within the specific region that is displayed on the left. FWE (peak-level; p < .05) corrected whole-brain effects were observed in the SPL, the cerebellum and the HG. Error bars indicate the observed standard error within each contrast. The FWE corrected analysis at p < .05 was used to obtain the above images. This image is reused from Jaap & Rose, 2024a licensed under CC by 4.0. Deed - Attribution 4.0 International - Creative Commons

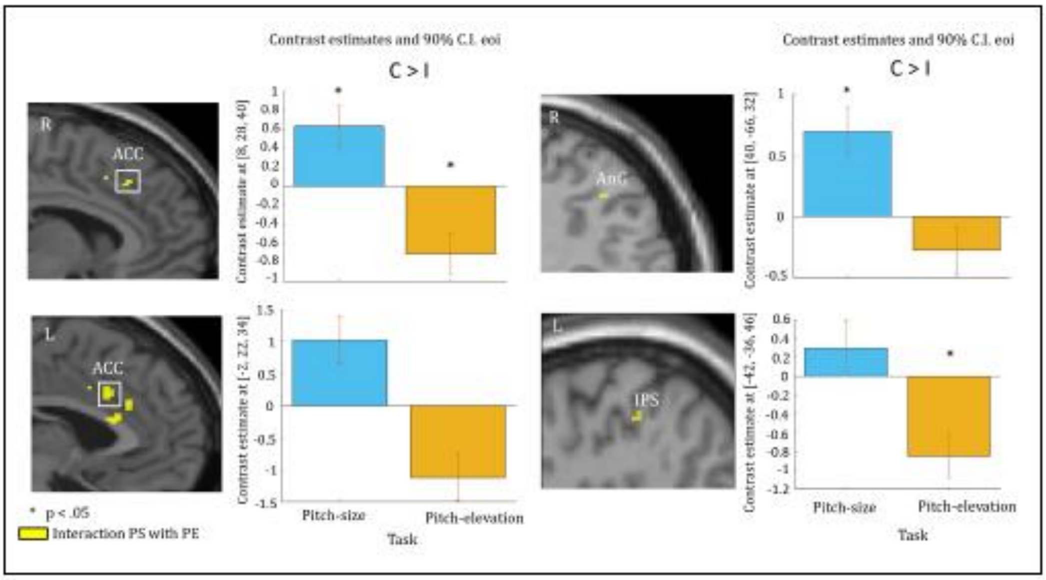


Figure 10. Effects of the interaction for pitch-size with pitch-elevation (left), which tested selectively in the pitch-size CMC for enhanced neural effects of congruent trials. Contrast estimates for the respective region are displayed on the right (blue = Pitch-size; orange = Pitch-elevation). Each bar reflects the difference in activation (C > I) within the specific region. For the ROIs, FWE corrected small volume effects (peak-level; p < .05) were observed in the ACC, the AnG and in the IPS. The white box in the images on the left indicate the area of the ACC which was inside our ROI. The observed standard error within each contrast is indicated by the error bars indicate. A T value of > 4 was chosen to obtain the above images. This image is reused from Jaap & Rose, 2024a licensed under CC by 4.0. Deed - Attribution 4.0 International - Creative Commons

2.3.4. Discussion

Overall, indicated by the classification task performed outside the scanner, participants classified the presented congruent and incongruent stimulus pairs in alignment with the CMC theory. The results suggest differential processing of congruent and incongruent stimuli starting already in early sensory systems. Greater effects in the HG were found for congruent stimuli in pitch-size and for incongruent stimuli in pitch-elevation. The ACC was involved in both CMCs, reflecting its role in the dynamic interplay between top-down and bottom-up multisensory processing (Benedict et al., 2002; Crottaz-Herbette, S; Menon, 2006). The AnG's role in multisensory integration and memory-related attention (Cabeza et al., 2012; Humphreys & Lambon Ralph, 2015) was evident, particularly in the pitch-size CMC.

Addressing the research goal of studying which mechanism potentially drive the studied CMCs, the observed results support the theory that statistical assumptions extracted from our environment provide a general basis for different CMCs given the observed effects in the AnG and ACC (Maimon et al., 2021; Pisanski et al., 2017; Spence, 2011; Spence & Sathian, 2020; Zeljko et al., 2019). However, further research is necessary to verify this assumption. Given that no region demonstrated significant activity supporting the language theory (tested in the left triangular and orbital parts of the IFG), our findings do not support the hypothesis that pitch-elevation CMCs are driven by language.

Addressing the research goal of studying the relationship of different CMCs, we observed no common congruency effects between the distinct CMCs, however clear differential effects were found. Therefore, we argue in favor of the idea that different mechanisms drive the integration of stimulus features at least within different audio-visual CMCs (Evans & Treisman, 2010; Spence, 2011; Spence & Sathian, 2020).

3. General Discussion & Future implications

The aim of this cumulative dissertation was to examine the role of certain contextual aspects on cognitive processes as potential influencing factors for these processes. Throughout this dissertation the role of context has been elaborated in various cognitive processes: In emotion processing, the transition from implicit to explicit learning and in the processing audio-visual crossmodal correspondences. As the results of each study has already been discussed individually, the focus of this section is on placing of the results in the broad field of context. Furthermore, based on the results presented throughout this dissertation, possible implications for future studies are elaborated.

In the first study, which focused on emotion processing, a potential contribution of internal contextual aspects to the valence processing of latter presented ambiguous facial expressions was investigated. It was observed that differences in oscillatory activity regarding latter choices of valences occur already previous to stimulus presentation. Furthermore, a positive correlation of negativity bias with this difference in pre-stimulus oscillatory activity was found. These results lead to the conclusion that neural activity, i.e. internal contextual aspects, preceding stimulus presentation influence emotion processing. In line with prior research in various fields highlighting the modulating role of pre-stimulus activity on subsequent stimulus processing (Taesler & Rose, 2022; Winterling et al., 2019), our findings emphasize the necessity of considering pre-stimulus neural activity to achieve a more comprehensive understanding of the cognitive processes under investigation.

Considering the findings of our first study, it would be interesting to further investigate which factors, such as expectancy of upcoming stimuli or mood, lead to differences in oscillatory activity of the perceived valences prior to the stimulus. Expectancy effects could be investigated by questionnaires as well as by manipulating the expectation of the upcoming stimulus (Baker et al., 2022; Sussman et al., 2017). This manipulation may be achieved through implementing different sounds instead of the neutral sound used in our study, which are associated with positivity (e.g. cheers) or negativity (e.g. boos or an error sound). However, the possibility that such sounds may influence mood must be carefully considered. According to a study by Manierka and colleagues (2021), mood fluctuations can affect the perception of the valence of facial expressions (Manierka et al., 2021). Therefore, future studies should investigate the role of mood in pre-stimulus oscillatory activity during the processing of facial expressions. In particular, it should be investigated whether mood fluctuations are reflected in pre-stimulus oscillatory activity and whether they show a causal correlation with subsequent valence ratings. We used a rather simple task design in our study. Considering that facial expressions are usually embedded in a situational context, future studies could explore the role of pre-stimulus oscillatory activity under realistic social conditions. Previous research has shown that, particularly for ambiguous facial expressions, the context in which the facial expression is embedded tremendously influences processing of the facial expression. It was observed that a threatening context biased towards fearful and a safe context biased towards positive facial emotion recognition (Bublatzky et al., 2020). It would be of great interest to test the transferability of our results to realistic social conditions. Especially for understanding as well as developing new approaches to treat mental disorders such as depression, in which the negative bias in facial emotion processing plays a major role (Beck, 2008; Joormann & Gotlib, 2006), performing further research on the role of pre-stimulus activity in emotion processing might be beneficial to gain new insight on the underlying mechanisms of these conditions.

The second study focused on the role of environmental contextual aspects on the transition of implicit to explicit memory in a sequential-association task. We found a greater amount of explicit memory achieved in the group that performed the task in an enriched environment compared to the group that performed the task in a conventional setup, which underpins the crucial role of environmental contextual factors on learning mechanisms.

It may be of interest to study encoding and retrieval of learnt knowledge in the same environment in future studies. In our study, the amount of learnt material was comparable between both learning environments. Previous studies showed that learning and retrieving the learnt material in the same environment facilities learning (Godden & Baddeley, 1975; Schwabe & Wolf, 2009). A greater amount of learnt material in VR compared to 2D environments could potentially be found if both encoding and retrieval were performed in VR. Methods such as EEG and fMRI would be of great interest to understand underlying neural processes which are involved in greater explicit than implicit knowledge acquired by enriched environments. However, it remains challenging to use modern VR devices when using common EEG or fMRI setups. The VR goggles are usually heavy and need to be fixated via straps on the participants head (see for example the HTC Vive or the Meta Quest). The pressure on the electrodes beneath the straps likely produce artifacts especially when the participant is moving. Restricting the participant's movements seems to be a neat way to avoid some of the artifacts, but it has the disadvantage that the full potential of VR cannot be used, i.e. the engagement with the VR environment is limited. This limitation of engagement with the environment also applies for the use of VR setups in fMRI. At least some workarounds have already been developed to address some of the mentioned challenges of combining of EEG and VR. These workarounds include for example using fewer electrodes or directly implanting electrodes into the VR device or into headbands (Hongbian, Li et al., 2024). These solutions are, however, not compatible with the normal EEG setups such as the active electrodes (actiCap slim, Brain Products, Gilching, Germany) that were used in our first study. However, in order to gain more insight into the potential benefits of VR or augmented reality on processes such as learning and memory, further studies are needed. Understanding the impact and potential benefits of technologies such as VR or AR on mechanisms such as learning is of great value for various sectors such as the games industry or for the development of new ways of teaching knowledge in schools (Kye et al., 2021).

When investigating the origin of distinct pitch-based CMCs in the third study, the environmental contextual aspect, especially the process of learning that certain stimulus combinations likely belong together, seems to be crucial in the process of acquiring such correspondences. On the other hand, we found that a cultural aspect, specifically language, might not be the main driver for these correspondences.

Future research is essential to further explore the context specificity of the CMC effect as well as the role of the environment in forming such CMCs. Previous studies have led to the suggestion that CMCs may be linked to Bayesian decision theory (Parise, 2016; Tong et al., 2020), wherein prior experiences shape the perception of upcoming stimuli. These studies suggest that CMCs are based on previous experience, which is in line with the results of our study. Additionally, research investigating whether CMCs are innate or learned provides strong evidence that they are predominantly shaped by learning (Parise & Spence, 2013; Parise, 2016; Parise et al., 2014; Spence & Deroy, 2012). Interestingly, in contrast to synesthesia, a phenomenon that is characterized by the occurrence of an experience in one sensory domain, which is accompanied by an involuntary experience within another sensory domain, e.g. specific letters elicit certain colors (Kadosh & Henik, 2006; Ramachandran & Hubbard, 2001), the pairing of the stimulus features in CMCs is not absolute (Chiou & Rich, 2012; Spence & Sathian, 2020). When two differently pitched tones are presented, the tone with the highest pitch is associated with the elevated object. Similarly, the small object will be matched with the higher tone and the bigger object will be matched with the lower tone as long as there is a relative difference between the tones and sizes. This also means that the assignment of congruent and incongruent, e.g. pitch-size or pitch-elevation pairs, is based solely on the context provided by the experiment (Brunetti et al., 2018; Chiou & Rich, 2012). Future studies should investigate to which extent CMCs may be a phenomenon that is bound to laboratory settings by further investigating origin of CMCs by also taking natural environments into account in studies. It also remains an open question to which extent certain CMCs may also depend on other factors than prior acquired knowledge such as for example task dependent stimulus salience (Chiou & Rich, 2012).

4. Conclusion

In the introduction of this dissertation, fields of research were introduced in which certain aspects of context have been shown to critically influence certain processes. The results presented and discussed in this dissertation emphasize the importance of considering context as a valuable factor influencing cognitive processes. This dissertation extended the existing knowledge by demonstrating that internal contextual aspects prior to the stimulus influence the processing of emotional valence and that environmental contextual factors facilitate the conscious access to implicit acquired knowledge and finally, it contributed to understanding of the underlying neural processes of CMCs.

Context is always present, however, its operationalization remains challenging due to its diffuse nature (Stark et al., 2018). It has been demonstrated that various aspects of context can tremendously influence cognitive processes. Rather than focusing on context a whole, it may be more useful to examine the role of specific aspects of context in cognitive processing. The knowledge that context encompasses a rich set of aspects whereby some of them may influence how the brain processes information more than others, may help to be able to find the missing pieces in understanding certain cognitive processes. Studying the influence of different aspects of context on cognitive processes may also provide insights into the underlying mechanisms of disorders such as depression and potentially contribute to modifying the way knowledge is taught in schools in the future.

5. Abbreviations

ACC	Anterior cingulate cortex
AnG	Angular gyrus
BDI-2	Beck depression inventory 2
C	Congruent
CA	Cornu Ammonis
CMC	Crossmodal correspondence
dB	Decibel
EEG	Electroencephalography
eoi	Effect of interest
EPI	Echo planar imaging
fMRI	functional magnetic resonance imaging
FWE	Family wise error
FWHM	Full widths half maximum
HG	Heschls’ gyrus
I	Incongruent
IPS	Intraparietal sulcus
ITI	Inter trial interval
LOPOCV	Leave-one-participant-out cross-validation
MVPA	Multivariate pattern analysis
ROI	Region of interest
RT	Reaction time
SD	Standard deviation
SEM	Standard error of the mean
SPL	Superior parietal lobule
SPM	Statistical parametric mapping
STAI	State trait anxiety inventory

TFA	Time frequency analysis
VR	Virtual reality

6. Tools & Bibliography

Tools

ChatGPT version 3.5 & 4, OpenAi: chatgpt.com

- Improve sentences and passages for their readability and flow
- Editing passages to fit the intended meaning of the passage
- Translating purposes

DeepL Translate, Deepl SE: https://www.deepl.com/translator

- Translate words, sentences and passages in order to verify their meaning

DeepL Write, Deepl SE: https://www.deepl.com/de/write

- Improve sentences and passages for their readability and flow

Scite_ Assistant: https://scite.ai/assistant

- Fine tuning literature search

Scite_ smart Citations plugin for Mozilla

- Fine tuning literature search

Bibliography

Alexander, B. K., Coambs, R. B., & Hadaway, P. F. (1978). The effect of housing and gender on morphine self-administration in rats. *Psychopharmacology*, *58*(2), 175–179. https://doi.org/10.1007/BF00426903

Anderson, B. A. (2016). The attention habit : how reward learning shapes attentional selection. *Annals of the New York Academy of Sciences*, *1369*, 24–39. https://doi.org/10.1111/nyas.12957

Baker, J., Gamer, M., Rauh, J., & Brassen, S. (2022). Placebo induced expectations of mood enhancement generate a positivity effect in emotional processing. *Scientific Reports*, *12*(1), 1–12. https://doi.org/10.1038/s41598-022-09342-2

Beck, A. T. (2008). The Evolution of the Cognitive Model of Depression and Its Neurobiological Correlates. *American Journal of Psychiatry*, *165*, 969–977. https://doi.org/https://doi.org/10.1176/appi.ajp.2008.08050721

Benedict, R. H. B., & Lockwood, A. (2002). Covert Auditory attention generates activation in the rostral/dorsal ACC. *Journal of Cognitive Neuroscience,* 14(4), 637–645.

Bowling, D. L., Garcia, M., Dunn, J. C., Ruprecht, R., Stewart, A., & Frommolt, K. (2017). Body size and vocalization in primates and carnivores. *Scientific Reports*, *7*, 41070.

Brunetti, R., Indraccolo, A., Del Gatto, C., Spence, C., & Santangelo, V. (2018). Are crossmodal correspondences relative or absolute? Sequential effects on speeded classification. *Attention, Perception, and Psychophysics*, *80*(2), 527–534. https://doi.org/10.3758/s13414-017-1445-z

Bublatzky, F., Kavcıoğlu, F., Guerra, P., Doll, S., & Junghöfer, M. (2020). Contextual information resolves uncertainty about ambiguous facial emotions: Behavioral and magnetoencephalographic correlates. *NeuroImage*, *215*. https://doi.org/10.1016/j.neuroimage.2020.116814

Cabeza, R., Ciaramelli, E., & Moscovitch, M. (2012). Cognitive Contributions of the Ventral Parietal Cortex: An Integrative Theoretical Account. *Trends in Cognitive Sciences*, 16(6), 338–352. https://doi.org/10.1016/j.tics.2012.04.008.Cognitive

Chelazzi, L., Perlato, A., Santandrea, E., & Libera, C. Della. (2013). Rewards teach visual selective attention. *Vision Research*, *85*, 58–72. https://doi.org/10.1016/j.visres.2012.12.005

Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, *41*(3), 339–353. https://doi.org/10.1068/p7161

Clos, M., Sommer, T., Schneider, S. L., & Rose, M. (2018). Enhanced transformation of incidentally learned knowledge into explicit memory by dopaminergic modulation. *PLoS ONE*, *13*(6), 1–15. https://doi.org/10.1371/journal.pone.0199013

Crottaz-Herbette, S; Menon, V. (2006). Where and When the Anterior Cingulate Cortex Modulates Attentional Response : Combined fMRI and ERP Evidence. *Journal of Cognitive Neuroscience*, 18(5), 766–780.

Curzon, P., Rustay, N. R., & Browman, K. E. (2009). Cued and Contextual Fear Conditioning for Rodents. In J. J. Buccafusco (Ed.), Methods of Behavior Analysis in Neuroscience. (2nd ed.). CRC Press/Taylor & Francis.

Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., Gur, R. C., & Mathiak, K. (2008).

Recognition profile of emotions in natural and virtual faces. In *PLoS ONE* (Vol. 3, Issue 11). https://doi.org/10.1371/journal.pone.0003628

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*, 1–12. https://doi.org/10.1167/10.1.6

Fanselow, M. S. (2010). From contextual fear to a dynamic view of memory systems. *Trends in Cognitive Sciences*, *14*(1), 7–15. https://doi.org/10.1016/j.tics.2009.10.008

Fredriksson, I., Adhikary, S., Steensland, P., Vendruscolo, L. F., Bonci, A., Shaham, Y., & Bossert, J. M. (2017). Prior exposure to alcohol has no effect on cocaine self-administration and relapse in rats : Evidence from a rat model that does not support the gateway hypothesis. *Neuropsychoharmacology*, *42*(5), 1001–1011. https://doi.org/10.1038/npp.2016.209

Gage, S. H., & Sumnall, H. R. (2019). Rat Park: How a rat paradise changed the narrative of addiction. *Addiction*, *114*(5), 917–922. https://doi.org/10.1111/add.14481

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *British Journal of Psychology*, *66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Guetta, R., & Loui, P. (2017). When music is salty: The crossmodal associations between sound and taste. *PLoS ONE*, *12*(3). https://doi.org/10.1371/journal.pone.0173366

Harmer, C. J., Phil, D., Sullivan, U. O., Ch, M. B. B., Favaron, E., Massey-chase, R., Ayres, R., Reinecke, A., Psych, D., Goodwin, G. M., Phil, D., Cowen, P. J., & Psych, F. R. C. (2009). Effect of Acute Antidepressant Administration on Negative Affective Bias in Depressed Patients. *American Journal of Psychiatry*, *166*(10), 1178–1184.

Harp, N. R., Freeman, J. B., & Neta, M. (2022). Mindfulness-based stress reduction triggers a long-term shift toward more positive appraisals of emotional ambiguity. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001173

Heilig, M., & Epstein, D. H. (2016). Time to connect: bringing social context into addiction neuroscience Markus. *Nature Reviews Neuroscience*, *17*(9), 592–599. https://doi.org/10.1038/nrn.2016.67

Hölle, D., Meekes, J., & Bleichner, M. G. (2021). Mobile ear-EEG to study auditory attention in everyday life. *Behavior Research Methods*, *53*, 2025–2036. https://doi.org/10.3758/s13428-021-01538-0

Hongbian, Li; Hyonyoung, Shin; Luis, Sentis; Ka-Chun, Siu; José del R., Millán; Nanshu, L. (2024). Combining VR with electroencephalography as a frontier of brain-computer interfaces. *Device*, *2*(6), 100425. https://doi.org/10.1016/j.device.2024.100425

Humphreys, G. F., & Ralph, M. A. L. (2015). Fusion and fission of cognitive functions in the human parietal cortex. *Cerebral Cortex*, *25*, 3547–3560. https://doi.org/10.1093/cercor/bhu198

Jaap, C., Maack, M. C., Taesler, P., Steinicke, F., & Rose, M. (2022). Enriched environments enhance the development of explicit memory in an incidental learning task. *Scientific Reports*, *12*(1), 1–11. https://doi.org/10.1038/s41598-022-23226-5

Jaap, C., & Rose, M. (2024a). Dissociable neuronal mechanism for different crossmodal correspondence effects in humans. *Acta Neurobiologiae Experimentalis*, 136–152. https://doi.org/10.55782/ane-2024-2439

Jaap, C., & Rose, M. (2024b). Relevance of pre-stimulus oscillatory activity for the perceived valence of emotional facial expressions. *Scientific Reports*, *14*(1), 1–17. https://doi.org/10.1038/s41598-024-69433-0

Jin, C. Y., Borst, J. P., & van Vugt, M. K. (2023). Decoding study-independent mind-wandering from EEG using convolutional neural networks. *Journal of Neural Engineering*, *20*(2). https://doi.org/10.1088/1741-2552/acc613

Joormann, J., & Gotlib, I. H. (2006). Is this happiness i see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology*, *115*(4), 705–714. https://doi.org/10.1037/0021-843X.115.4.705

Kadosh, R. C., & Henik, A. (2006). A common representation for semantic and physical properties. *Experimental Psychology*, *53*(2), 87–94. https://doi.org/10.1027/1618-3169.53.2.87

Kappas, A., Krumhuber, E. & Küster, D. (2013). 6 Facial behavior. In J. Hall & M.

Knapp (Ed.), Nonverbal Communication (pp. 131-166). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110238150.131

Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, *21*(4), 373–387. https://doi.org/10.1162/PRES_a_00124

Kim, W. Bin, & Cho, J. (2020). Encoding of contextual fear memory in hippocampal–amygdala circuit. *Nature Communications*, *11*(1382). https://doi.org/10.1038/s41467-020-15121-2

Kirsch, I. (2019). Placebo effect in the treatment of depression and anxiety. *Frontiers in Psychiatry*, *10*, 1–9. https://doi.org/10.3389/fpsyt.2019.00407

Kirsch, I., Moore, T. J., Scoboria, A., & Nicholls, S. S. (2002). The emperor's new drugs: An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. Prevention & Treatment, 5(1), Article 23. https://doi.org/10.1037/1522-3736.5.1.523a

Knierim, J. J., Neunuebel, J. P., & Deshmukh, S. S. (2014). Functional correlates of the lateral and medial entorhinal cortex: Objects, path integration and local - Global reference frames. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1635). https://doi.org/10.1098/rstb.2013.0369

Korzeniowska, A. T., Root-Gutteridge, H., Simner, J., & Reby, D. (2019). Audio–visual crossmodal correspondences in domestic dogs (Canis familiaris). *Biology Letters*, *15*(11), 20190564. https://doi.org/10.1098/rsbl.2019.0564

Kragel, P. A., Knodt, A. R., Hariri, A. R., & LaBar, K. S. (2016). Decoding Spontaneous Emotional States in the Human Brain. In *PLoS Biology*, 14(9): e2000106. https://doi.org/10.1371/journal.pbio.2000106

Kye, B., Han, N., Kim, E., Park, Y., & Jo, S. (2021). Educational applications of metaverse: possibilities and limitations. Journal of Educational Evaluation for Health Professions, 18, 32. https://doi.org/10.3352/

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children’s speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. https://doi.org/10.1121/1.426686

Li, J., Ma, H., Yang, H., Yu, H., & Zhang, N. (2023). Cognitive bias modification for adult’s depression: A systematic review and meta-analysis. *Frontiers in*

Psychology, *13:*968638. https://doi.org/10.3389/fpsyg.2022.968638

Liang, Y., & Liu, B. (2020). Cross-subject commonality of emotion representations in dorsal motion-sensitive areas. *Frontiers in Neuroscience*, *14*:567797. https://doi.org/10.3389/fnins.2020.567797

Lisman, J., Buzsáki, G., Eichenbaum, H., Nadel, L., Ranganath, C., & Redish, A. D. (2017). Viewpoints: how the hippocampus contributes to memory, navigation and cognition. *Nature Neuroscience*, *20*(11), 1434–1447. https://doi.org/10.1038/nn.4661

Liuzzi, A. G., Bruffaerts, R., Peeters, R., Adamczuk, K., Keuleers, E., De Deyne, S., Storms, G., Dupont, P., & Vandenberghe, R. (2017). Cross-modal representation of spoken and written word meaning in left pars triangularis. *NeuroImage*, *150*, 292–307. https://doi.org/10.1016/j.neuroimage.2017.02.032

Lüddecke, R., & Felnhofer, A. (2022). Virtual reality biofeedback in health: A scoping review. *Applied Psychophysiology and Biofeedback*, *47*(1), 1–15. https://doi.org/10.1007/s10484-021-09529-9

Maimon, N. B., Lamy, D., & Eitan, Z. (2021). Space oddity: musical syntax is mapped onto visual space. *Scientific Reports*, *11*(1), 1–17. https://doi.org/10.1038/s41598-021-01393-1

Manierka, M. S., Rezaei, R., Palacios, S., Haigh, S. M., & Hutsler, J. J. (2021). In the mood to be social: Affective state influences facial emotion recognition in healthy adults. *Emotion*, *21*(7), 1576–1581. https://doi.org/10.1037/emo0000999

Maren, S., Phan, K. L., & Liberzon, I. (2013). The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nature Reviews Neuroscience*, *14*(6), 417–428. https://doi.org/10.1038/nrn3492

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

McCormick, K., Lacey, S., Stilla, R., Nygaard, L. C., & Sathian, K. (2018). Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia*, *112*, 19–30. https://doi.org/10.1016/j.neuropsychologia.2018.02.029

Milad, M. R., Pitman, R. K., Ellis, C. B., Gold, A. L., Shin, L. M., Lasko, N. B., Zeidan, M. a, Handwerger, K., Orr, S. P., & Rauch, S. L. (2009). Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biological Psychiatry*, *66*(12), 1075–1082. https://doi.org/10.1016/j.biopsych.2009.06.026

Misselhorn, J., Daume, J., Engel, A. K., & Friese, U. (2016). A matter of attention: Crossmodal congruence enhances and impairs performance in a novel trimodal matching paradigm. *Neuropsychologia*, *88*, 113–122. https://doi.org/10.1016/j.neuropsychologia.2015.07.022

Mobbs, D., Weiskopf, N., Lau, H. C., Featherstone, E., Dolan, R. J., & Frith, C. D. (2006). The Kuleshov Effect: the influence of contextual framing on emotional attributions. *Social Cognitive and Affective Neuroscience*, *1*(2), 95–106. https://doi.org/10.1093/scan/nsl014

Nasir, S., Rosenthal, D., & Moore, T. (2011). The social context of controlled drug use amongst young people in a slum area in Makassar, Indonesia. *International Journal of Drug Policy*, *22*(6), 463–470. https://doi.org/10.1016/j.drugpo.2011.10.006

Neta, M., Berkebile, M. M., & Freeman, J. B. (2021). The dynamic process of ambiguous emotion perception. *Cognition and Emotion*, *35*(4), 722–729. https://doi.org/10.1080/02699931.2020.1862063

Neta, M., & Brock, R. L. (2021). Social connectedness and negative affect uniquely explain individual differences in response to emotional ambiguity. *Scientific Reports*, *11*(3870). https://doi.org/10.1038/s41598-020-80471-2

Neta, M., Tong, T. T., Rosen, M. L., Enersen, A., Kim, M. J., & Dodd, M. D. (2016). All in the first glance: first fixation predicts individual differences in valence bias. *Cognition and Emotion*, *31*(4), 772–780. https://doi.org/10.1080/02699931.2016.1152231

Neta, M., & Whalen, P. J. (2010). The primacy of negative interpretations when resolving the valence of ambiguous facial expressions. *Psychological Science*, *21*(7), 901–907. https://doi.org/10.1177/0956797610373934

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527.

https://doi.org/10.1016/j.tics.2007.09.009

Onoda, K., Okamoto, Y., Shishida, K., Hashizume, A., Ueda, K., Yamashita, H., & Yamawaki, S. (2007). Anticipation of affective images and event-related desynchronization (ERD) of alpha activity: An MEG study. *Brain Research*, *1151*(1), 134–141. https://doi.org/10.1016/j.brainres.2007.03.026

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Parise, C., & Spence, C. (2013). Audiovisual cross-modal correspondences in the general population. In J. Simner & E. M. Hubbard (Eds.), *The Oxford handbook of synesthesia* (pp. 790–815). Oxford University Press

Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. In *Multisensory Research,* 29,7-28. Brill Academic Publishers. https://doi.org/10.1163/22134808-00002502

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, *220*(3–4), 319–333. https://doi.org/10.1007/s00221-012-3140-6

Parise, C. V, Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(16), 6104–6108. https://doi.org/10.1073/pnas.1322705111

Park, G., Vasey, M. W., Kim, G., Hu, D. D., & Thayer, J. F. (2016). Trait anxiety is associated with negative interpretations when resolving valence ambiguity of surprised faces. *Frontiers in Psychology*, *7*, 1–8. https://doi.org/10.3389/fpsyg.2016.01164

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception*, *41*(7), 854–861. https://doi.org/10.1068/p7225

Peelen, M. V., & Downing, P. E. (2023). Testing cognitive theories with multivariate pattern analysis of neuroimaging data. *Nature Human Behaviour*, *7*(9), 1430–1441. https://doi.org/10.1038/s41562-023-01680-z

Penton-Voak, I. S., Munafò, M. R., & Looi, C. Y. (2017). Biased facial-emotion perception in mental health disorders: A possible target for psychological intervention? *Current Directions in Psychological Science*, *26*(3), 294–301. https://doi.org/10.1177/0963721417704405

Petro, N. M., Tong, T. T., Henley, D. J., & Neta, M. (2018). Individual differences in valence bias: FMRI evidence of the initial negativity hypothesis. *Social Cognitive and Affective Neuroscience*, *13*(7), 687–698. https://doi.org/10.1093/scan/nsy049

Pisanski, K., Isenstein, S. G. E., Montano, K. J., O'Connor, J. J. M., & Feinberg, D. R. (2017). Low is large: spatial location and pitch interact in voice-based body size estimation. *Attention, Perception, and Psychophysics*, *79*(4), 1239–1251.

Place, R., Farovik, A., Brockmann, M., & Eichenbaum, H. (2016). Bidirectional prefrontal-hippocampal interactions support context-guided memory. *Nature Neuroscience*, *19*(8), 992–994. https://doi.org/10.1038/nn.4327

Pool, E. R., Tord, D. M., Delplanque, S., Vuilleumier, P., & Sander, D. (2022). Differential contributions of ventral striatum subregions to the motivational and hedonic components of the affective processing of reward. *The Journal of Neuroscience 42*(13), 2716–2728.

Ramachandran, V. S., & Hubbard, E. M. (2001). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1470), 979–983. https://doi.org/10.1098/rspb.2000.1576

Roehe, M. A., Kluger, D. S., Schroeder, S. C. Y., Schliephake, L. M., Boelte, J., Jacobsen, T., & Schubotz, R. I. (2021). Early alpha/beta oscillations reflect the formation of face-related expectations in the brain. *PLoS ONE*, *16*(7 July), 1–18. https://doi.org/10.1371/journal.pone.0255116

Roelofs, A., Van Turennout, M., & Coles, M. G. H. (2006). Anterior cingulate cortex activity can be independent of response conflict in Stroop-like tasks. *PNAS,* 103(37), 13884–13889.

Rose, M., Haider, H., & Büchel, C. (2010). The emergence of explicit memory during learning. Cerebral Cortex, 20(12), 2787–2797. https://doi.org/10.1093/cercor/bhq025

Ru, D., & Frensch, P. A. (2008). How incidental sequence learning creates reportable knowledge: The role of unexpected events. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 34(5), 1011–1026. https://doi.org/10.1037/a0012942

Rudy, J. W. (2009). Context representations, context functions, and the parahippocampal-hippocampal system. *Learning and Memory*, *16*(10), 573–585. https://doi.org/10.1101/lm.1494409

Saarimäki, H., Glerean, E., Smirnov, D., Mynttinen, H., Jääskeläinen, I. P., Sams, M., & Nummenmaa, L. (2022). Classification of emotion categories based on functional connectivity patterns of the human brain. *NeuroImage*, 247, 118800. https://doi.org/10.1016/j.neuroimage.2021.118800

Salvidegoitia, M. P., Jacobson, N., Bauer, A.-K. R., Griffiths, B., Hanslmayr, S., & Debener, S. (2019). Out and about : Subsequent memory effect captured in a natural outdoor environment with smartphone EEG. *Psychophysiology*, *56*(10), e13331. https://doi.org/https://doi.org/10.1111/psyp.13331

Sanz, L. R. D., Vuilleumier, P., & Bourgeois, A. (2018). Cross-modal integration during value-driven attentional capture. *Neuropsychologia, 120*, 105–112. https://doi.org/10.1016/j.neuropsychologia.2018.10.014

Schwabe, L., & Wolf, O. T. (2009). The context counts: Congruent learning and testing environments prevent memory retrieval impairment following stress. *Cognitive, Affective, & Behavioral Neuroscience*, 9(3), 229–236. https://doi.org/10.3758/CABN.9.3.229

Smith, M. A. (2012). Peer influences on drug self-administration: Social facilitation and social inhibition of cocaine intake in male rats. *Psychopharmacology*, *224*(1), 81–90. https://doi.org/10.1007/s00213-012-2737-6

Smith, S. A., & Mulligan, N. W. (2021). Immersion, presence, and episodic memory in virtual reality environments. *Memory, 29*(10), 1–14 https://doi.org/10.1080/09658211.2021.1953535

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203–220. https://doi.org/10.3758/BF03196157

Sommer, T., Rose, M., Weiller, C., & Büchel, C. (2005). Contributions of occipital, parietal and parahippocampal cortex to encoding of object-location associations. *Neuropsychologia, 43*, 732–743.

https://doi.org/10.1016/j.neuropsychologia.2004.08.002

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attention, Perception & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Spence, C. (2019). On the relative nature of (pitch-based) crossmodal correspondences. *Multisensory Research*, *32*(3), 235–265. https://doi.org/10.1163/22134808-20191407

Spence, C., & Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *I-Perception*, 3(5), 316–318. https://doi.org/10.1068/i0526ic

Spence, C., Ngo, M. K., Percival, B., & Smith, B. (2013). Crossmodal correspondences: Assessing shape symbolism for cheese. *Food Quality and Preference*, *28*(1), 206–212. https://doi.org/10.1016/j.foodqual.2012.08.002

Spence, C., & Parise, C. V. (2012). The cognitive neuroscience of crossmodal correspondences. *I-Perception*, *3*(7), 410–412. https://doi.org/10.1068/i0540ic

Spence, C., & Sathian, K. (2020). Audiovisual crossmodal correspondences: behavioral consequences and neural underpinnings. In Sathian, K and Ramachandran, VS (Ed.), *Multisensory Perception: From Laboratory to Clinic* (pp. 239–258). https://doi.org/10.1016/B978-0-12-812492-5.00011-5

Staresina, B. P., Duncan, K. D., & Davachi, L. (2011). Perirhinal and parahippocampal cortices differentially contribute to later recollection of object- and scene-related event details. *Journal of Neuroscience*, *31*(24), 8739–8747. https://doi.org/10.1523/JNEUROSCI.4978-10.2011

Stark, S. M., Reagh, Z. M., Yassa, M. A., & Stark, C. E. L. (2018). What's in a context? Cautions, limitations, and potential paths forward. *Neuroscience Letters*, *680*, 77–87. https://doi.org/10.1016/j.neulet.2017.05.022

Staudigl, T., & Hanslmayr, S. (2013). Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. *Current Biology*, *23*(12), 1101–1106. https://doi.org/10.1016/j.cub.2013.04.074

Strickland, J. C., & Smith, M. A. (2014). The effects of social contact on drug use: Behavioral mechanisms controlling drug intake. *Experimental and Clinical Psychopharmacology*, *22*(1), 23–34. https://doi.org/10.1037/a0034669

Strube, A., Rose, M., Fazeli, S., & Büchel, C. (2021). The temporal and spectral characteristics of expectations and prediction errors in pain and thermoception. *eLife*, *10*, e62809. https://doi.org/10.7554/eLife.62809

Sussman, T. J., Szekely, A., Hajcak, G., & Mohanty, A. (2016). It's all in the anticipation: How perception of threat is enhanced in anxiety. *Emotion*, *16*(3), 320–327. https://doi.org/10.1037/emo0000098

Taesler, P., Jablonowski, J., Fu, Q., & Rose, M. (2019). Modeling implicit learning in a cross-modal audio-visual serial reaction time task. *Cognitive Systems Research*, *54*, 154–164. https://doi.org/10.1016/j.cogsys.2018.10.002

Taesler, P., & Rose, M. (2022). Multivariate prediction of pain perception based on pre - stimulus activity. *Scientific Reports*, 12, 3199. https://doi.org/10.1038/s41598-022-07208-1

Tong, J., Li, L., Bruns, P., & Röder, B. (2020). Crossmodal associations modulate multisensory spatial integration. *Attention, Perception, and Psychophysics*, *82*(7), 3490–3506. https://doi.org/10.3758/s13414-020-02083-2

Venniro, M., Zhang, M., Caprioli, D., Hoots, J. K., Golden, S. A., Heins, C., Morales, M., Epstein, D. H., & Shaham, Y. (2018). Volitional social interaction prevents drug addiction in rat models. *Nature Neuroscience*, *21*(11), 1520–1529. https://doi.org/10.1038/s41593-018-0246-6

Wager, T. D., & Atlas, L. Y. (2015). The neuroscience of placebo effects: Connecting context, learning and health. *Nature Reviews Neuroscience*, *16*(7), 403–418. https://doi.org/10.1038/nrn3976

Wagner, U., Steiner, E., Hartmann, C., & Braun, K. (2021). Crossmodal correspondences between color, smell, and texture: Investigating the sensory attributes of a body lotion. *Marketing, Zeitschrift Fur Forschung Und Praxis*, *42*(2), 19–34. https://doi.org/10.15358/0344-1369-2020-2-19

Wessel, J. R., Haider, H., & Rose, M. (2012). The transition from implicit to explicit representations in incidental learning situations: More evidence from high-frequency EEG coupling. *Experimental Brain Research*, *217*(1), 153–162. https://doi.org/10.1007/s00221-011-2982-7

Willems, R. M., & Peelen, M. V. (2021). How context changes the neural basis of perception and language. *IScience*, *24*(5), 102392. https://doi.org/10.1016/j.isci.2021.102392

Winterling, S. L., Shields, S. M., & Rose, M. (2019). Reduced memory-related ongoing oscillatory activity in healthy older adults. *Neurobiology of Aging*, *79*, 1–10. https://doi.org/10.1016/j.neurobiolaging.2019.03.012

Wittkamp, C. A., Wolf, M.-I., & Rose, M. (2024). The neural dynamics of positive and negative expectations of pain. *eLife, 13, RP97793.* https://doi.org/10.7554/eLife.97793.3

Zeljko, M., Kritikos, A., & Grove, P. M. (2019). Lightness/pitch and elevation/pitch crossmodal correspondences are low-level sensory effects. *Attention, Perception, and Psychophysics*, *81*(5), 1609–1623. https://doi.org/10.3758/s13414-019-01668-w

Zheng, J., Skelin, I., & Lin, J. J. (2022). Neural computations underlying contextual processing in humans. *Cell Reports*, *40*(12), 111395. https://doi.org/10.1016/j.celrep.2022.111395

7. Publications

7.1. Study 1: Processing of emotional facial expressions is affected by preceding oscillatory activity

Jaap C, Rose M. (2024). Relevance of pre-stimulus oscillatory activity for the perceived valence of emotional facial expressions. *Scientific Reports*. 14(1):19263. https://doi.org/10.1038/s41598-024-69433-0

www.nature.com/scientificreports

scientific reports

OPEN

Relevance of pre-stimulus oscillatory activity for the perceived valence of emotional facial expressions

Carina Jaap & Michael Rose ✉

The interpretation of emotional facial expressions is crucial in everyday social interactions, and rapid processing of these expressions is necessary. Although extensive research has shed light on the mechanisms involved in facial expression processing, there is limited research on the potential role of the state of neural activity that directly precedes the occurrence of a face. Here, we investigated the potential modulatory role of pre-stimulus oscillatory activity in emotional facial expression processing. We tested emotional facial processing in two experiments, one utilizing artificial and the other natural facial expressions. The participants had to evaluate the emotional valence of the presented ambiguous facial expressions. In a univariate analysis, differences in the oscillation activity of the later rated valence of the faces were observed in both experiments, and these differences were observed even before the presentation of the facial expressions. Importantly, two different multivariate approaches directly supported the relevance of pre-stimulus oscillatory activity by exclusively using pre-stimulus oscillatory data to predict the perceived valence of the latter rated facial expression across the two experiments within as well as across subjects. The behavioral data shows the often observed negativity bias, i.e. ambiguous faces resulted in the tendency to rate them as negative. This negativity bias was related to neural activity modulations in the pre-stimulus period and also within post-stimulus processing related activity. These findings underscore the significance of pre-stimulus oscillatory activity in facial expression processing, indicating a functional role of ongoing neural states that affects the processing of facial expressions and constitute a basis for the well described negativity bias.

Keywords Emotional facial expression, Pre-stimulus oscillatory activity, Multivariate analysis, Negativity bias

Abbreviations

BDI-2	Beck depression inventory 2
dB	Decibel
EEG	Electroencephalography
fMRI	Functional magnetic resonance imaging
ITI	Inter trial interval
LOPOCV	Leave-one-participant-out cross-validation
LPP	Late positive potential
MVPA	Multivariate pattern analysis
SD	Standard deviation
STAI	State trait anxiety inventory
TFA	Time frequency analysis

Department for Systems Neuroscience, NeuroImage Nord, University Medical Center Hamburg Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany. ✉email: rose@uke.de

Human faces are one of the main visual stimuli we encounter on a daily basis. Alongside information about a person's identity, facial expressions play a crucial role in social interactions as they provide insight about a person's affective state[1–3]. As facial expressions contain such valid information and are essential for social communication, humans become very precise in classifying them[3,4]. Not only are humans precise in recognizing emotional facial expressions, they are also fast in processing them [3,5,6]. Previous studies using EEG have shown that effects of emotional facial expressions can be observed as early as 50–100 ms after stimulus presentation[5]. These early effects were found primarily extending until 200 ms posterior to an emotional facial expression[5–13], and were found for both positive and negative emotional expressions[9,12,14]. Studies using neural imaging have proposed a theory of a close relationship between the amygdala and early facial emotion processing, especially in fear [15,16]. It has been hypothesized that the fast and automatic processing of emotional facial expressions occurs via processing pathways of the amygdala and thalamus [16,17], amongst other regions. In addition to these early effects, late effects have been observed in emotional face processing, typically emerging around 300–400 ms after stimulus presentation[11,18–20]. It is hypothesized that early processing operates through automatic, bottom-up mechanisms driven by visual processing, whereas later processes are likely influenced by top-down mechanisms [11].

However, emotional processing is strongly affected by different stimulus-independent factors like mood [21], and mood dependent disorders like anxiety [22–24],depression[25], social connectedness [26] and stress [27]. The modulatory effect of stimulus-independent factors on the processing of emotional stimuli can be assumed to be reflected in pre-stimulus neural activity, but this pre-stimulus neural activity has not been directly examined for emotional face processing yet. Nevertheless, evidence from studies manipulating the anticipation of certain subsequent emotional[28,29] or facial[30] stimuli suggest that pre-stimulus oscillations differ depending on the anticipated emotional stimulus. Research in other domains has already emphasized the significance of pre-stimulus activity in the processing of forthcoming stimuli. Studies on pain perception, for instance, have revealed that an individual's brain oscillations preceding the experience of a painful stimulus can notably predict their perception of pain intensity[31,32]. Studies addressing memory encoding [33–36] or visual perception [33,37] have also found a modulation of stimulus processing by oscillations preceding the stimulus.

Another important factor to consider in emotion processing is the interpretation bias. While humans demonstrate proficiency in categorizing distinct facial expressions, an interpretation bias may arise under certain conditions. An interpretation bias entails perceiving the presented emotional facial expression as either more positive or more negative than intended [26,38]. To study interpretation bias in face processing, facial stimuli such as morphs between two unambiguous emotions [39] or faces expressing surprise [23,38] are often used. These facial expressions entail a unique challenge since they can evoke both positive and negative interpretations. When presenting an ambiguous expression, the ideal interpretation of the rating of these would be evenly split between positive and negative valence. However, ambiguous expressions can elicit an interpretation bias, particularly when no context is given. This bias tends to be negative and is therefore also referred to as negativity bias [38,40]. A negativity bias has been observed in both clinical [25] as well as healthy populations[26,40,41].

In clinical populations, i.e. individuals with a mood disorder, an attention bias for negative expression as well as a specific sensitivity for sadness was observed when categorizing facial expressions [25,42,43]. However, with interventions such as a cognitive bias modification, the negativity bias can be modulated as some studies show [44,45]. The ability to modify a negativity bias in mood disorders, such as depression, is crucial since a negative information processing bias is believed to play a major role in the cause and treatment of these disorders[46]. Identifying and adjusting the origins of negativity bias would be a significant advancement in conventional therapeutic approaches. This is particularly important given the unwanted side effects associated with conventional therapeutic substances and their high placebo response [47,48].

In healthy participants, a negativity bias is observed in studies in which ambiguous stimuli are presented without a specific context [41]. This negativity bias can be modified, e.g. by practicing mindfulness, which leads to more shifts towards positive interpretations of ambiguity[27]. In addition to mindfulness training, age also appears to have a positive effect on the processing of ambiguous emotions in healthy participants[49–51]. Further, it was shown that the manipulation of expectancy prior to an emotional face can also shift the consecutive emotional processing[52]. These results indicate the importance of stimulus-independent processes, and it can be assumed that a relevant aspect of the processes underlying a negativity bias is reflected in pre-stimulus activity. Therefore, we aimed to examine the modulatory impact of pre-stimulus activity on the processing of subsequent emotional facial expressions.

Using EEG, we address the questions of whether pre-stimulus oscillations are relevant for the perceived valence of subsequent ambiguous facial expressions and whether these pre-stimulus oscillations correlate with the commonly found negativity bias. To test this, all participants were presented with facial stimuli that showed surprised facial expressions and were asked to classify the perceived valence of the faces.

We selected two distinct sets of facial stimuli to replicate the study within participants. Consequently, our study consisted of two experiments, with each participant engaging in both. In Experiment 1, we presented artificial faces and in Experiment 2, we present natural human faces from the Karolinska database [53] (see “Materials and methods” section for more information on the stimuli). Given that both natural and artificial emotional facial expressions are recognized equally well, as observed in a study in which participants rated different emotions in facial expressions displayed on natural as well as artificial faces[54], we did not anticipate substantial differences between the results of the two experiments. The rationale behind employing two distinct sets of stimuli was to gain a more comprehensive understanding of the potential involvement of pre-stimulus oscillations on subsequent emotion processing across different stimulus attributes. Consequently, the objective was not to investigate the differences between the two experiments but rather to assess the probable involvement of pre-stimulus oscillations on emotion processing independent of the nature of the emotional expression that is subsequently presented.

Therefore, to achieve a more general conclusion about the relevance of pre-stimulus neural activity in latter emotion processing, we used two multivariate classification approaches. One approach was performed at subject- and the other at group-level. The latter was performed to gain insight on a probable participant and stimulus overarching pattern in the pre-stimulus period between differently perceived valences of subsequent presented emotional facial expressions. In both approaches the classifier was trained on pre-stimulus oscillatory activity data of Experiment 1 and tested on pre-stimulus oscillatory activity data of Experiment 2. The multivariate approach also reflects the multiple contributors that can be assumed to constitute emotional processes and corresponding modulatory factors like stress [27], anxiety[23] or mood [21]. Therefore, a multivariate approach may provide valuable insights into the nuanced differences of complex emotional processing compared to a univariate approach.

Materials and methods

Participants

Fifty-eight participants took part in this study (37 females, age = 26.6 ± 4.1 years). All our methods were carried out in accordance with the relevant ethical guidelines and regulations and the experimental protocols were approved by the Ethics Committee of the General Medical Council Hamburg (PV7022). The participants were recruited via an online job platform. All participants gave their written informed consent prior to the experiment and were paid an expense allowance of 10 €/h. The inclusion criteria for this study were 18–35 years of age, physically and mentally healthy, normal or corrected to normal vision and normal hearing. Further exclusion criteria contained frequent use of medication except for birth control.

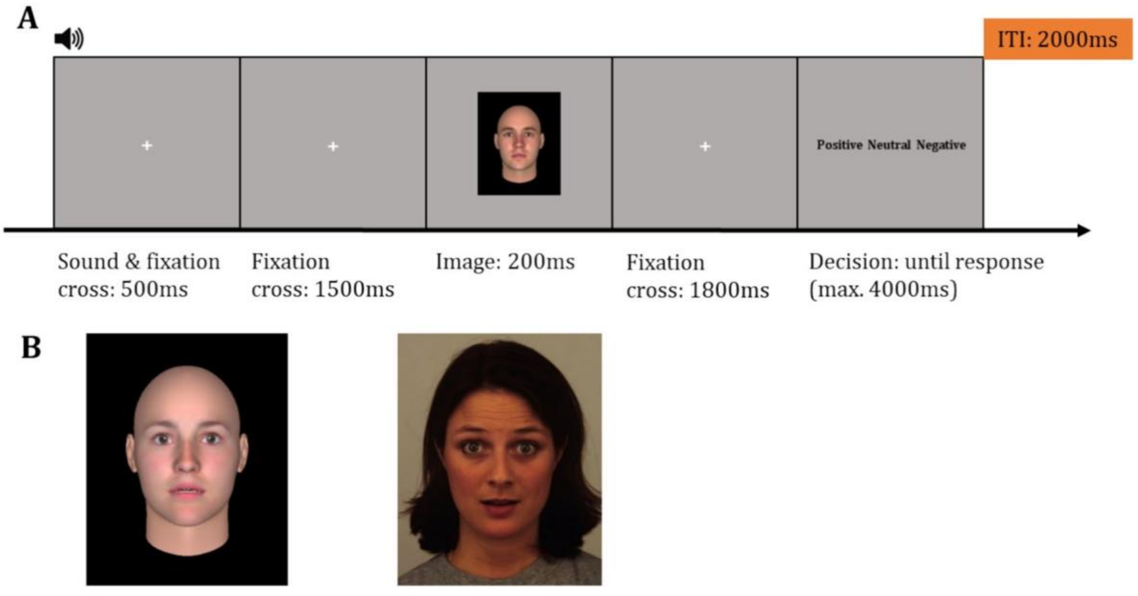
In Experiment 1, 11 participants had to be excluded from the final sample (5 due to a lack of remaining trials after pre-processing, 4 due to technical issues and 1 due to excessive movement), resulting a final sample size of 48 participants (33 females, age = 26.6 ± 4.1 years). In Experiment 2, 19 participants had to be excluded from the final sample (7 due to technical issues, 10 due to lack of remaining trials after preprocessing, and 2 due to excessive movement). This resulted in a final sample size of 39 participants (24 females, age = 26.7 ± 4.2 years).

Procedure

After giving written informed consent, each participant filled out the germane versions of the Becks depression inventory 2 with 21 items[55] (BDI-2) as well as the state trait anxiety inventory with 20 items each[56] (STAI-Y). After filling out the inventories, the participants were seated onto a comfortable chair in a dimly lit, and shielded room. An oral instruction of the task was given by the experimenter, followed by a written instruction on the computer screen. The participants were instructed to decide whether the perceived valence of a presented face was negative or positive. The participants were also informed about a third choice labelled as 'neutral'. However, they were also informed that the facial expressions would never be neutral and that they should ideally choose between a positive or negative valence for the facial expression. If the participants were unsure, they were advised to choose 'neutral' instead of guessing one of the two valence options. To log their answer, the participants had to click on one of three buttons on a mouse. The participants were instructed to choose between a negative or positive valence using the left and right buttons respectively, while the middle mouse button could be used to choose the neutral option. The buttons for negative and positive valences were marked in red and green respectively. The participants were informed that the task was not a speeded decision task, but that they were expected to make a decision within a maximum of 4000 ms. Each trial began with a 2000 ms presentation of a fixation cross. Then, a neutral sound was played for 500 ms. The fixation cross was still present while the sound played and remained on the screen for 1500 ms after the sound had ended. Next, an image of a face was presented for 200 ms (see Fig. 1). Subsequently, a fixation cross was presented for 1800 ms, followed by the response selection. The participants were instructed to select the valence they perceived on the face. The trial ended when the participant clicked a button on the mouse, or after 4000 ms if there was no response. There was an inter-trial interval (ITI) of 2000 ms between trials. The experiment consisted of two parts. In Experiment 1, 14 different artificial faces were presented in 210 trials (for more information see "Stimuli" section). In Experiment 2, 68 different natural human faces were presented. Each face was only presented once, therefore Experiment 2 consisted of 68 trials. We chose not to repeat the natural faces to avoid possible memory effects of the previously rated valence, since the facial features were quite distinct between the natural faces. The order of the two parts remained the same for each participant, as the real human faces were more salient and might influence how the artificial faces would be perceived. There was a short break between the two experiments and breaks were also provided within each experiment.

Stimuli

The visual stimuli in Experiment 1 consisted of 14 artificial faces (7 male, 7 female) created with FaceGen (FaceGen Modeller Core, Version 3.22). The age parameter for all generated faces was set to 20, and the nationality was specified as 'typical European'. To induce a slightly more natural, i.e. imperfect, appearance, an asymmetry value of 0.5 was applied. To create emotional expressions, we used FaceGen's pre-defined discrete emotional expressions, which have intensities ranging from 0 to 100%. Each of our faces was set to 60% surprise in the FaceGen Modeller. This percentage resulted in a balanced distribution of valence ratings in the pilot measurements. The visual stimuli in Experiment 2 consisted of 68 natural faces (34 male, 34 female) from the Karolinska database[57]. All images consisted of frontal faces from the Karolinska Directed Emotional Faces A series showing surprised faces. The images (6° 17′ 0.77″) were presented on a computer screen (23″, refresh rate 60 Hz, ~ 1 m distance from the participant). The images and sounds were presented with Presentation® software (Version 22.01, Neurobehavioral Systems, Inc., Berkeley, CA). The neutral sound was created with Audacity® software (Version 2.4.1.) and consisted of a 500 Hz sine wave as a base, with a fade out towards the end of the 500 ms duration. The sound was presented through two speakers (HD 201, Sennheiser, Germany) one on each side of the screen. The volume of the sound was adjusted to a comfortable level for each participant.



Experiment 1

Experiment 2

Figure 1. (**A**) Timeline of a trial in the experiment with an example of an image used in Experiment 1. The timeline of trials was identical in both experiments. (**B**) Example images used in the experiments. Experiment 1 contained artificial images of faces created with FaceGen (left) and Experiment 2 natural images of faces from the Karolinska data base[57] (https:// kdef. se/; right, shown here as an example is AF01SUS). All images resemble surprised faces.

Analysis of behavioral data

The analysis was performed using R (Version 4.3.1). We computed the percentage of negative ratings in relation to all ratings for each participant individually in each experiment in the final sample. Subsequently, we calculated the means and standard deviations of the percentage of negative ratings relative to all ratings separately for each experiment. For the participants in the final sample whose data sets from both experiments were usable for EEG analysis (N = 36), we performed group-level statistics. We performed a paired t-test on the percentage of negative ratings relative to all ratings across both experiments. We then computed a Pearson correlation coefficient between the percentage of negative ratings (in relation to all ratings) and the BDI-2 as well as the STAI within each experiment.

Analysis of EEG data

Acquisition

The EEG signal was recorded from 64 active electrodes (actiCap slim, Brain Products, Gilching, Germany) placed according to the 10–20 system. The recording took place in an electrically shielded and sound-attenuated room. The ground electrode was placed at the inion (Iz). The online reference was at FCz and was later offline rereferenced to a common average reference. To monitor eye movements, three EOGs were placed on the face. Two were placed on the canthi of the eyes and one was placed under the left eye. Fp1 was used as an additional EOG above the left eye. The electrodes were filled with high-viscosity electrolyte-gel for active electrodes (SuperVisc, EasyCap GmbH). Electrode impedances were kept below 10 kΩ. The EEG signal was amplified with a low cutoff frequency of 0.5 Hz and was recorded with a sampling rate of 500 Hz.

Preprocessing

The EEG signal was processed offline using the FieldTrip toolbox [58] (Version 20220632; https://www.fieldtriptoolbox.org/) implemented in MATLAB (Version 2020a, The MathWorks Inc, Natick, Massachusetts, USA). Epochs from –2000 to 2000 ms relative to the onset of the face stimulus were extracted over the whole dataset. Manual artifact rejection was applied. During manual artifact rejection, we eliminated trials that contained muscle noise, eye blinks in close proximity to the stimulus onset and trials in which the behavioral responses were given too early, i.e. responses given more than 200 ms before the decision window began, which was presented 1800 ms after the presentation of the face ended. Afterwards, an independent component analysis was performed to remove eye movement, blinks, cardiac, and muscle artifacts based on visual inspection of the time course, spectrum, and topography of each component. The re-referencing was then performed followed by another manual artifact detection and rejection followed to remove any remaining artifacts. In rare cases, where a channel was interpolated, the channel was removed prior to the independent component analysis and interpolated after the re-referencing[59]. After the pre-processing was finished, the remaining trials were separated based on the valence response of the participant, i.e. negative and positive ratings of the faces. In Experiment 1, only trials of ambiguous-rated faces, i.e. faces rated as both negative and positive, were considered in the further analysis. In Experiment 1, in which we presented 210 trials in total, a mean of 81 (*SD* = 21) trials remained for the negative and a mean of 68 (*SD* = 28) remained for the positive-rated faces. For the analysis of the time–frequency EEG data, only trial numbers that were equalized between the two perceived valences were used. The trial number of the rated valence with more trials was adjusted to match the trial number of the rated valence with fewer trials by randomly selecting an appropriate number of trials from the larger set of trials (mean number of trials for both valences = 59; *SD* = 20). The inclusion criterion for the analysis of the effective number of trials per rated valence for each participant was 10. In Experiment 2 all natural facial stimuli were presented only once and therefore all remaining trials after the pre-processing were taken into account for the further analysis. In Experiment 2, in which we presented 68 trials in total, a mean of 38 (*SD* = 9) trials remained for the negative and a mean of 19 (*SD* = 6) remained for the positive-rated faces. For the analysis of the time–frequency EEG data, only trial numbers that were equalized between the two perceived valences were used. The trial number of the rated valence with more trials was adjusted to match the trial number of the rated valence with fewer trials by randomly selecting an appropriate subset from the larger set of trials (TFA; mean number of trials in both valences = 19; *SD* = 6). The inclusion criterion for the analysis for the effective number of trials per valence for each participant was 10. We applied the following steps of analysis concerning the time–frequency decomposition consistently across both experiments.

Time–frequency decomposition

We conducted a time–frequency decomposition using the method *mtmconvol* in the MATLAB toolbox FieldTrip[58] (https://www.fieldtriptoolbox.org). The decomposition encompassed frequencies ranging from 2 to 120 Hz with a frequency resolution of 1 Hz for the time-interval of -2000 to 2000 ms. We applied a wavelet convolution with a Hanning taper. A 500 ms sliding window was applied, moving in 50 ms time steps. The temporally and spectrally decomposed data underwent normalization through the application of a baseline correction expressed in decibels (dB). The raw power values were corrected by calculating the ratio between the power values at each time point for each frequency band and the mean power of the baseline period for the same frequency band. The resulting Ratio values were then converted to dB by taking their base 10 logarithm and multiplying by 10 [59]. The baseline period used for this approach spanned from − 1500 to − 500 ms relative to stimulus onset. Baseline correction was performed utilizing the implemented function within FieldTrip [58] toolbox (https://www .fiel dtrip toolbox .or g).

Statistical analysis of EEG data

Group-level analysis of time–frequency decomposed data

Our aim was to test whether pre-stimulus oscillations are related to processing of later presented ambiguous facial expressions by studying differences of the time–frequency spectra in dependence to perceived valence of emotional facial expressions. For this purpose, two experiments with different stimulus material were performed. The statistical analysis of the time–frequency data was performed separately for each experiment. For the grouplevel analysis, paired samples t-tests were performed on the power spectrum of the temporally and spectrally decomposed data for the two emotional valences, i.e. negative and positive, using the implemented Fieldtrip function. The calculated t-values were then corrected using nonparametric cluster-based permutation tests as implemented in the Fieldtrip toolbox[60] (method = montecarlo, cluster threshold = nonparametric-individual, alpha = 0.025, number of randomizations: 2000; frequency: 2 to 60 Hz). We performed the analysis over the time-window of − 500 to 1500 ms relative to stimulus onset.

Correlation between oscillatory data and interpretation bias

We aimed to determine whether negativity bias correlates with the differences of the time–frequency spectra of the two perceived valences of facial expressions, independent of the different facial stimuli used in the experiments. We were interested in this correlation irrespective of the different facial stimuli used in the experiments. Therefore, we combined the time–frequency spectra of both experiments. To test for a relation of the negativity bias, i.e. negative ratings in percent in relation to all ratings, and the difference in oscillatory power in perceived valence (negative–positive rated faces), we used the implemented correlation function in FieldTrip [58] (https:// www.fiel dtrip t oolbox. org). To prepare for the correlation analysis, the difference in oscillatory power between negative and positive rated faces was calculated for each participant for each channel-time–frequency data point in the range of 2–60 Hz and − 500 to 1500 ms relative to stimulus onset. The trials were the same as those used in the main group level analysis, i.e. the number of positive and negative rated trials were equalized within participant. The difference values were then correlated with the percentage of negative ratings (in relation to all ratings) of each participant by computing a Pearson correlation coefficient. The cluster-based permutation correction described in *Group-level analysis of time–frequency decomposed data* was applied to the data to correct for multiple comparisons.

Multivariate pattern analysis (MVPA)

In order to test the differences in the time–frequency spectrum in the pre-stimulus time-window in dependence of the perceived valences across both experiments[61] at subject-level, cross-decoding was performed utilizing the MVPA Light toolbox in Matlab [62]. The trials from each participant, used in the univariate analysis of each experiment, were also employed for the MVPA. This means, that the trial numbers of each of the two perceived valences (positive and negative) were equalized within participants, as in the univariate group-level analysis. However, no baseline correction was applied to the data, since single-trial data are used for the MVPA and all critical information should remain in the data. Only participants for whom we had data from both experiments were included in the cross decoding analysis (N = 36; mean trials Experiment 1 = 62, *SD* = 20; mean trials Experiment 2 = 19, *SD* = 6).

The data from each participant in Experiment 1 were utilized to train a Support Vector Machine (SVM). The trained SVM was then tested on the data of Experiment 2 from the same participant. The time-window of interest ranged from − 1000 to − 200 ms before visual stimulus onset to ensure that the decoding is based on pure pre-stimulus data. This time-window ensured the testing of purely pre-stimulus-driven differences in perceived valence and allowed for the assessment of the stability of the effect over a large window of 800 ms. In contrast to testing the MVPA over the window of − 2000 to − 1000 ms, this window was chosen to test the activity directly preceding and therefore in direct relation to the onset of the stimulus, without any overlap with the post-stimulus window. On the subject level, the high dimensional time–frequency decomposed data were reduced to two dimensions by averaging across the time domain (channels = 60; frequency = 2–60 Hz). Both channel and frequency served as features for the SVM and accuracy was used as metric. The hyperparameter *C* was obtained through grid search implemented in the MVPA Light toolbox. For the statistical analysis, we performed a binomial test on the accuracy results of each participant using the MVPA Light toolbox[62]. Within the binomial test, a binomial distribution is used to calculate the p-value of each accuracy[61]. The output is a mask, in which all accuracies that reach significance are used for further analyses. Subsequently, we identified the highest accuracy value within this mask for each participant and computed the general mean accuracy of these peak accuracies across all participants. If no values remained within the binomial mask, we utilized the maximum accuracy from the original accuracy results. To test whether the accuracies over all participants are different from chance level, the peak accuracies from the binomial analysis were compared to the chance level of 0.5, since accuracies range between 0 and 1, using a one-sample t-test.

In order to test the differences in the time–frequency spectrum in the pre-stimulus time-window in dependence of the perceived valences across both experiments [61] on group level, we performed cross-decoding implementing leave-one-participant-outcross-validation (LOPOCV) using the MVPA Light toolbox in Matlab [62] as a second MVPA approach.

The number of trials, participants, and the pre-stimulus window used for the MVPA at group-level remained the same as for the MVPA performed at subject-level. On the subject level, the high dimensional time–frequency decomposed data were reduced to two dimensions by averaging across the time domain (channels = 60; frequency = 2–60 Hz). The data was then z-transformed. Both channel and frequency served as features for the SVM and accuracy was used as metric. The hyperparameter *C* was obtained through grid search implemented in the MVPA Light toolbox. The SVM was trained with the pooled data of 35 participants. The trained SVM was then tested on the data of Experiment 2 of the left out participant. This procedure was repeated until the data of each participant was once used as test data.

The statistical analysis of the MVPA at subject-level was implemented again on the results of the MVPA performed at group-level.

Results

Behavioral results

During the experiment, the participants saw pictures of different faces (artificial faces in Experiment 1 and natural faces in Experiment 2). The faces expressed an ambiguous emotion and the participants had to rate these emotional expressions as positive or negative. We expected to see a negativity bias in the face ratings and this assumption was confirmed in our behavioral data, i.e. more faces were rated as negative in both experiments and this negativity bias was highly variable across participants.

The mean percentage of negative ratings in percent from all ratings over all participants, i.e. the negativity bias, was lower for Experiment 1 (M = 55%, *SD* = 14%; data range = 32–87%; N = 48; see Fig. 2) compared to Experiment 2 (M = 65%, *SD* = 11%; data range = 38–82%; N = 39; see Fig. 2). An overall wide spread distribution across the participants' ratings in both experiments was observed (see Fig. 2). The group statistic over both experiments showed a significant difference between the negativity bias in the two experiments ($t_{35} = -3.77$; $p < 0.001$).

Since depression as well as anxiety are known to influence emotion processing, each participant filled out both the BDI-2 as well as the STAI-Y questionnaires [23,25]. The mean scores for the questionnaires are displayed in Table 1. The potential range for the BDI-2 scores are 0 to 63, while in our sample, the scores ranged from 0 to 12. As for the STAI-S and STAI-T, their possible ranges are 20–80 respectively. In our sample, the STAI-S scores ranged from 20 to 41, and the STAI-T scores ranged from 20 to 44.

To test if there is an association between the negativity bias and depressive or anxious tendencies, we computed a Pearson correlation coefficient between the negativity bias and the BDI-2/STAI-S/STAI-T scores. We found no significant correlation between the ratings and the BDI-2 or STAI-S/STAI-T scores in both experiments (see Table 2).

Electrophysiology results

Our EEG results show clear differences between subsequently positively and negatively rated ambiguous facial expressions in pre- as well post-stimulus time intervals. These differences in oscillatory power indicated an overall stronger effect for facial expressions subsequently rated as negative. The cluster-statistic of the comparison of the time–frequency spectra of the two perceived valences, resulted in one significant positive cluster in Experiment 1 ($p = 0.0135$) as well as Experiment 2 ($p < 0.001$). In Experiment 1, the cluster comprised all 60 channels and frequency ranges from 2 to 49 Hz. The peak within the pre-stimulus time-window was observed in channel FC4 (4 Hz, $t_{47} = 3.4$; see Fig. 3C), and the peak within the post-stimulus time-window was observed in channel CPz (16 Hz, t_{47} = 5.7). In Experiment 2, the cluster comprised all 60 channels and frequency ranges from 2 to 39 Hz. The peak within the pre-stimulus time-window was observed in channel AF8 (23 Hz, t_{38} = 4.6; see Fig. 4C), and the peak within the post-stimulus time-window was observed in channel F1 (16 Hz, t_{38} = 6.3). An overview of the results of all channels for both Experiment 1 and 2 can be found in Fig. 1 of the supplementary material.

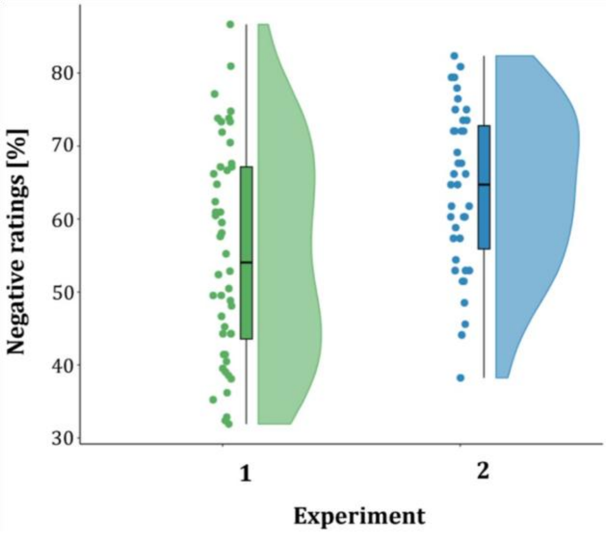


Figure 2. The distribution of negative ratings in percent from all ratings for Experiment 1 (N = 48; data range = 32–87%; left) and Experiment 2 (N = 39; data range = 38–82%; right) faces that expressed ambiguous emotions. Each dot represents the negative rating in percent from all ratings of a participant.

Experiment	**Questionnaire**		
	BDI-2	**STAI-S**	**STAI-T**
1 (artificial)	4.04 ± 3.33	30.6 ± 5.42	30.58 ± 6.71
2 (natural)	4.13 ± 3.29	31.18 ± 5.82	31.56 ± 7.81

Table 1. Mean and standard deviation of the BDI-2 and STAI-S/STAI-T subscales. As the group sizes differed between the experiments, the means and standard deviations were calculated for each experiment separately. The potential range for the BDI-2 scores are l0 to 63, while in our sample, the scores ranged from 0 to 12. As for the STAI-S and STAI-T, their possible ranges are 20–80 respectively. In our sample, the STAI-S scores ranged from 20 to 41, and the STAI-T scores ranged from 20 to 44.

	Questionnaire		
Experiment	**BDI-2**	**STAI-S**	**STAI-T**
1 (artificial)	− 0.175	0.106	0.149
2 (natural)	0.041	− 0.21	− 0.081

Table 2. Results of the computed Pearson correlation coefficient between the BDI-2 as well as STAI-S/ STAI-T subscales with the negative ratings in percent from all ratings for Experiment 1 (artificial faces) and Experiment 2 (natural faces). Displayed are the correlation coefficient values.

Crucially, in both experiments, we observed an effect in the alpha-band (~ 8–12 Hz) within − 500 to 0 ms relative to stimulus onset for facial expressions latter perceived as negative (see Figs. 3A, 4A). The topographical organization of the alpha-band effect shared similarities in both experiments, as illustrated in Figs. 3A and 4A. Within the pre-stimulus period, the effect in the alpha-band was predominantly observed in the left centroparietal regions (see Fig. 3A, exemplary channel CP3 & Fig. 4A, exemplary channel CP5). This observation suggests that the ongoing activity affects the emotional processing of latter occurring ambiguous facial expressions.

In addition to the shared pattern observed in the pre-stimulus period, differential effect-patterns between the experiments were observed in lower and higher frequencies near the alpha-band in the pre-stimulus period. In Experiment 1, the effect in the alpha-band was accompanied by effects in the theta-band in the centro-parietal and frontal regions (3–7 Hz; see Fig. 3A, exemplary channel CP3 & Fig. 3C, exemplary channel FC4), whereas for natural faces in Experiment 2, the effects were also observed with the beta-band in the centro-parietal and frontal regions (13–17 Hz; see Fig. 4A, exemplary channel CP5).

Furthermore, the presentation of natural faces in Experiment 2 was accompanied by an increase in the beta (~ 22–24 Hz) and gamma (~ 30–35 Hz) frequency bands in the pre-stimulus period, as shown in Fig. 4B, C. The increased beta-band effect was primarily located in central and right frontal regions (see Fig. 4C, exemplary channel AF8), while the increased effect in the gamma-band was mainly observed in the left centro-parietal area (see Fig. 4B, exemplary channel C3). Both effects extended over substantial parts of the pre-stimulus interval of 500 ms (see Fig. 4B, C).

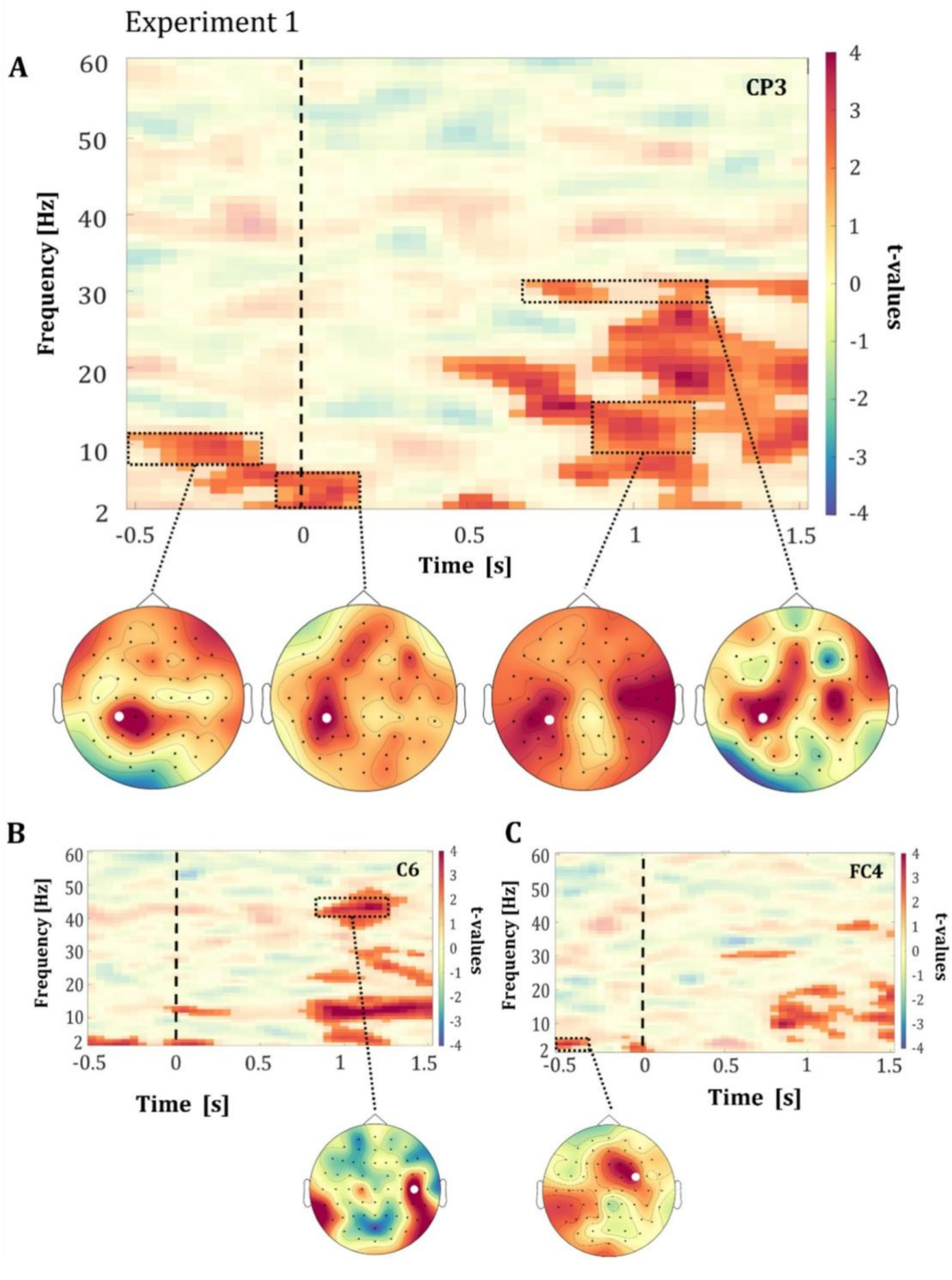


Figure 3. Pre- and post-stimulus results from Experiment 1. Time–frequency representations (TFRs) showing the t-values of difference in power between negatively and positively rated facial expressions across all participants in (**A**) channel CP3, (**B**) channel C6, and (**C**) channel FC4. The facial stimulus was presented for 200 ms with its onset at 0 ms, marked by a dashed line. The motor response was required from 2000 ms after stimulus onset onwards. The highlighted areas represent the significant t-values ($p < 0.05$, corrected) and nonsignificant t-values are transparent. The topographic plots show the spatial distribution of the t-values in the respective time–frequency windows, which are marked with a rectangle. The white dot in the topographic plots marks the position of the channel that is shown in the TFR above.

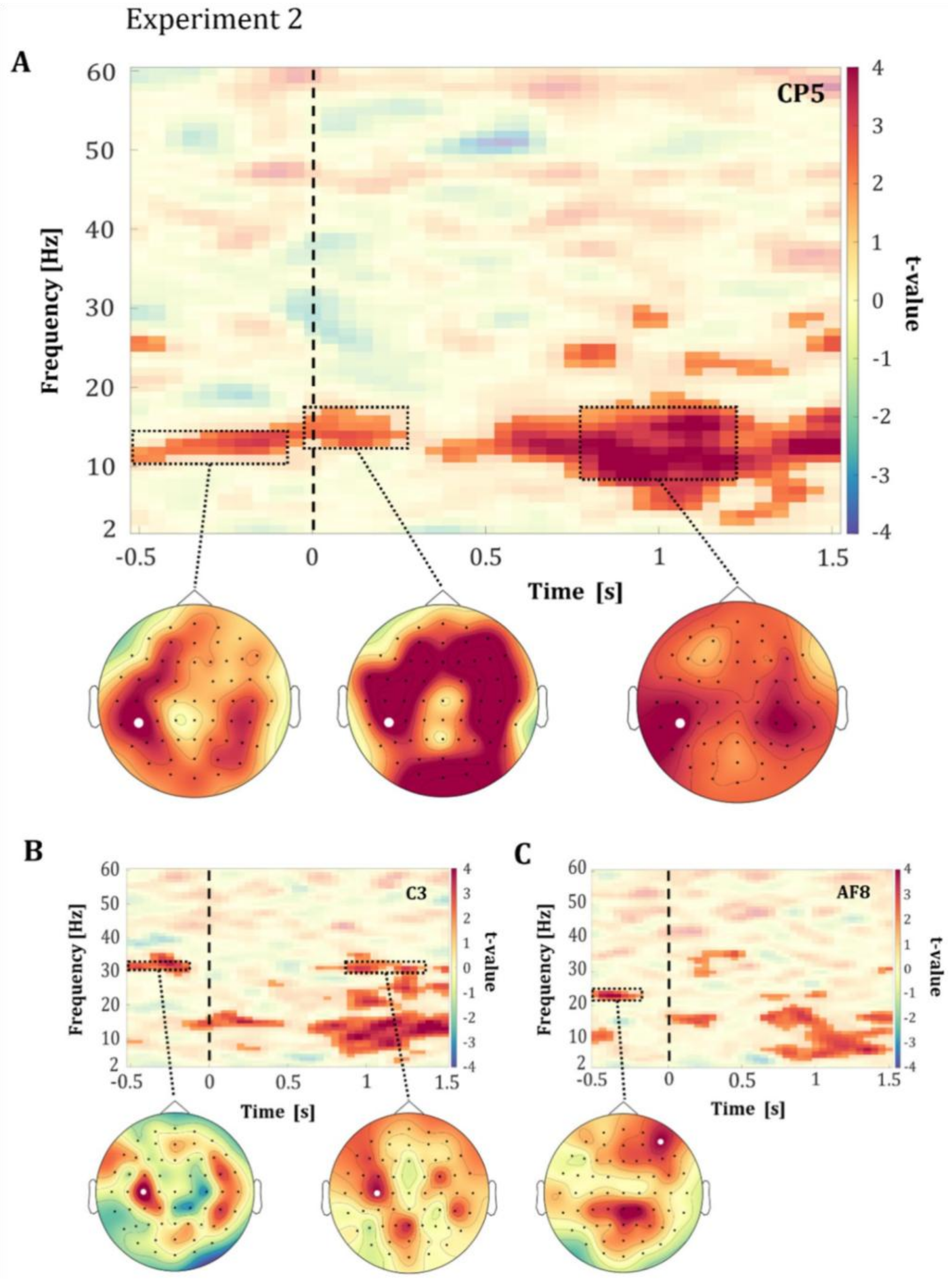


Figure 4. Pre- and post-stimulus results from Experiment 2. TFRs showing the t-values of the difference in power between negatively and positively rated facial expressions across all participants in (**A**) channel CP5, (**B**) channel C3 and (**C**) channel AF8. The facial stimulus was presented for 200 ms with its onset at 0 ms, marked by a dashed line. The motoric response was required from 2000 ms after stimulus onset onwards. The highlighted areas represent the significant t-values ($p < 0.05$, corrected) and non-significant t-values are transparent. The topographic plots show the spatial distribution of the t-values in the respective time–frequency windows, which are marked with a rectangle. The white dot in the topographic plots mark the position of the channel that is shown in the TFR above.

In addition to these pre-stimulus effects, also early stimulus processing revealed differences between positively and negatively rated facial expressions in both experiments. In Experiment 1 the previously described effects in the theta-band extends from the pre-stimulus into an early post-stimulus time-window (~ 3–7 Hz;

see Fig. 3A). In Experiment 2, the described beta-band effect also extends from the pre-stimulus into early stimulus processing (~ 13–17 Hz; see Fig. 4A). Within both experiments the topographic organizations of these effects is focused on the central and centro-parietal regions. However, these effects also spread to the frontal and occipital regions, particularly evident in Experiment 2 (see Fig. 4A, exemplary channel CP5). These early differential effects between the latter rated valences, observed in both experiments, indicate a connection between pre-stimulus and early stimulus processing regarding the processing of ambiguous emotional facial expressions.

In both experiments, large post-stimulus effects were observed starting from 500 ms relative to stimulus onset onwards, which ranged from theta to gamma frequency bands (~ 3–30 Hz; see Fig. 3A, exemplary channel CP3; Fig. 4B, exemplary channel C3). The topographic organization of these effects share similarities in both experiments. Alpha–beta-band effects were observed predominantly in central and centro-parietal regions, extending to frontal regions. Effects in the higher beta-bands (~ 30 Hz) were observed mainly in left and right central and centro-parietal regions, extending to frontal and parietal regions. In Experiment 1, we additionally observed an increase in effect size in the gamma frequency band (~ 37–49 Hz) starting at ~ 800 ms relative to stimulus onset (see Fig. 3B, exemplary channel C6). This effect was mainly observed in the right central, centro-parietal and temporo-parietal regions.

Representation of the negativity bias

As expected, we observed a negativity bias in both experiments with a wide spread distribution of ratings across the participants. We used this information to evaluate the neural representation of the negativity bias across both experiments. Specifically, the relationship between the percentage of total negative ratings per participant and the difference in the differential time–frequency spectrum of negatively and positively perceived facial expressions was computed. The two experiments were combined despite the different strength of negativity bias and the likely differences in the processing of the stimuli. This was done with the intention of investigating a potential general correlation between negativity bias and neural oscillations in the perceived valence of emotional facial expressions, specifically focusing on the pre-stimulus period, irrespective of the stimulus. As for all comparisons, we used an equal number of trials per participant for positive and negative ratings to avoid an influence of differences in trial numbers on the statistical analyses (see "Materials and methods" section for details). The cluster-statistic of the relation between the negative ratings and the time–frequency decomposed data of the differently perceived valences resulted in one significant positive cluster ($p < 0.001$). The cluster comprised all 60 channels and spanned the frequency range from 2 to 60 Hz. The peak in this cluster within the pre-stimulus time-window was observed in channel PO3 (57 Hz; $r = 0.35$), the peak in the cluster within the post-stimulus time-window was observed in channel P8 (11 Hz; $r = 0.51$). An overview of the results of all channels can be found in Fig. 1 (lower plot) of the supplementary material.

We observed clear positive correlations between the negativity bias and the oscillatory power in broad time intervals before and after stimulus presentation over theta to gamma frequency bands.

Within the pre-stimulus period a relation of the negativity bias with the oscillatory activity was observed in the theta-band (~ 3–7 Hz) within parietal and frontal regions (see Fig. 5A, exemplary channel P8), in the beta-band (~ 17–23 Hz) within occipital and temporal regions (see Fig. 5C, exemplary channel O1) and higher gamma-band (~ 50–60 Hz) within centro-parietal, temporo-parietal and frontal regions (see Fig. 5B, exemplary channel CP3). These effects were observed in the 500 ms preceding stimulus onset (see Fig. 5A–C).

In addition to the evident pre-stimulus effects, we also observed a positive association of neural activity with negativity bias within the post-stimulus interval. We observed broad effects starting at ~ 600 ms relative to stimulus onset within the alpha- (8–13 Hz) and beta-band (19–24 Hz) (see Fig. 5A). The post-stimulus patterns illustrating the relationship between the negativity-bias and the oscillatory difference in perceived valence share similarities with the post-stimulus effects observed in the time–frequency decomposed data of the oscillatory differences in perceived valence (see Figs. 3A, 4B). This strongly supports the assumption that processing of emotional faces may be modulated by a negativity bias.

For the alpha-band effect, we observed a broad topographic distribution with prominent extension over the parietal, centro-parietal and frontal regions (see Fig. 5A, exemplary channel P8). For beta-band effects, we observed a broad topographical distribution with prominent extension over the frontal, fronto-central, parietal, and occipital regions (see Fig. 5A, exemplary channel P8).

MVPA results

To evaluate the relevance of pre-stimulus oscillatory activity for the subsequent processing of emotional expressions of differently perceived valences, two MVPA approaches were applied, one at the subject and one at the group-level. Both MVPA analyses utilized a cross-decoding approach.

In the MVPA at the subject level, a SVM was trained on the time–frequency decomposed data of each participant separately in Experiment 1. Subsequently, we tested this trained SVM on the time–frequency decomposed data for each participant separately in Experiment 2. This analysis aimed to ascertain whether similar patterns existed within participants and across both experiments, i.e. irrespective of the implemented stimuli, during the pre-stimulus time-window (−1000 to − 200 ms) without any overlap of the post-stimulus window. The results of our analysis indicated that the SVM trained on the time–frequency patterns preceding the subsequently negatively and positively rated faces from Experiment 1 was able to accurately decode the patterns preceding the subsequently negatively and positively rated faces from Experiment 2. The peak accuracy was above a chance level of 0.5 for each participant (see Fig. 6). The mean peak accuracy across the entire group (N = 36) was 0.739 (*SD* = 0.053; data range = 0.66–0.86). Furthermore, the peak accuracies across all participants were found to be significantly different from chance level ($t_{35} = 26.856$, $p < 0.001$).

To exclude a possible contamination of the pre-stimulus period by oscillatory activity related to the perceived valence of the previous trial, we performed an MVPA. We used a cross-decoding approach to the data from Experiment 1 (N = 36), which had already been used in the MVPAs performed at the subject and group level.

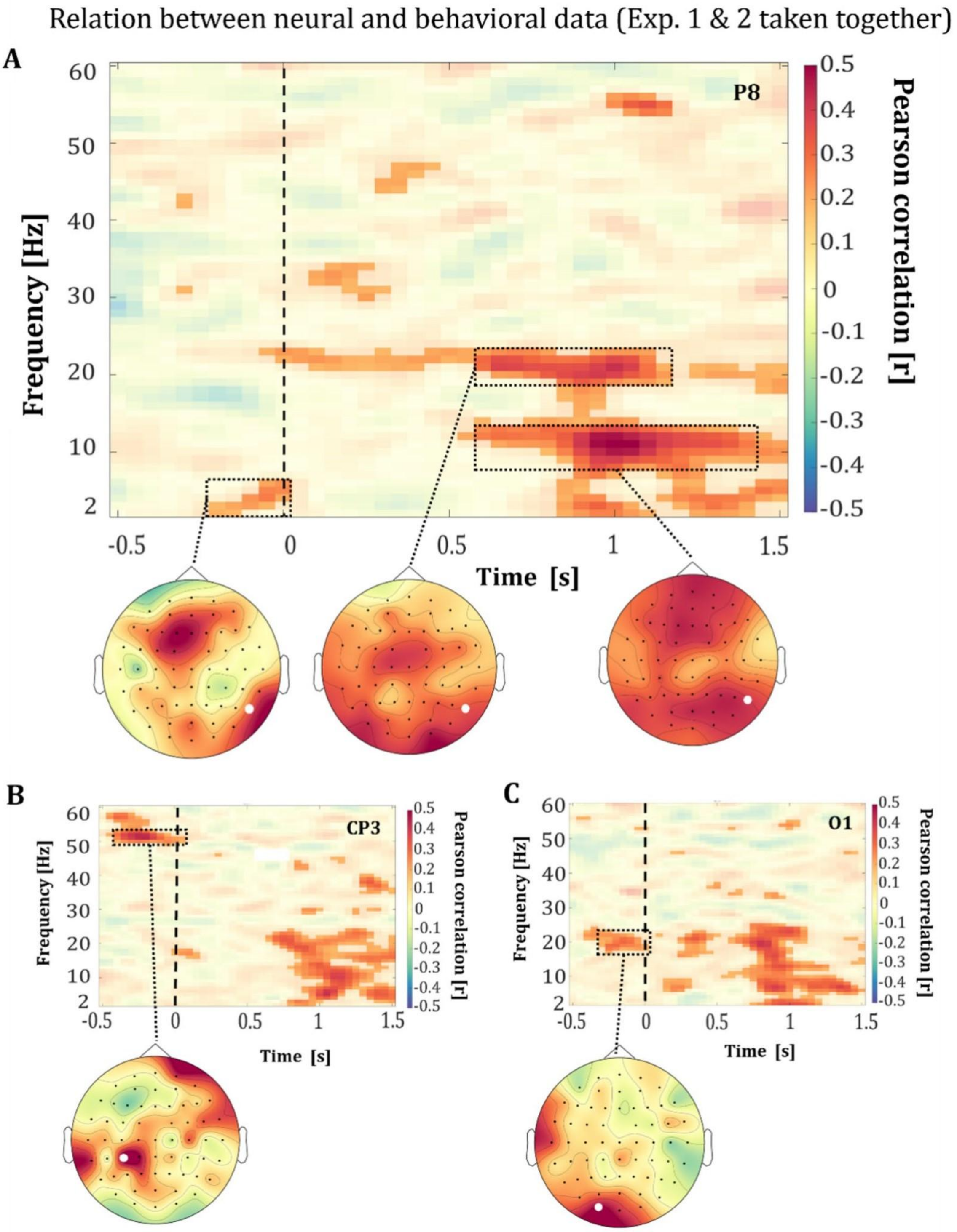


Figure 5. Relation between the neural and behavioral data over both experiments. TFRs demonstrating the representation between the negative ratings of facial expressions in percent in relation to all ratings and the difference (negatively and positively rated facial expressions) in oscillatory power of the face ratings in (**A**) in channel P8, (**B**) channel CP3, and (**C**) channel O1. The facial stimulus was presented for 200 ms with its onset at 0 ms**,** marked by a dashed line. The motor response was required from 2000 ms after stimulus onset onwards. The highlighted areas represent the significant correlation coefficients (p < 0.05, corrected) and non-significant correlation coefficients are transparent. The topographic plots show the spatial distribution of the correlation coefficients in the respective time–frequency windows, which are marked with a rectangle. The white

dot in the topography marks the position of the channel that is shown in the TFR above.

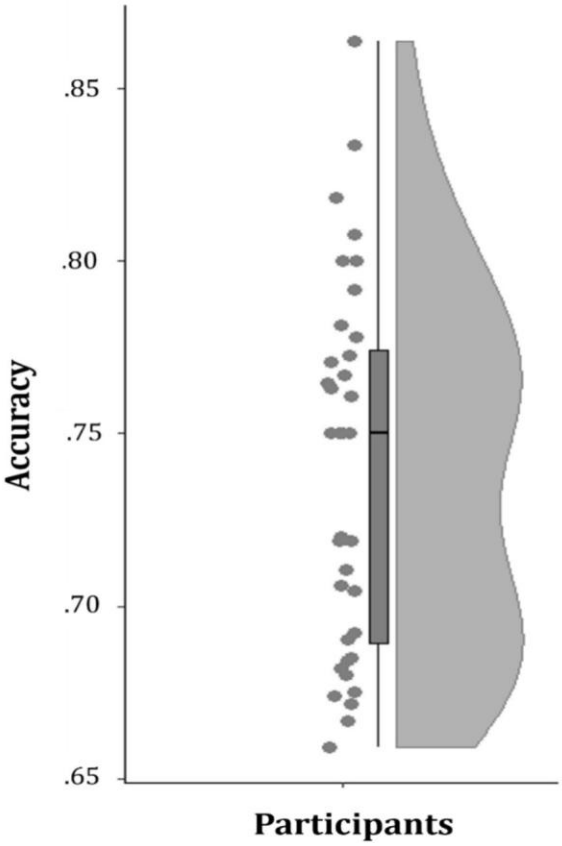


Figure 6. The distribution of peak accuracies per participant in the binomial statistic obtained from the within subjects cross experiment decoding approach. Each dot represents the peak accuracy of a participant (data range = 0.66–0.86). The results indicate that cross-decoding performed well, with accuracies above chance level (0.5) for each participant and across participants (t_{35} = 26.856, p < 0.001).

First, trials with neutral or missed previous ratings were excluded from analyses. Then we used the labels of the rating from the previous trial to train an SVM (metric = accuracy). We then tested the trained SVM using the current rating labels of the trials. This approach allowed us to investigate whether an SVM trained on oscillatory data labeled with the previous trial ratings could successfully decode the pre-stimulus data with the correct labels, thereby assessing the influence of the previous rating on the current rating.

To evaluate the performance of this MVPA, we conducted a second MVPA using the labels of the current ratings for both training and testing an SVM (metric = accuracy; cross-validation = k-folds with 5 folds and 4 repetitions).

Mean peak accuracy across the entire group in the MVPA trained on the labels of the previous ratings performed within participants of Experiment 1 was 0.547 (chance level = 0.5, SD = 0.048). Mean peak accuracy across the entire group in the MVPA trained on the labels of the current ratings performed within participants of Experiment 1 was 0.650 (SD = 0.045). The results of the paired t-test conducted between these MVPAs were found to be statistically significant (t_{35} = 8.131, p <

0.001). We conclude that the activity measured in the prestimulus period is not significantly influenced by the previous stimulus evaluation. This conclusion is supported by the worse performance of the MVPA trained with the labels of the previous ratings compared to the MVPA trained with the labels of the current ratings, as well as the significant difference in performance between these two MVPAs.

In the MVPA performed at the group level, a cross-decoding approach utilizing LOPOCV was applied on exclusively pre-stimulus oscillatory activity. This analysis aimed to ascertain whether similar patterns existed across participants and experiments. The same time interval (− 1000 to − 200 ms) utilized in the MVPA at subjectlevel was employed. Using a cross-decoding approach, we trained a SVM on the time–frequency decomposed data using a LOPOCV approach. This means, that the SVM was trained on the pooled data of Experiment 1 of 35 participants. The trained SVM was then tested on the data set of the remaining participant with data of Experiment 2. This was repeated until the data of each participant was used as test data set. The peak accuracy was above chance level of 0.5 for each iteration. The mean peak accuracy across all iterations was 0.69 (*SD* = 0.056; data range = 0.58–0.82). Furthermore, the peak accuracies across all iterations were found to be significantly different from chance level (t_{35} = 21.4, $p < 0.001$) again indicating a general relevance of pre-stimulus activity for the perceived valence of different face stimuli.

Discussion

The objective of our study was to examine the relevance of pre-stimulus oscillatory activity for the processing of latter presented emotional facial expressions of differently perceived valences. The results showed differences between subsequently perceived emotional valences in the oscillations before the presentation of the facial expression, as well as early and late effects following stimulus onset.

The involvement of pre-stimulus activity in the processing of emotional faces was replicated using different stimulus material**.** The relevance of this activity was demonstrated through two multivariate approaches applied selectively on the pre-stimulus data, with one experiment serving as the training dataset and the other as a test dataset.

Finally, across the participants the observed behavioral negativity bias was also positively related to neural oscillations in the pre-stimulus period, i.e. prior to the appearance of the stimulus, underscoring the significance of ongoing activity for latter emotional processing.

Perceived valence of latter presented emotional expressions is influenced by pre-stimulus oscillations

As hypothesized, we observed late and early post-stimulus effects triggered by emotional facial expressions of latter rated positive and negative valence in both experiments. Such subsequent stimulus effects were also reported by previous EEG studies on the processing of emotional facial expressions [5,11,12,14,18–20]. The disparities in oscillations, which we observed between negative and positive emotional facial expressions, particularly in the later post-stimulus time window within the alpha to beta bands in both experiments, could be attributed to differences in processing the perceived valences. Previous research using images from the International Affective Picture System (IAPS), has shown that stimuli with a more negative valence tend to produce stronger effects in the alpha and beta bands[63,64].

Importantly, our study adds a novel finding: We observed valence-dependent differences in pre-stimulus activity for subsequently presented ambiguous facial expressions. These differences in pre-stimulus activity were observed for both artificial facial expressions in Experiment 1 and natural facial expressions in Experiment 2. Our finding indicates the potential relevance of pre-stimulus neural dynamics in modulating subsequent emotional processing. The modulatory role of oscillations on a subsequent stimulus has been already proven in various studies[31,33–36,65,66]. For instance, in the field of memory, it has been discovered that pre-stimulus oscillations play a role in influencing the successful encoding of subsequent stimuli [34,66,67]. The causal role of pre-stimulus oscillations on subsequent stimulus modulations was further supported by studies successfully using a brain computer interface to enhance memory encoding and visual perception of subsequent stimuli [34,35].

Our study extends this understanding to the domain of emotional processing, demonstrating the influence of pre-stimulus oscillations on the processing of emotional facial expressions. Notably, our results emphasize the role of the alpha band in processing emotional facial expressions of differently perceived valences. We observed significant valence-dependent differences in oscillatory activity preceding the visual stimulus in the alpha band (~ 8–12 Hz) across both experiments. The alpha band was initially believed to primarily play a role in simple visual processing, with alpha synchronization occurring when the eyes are closed and alpha de-synchronization, i.e. suppression, in response to opening the eyes[68]. However, research over the past decades has shown that alpha band activity is involved in several cortical processes such as long-term memory[69], or modulation of visual perception[33,70]. Furthermore, previous research on affective stimuli [64,71] and face[30] processing has elucidated the involvement of alpha oscillations in predictive processing. In our study, however, there was no experimental manipulation of the expectation of the valence of facial expressions. Although not definitive, it can be speculated that the differentiation observed in the alpha band prior to the stimulus was caused by individual, valencedependent fluctuating expectations. On the other hand, the effects observed during the pre-stimulus period, could also originate from fluctuations in emotional states. This theory is supported by a study that examined spontaneous emotional states during resting-state. Kragel et al. conducted an fMRI study in which they successfully decoded distinct emotional states occurring spontaneously during resting-state[72]. We therefore assume that spontaneous fluctuations in emotional states could influence subsequent processing of facial expressions in our study and are related to the effects observed in the oscillatory pre-stimulus activity.

Effects within regions previously identified in studies focusing on post-stimulus emotion processing, including frontal, central, centro-parietal, temporal and occipital regions[8,9,19] were observed in our study in pre- as well as post-stimulus periods. Despite numerous studies on processing emotional facial expressions, it remains speculative whether the effects we observed during the pre-stimulus period originate from the same regions observed in direct stimulus processing. Considering that participants were instructed to provide a motor response 2000 ms after the visual stimulus onset, and trials involving a motor responses prior to

200 ms before the requested response were excluded, it is reasonably to concluded that the observed results are not attributable to motor responses.

In addition to the commonalities observed across experiments, we also found differences in activity during the pre-stimulus interval between experiments. In Experiment 1 alpha band effects accompanied by theta oscillations were observed, while Experiment 2 showed alpha effects accompanied by beta and gamma oscillations. It is unlikely that these differences were caused by differences of the processed emotions, as both stimulus sets comprised surprised faces. The disparities observed in the pre-stimulus oscillations across our experiments could potentially stem from the contrast between the use of artificial faces as stimuli in Experiment 1 and natural faces in Experiment 2. This assumption is supported by findings of an EEG study in which less arousal for artificial faces was observed [73] and an fMRI study in which less activity in the amygdala for artificial faces compared to natural human faces was observed [74]. Furthermore, in Experiment 1, the same 14 faces were presented throughout, while in Experiment 2, each face image was distinct, which could have established differential cognitive settings that may affect pre-stimulus effects. Nevertheless, it is noteworthy that common effects emerged across both experiments during both the pre-stimulus and post-stimulus phases.

To assess both the similarities and differences between the two experiments and to evaluate their statistical relationship, as well as to gage the generalizability of our pre-stimulus effects, we employed two different multivariate approaches. One MVPA was used to test for experiment-independent differences at the subject-level, while a second MVPA was implemented to test for experiment and subject independent differences at the group level. Specifically, this latter approach tested whether there is an overarching pattern within the pre-stimulus window that is consistent across both experiments and subjects, in relation to the valences of the subsequently rated emotional facial expressions. In both approaches we utilized data from one experiment for training and data from the other experiment for testing purposes. With both MVPA approaches accuracies significantly differing from chance level were observed.

The multivariate approach takes advantage of unique individual differences that a univariate analysis cannot identify. Emotion processing is subject to inter-individual differences, as shown by studies on mood-related disorders [25,75,76] and healthy participants [26,40]. In addition, a variety of different processes are involved in the processing of emotions [77,78]. However, both MVPAs, at subject- and at group-level, showed accuracies significantly differing from chance level. Remarkably, we observed reliable classification above chance for all participants over a pre-stimulus time window of 1000 ms, with a mean accuracy of 0.739 within subjects and 0.699 on the group-level. With the use of multivariate approaches we were able to highlight the generalizability of our results across different stimulus sets as well as participants, providing further evidence for the modulatory role of pre-stimulus oscillations on processing of emotional facial expressions. Our findings allow us to assume that overarching patterns exist in oscillations during the pre-stimulus period between differently perceived valences of emotional facial expressions.

Representation of the negativity bias

As hypothesized, we observed a negativity bias in valence ratings across both natural facial expressions in Experiment 1 and artificial facial expressions in Experiment 2. The observed bias in our study is comparable with findings of other studies on surprised faces. We found a mean of 55% and 65% for negative ratings of artificial faces in Experiment 1 and natural faces in Experiment 2 respectively. In their 2018 study on individual differences in the valence bias, Petro and colleagues reported a valence bias of natural surprised faces with a mean of 59.1% negative ratings in 57 participants [40]. In their 2021 study on individual differences in response to emotional ambiguity, Neta and Brock reported 63.59% negative ratings for surprised faces in natural faces [26]. Therefore, we see a typical strength of the negativity bias in healthy participants for surprised facial expressions.

The neural data reflected the observed negativity bias, evident in the positive association of negative valence ratings with the effect in neural activity between the valence ratings. This positive association spanned over both early and late post-stimulus time frames and across all frequency bands from theta to gamma (see Fig. 5A). This finding suggests that the processing of emotional facial expressions is modulated by the negative bias. Supporting this interpretation, an fMRI study revealed a correlation of the negative bias with the neural activity in the left middle frontal cortex, which exhibited a stronger effect with higher overall negative ratings of surprised faces compared to neutral faces [40]. Moreover, we found a positive relationship between the negativity bias and pre-stimulus oscillations, suggesting that stimulus-independent oscillations already initiated biased processing of perceived valence (see Fig. 5A–C). The concept of the negativity bias has spurred various theoretical frameworks [40,79,80]. One prominent theory, the initial negativity hypothesis, posits that emotional facial expressions with an inherent ambiguity like surprise are initially processed as negative, with subsequent cognitive processes potentially shifting valence judgments towards positivity [40]. We propose that stimulus-independent oscillations observed in the pre-stimulus period may further modulate the process described in the initial negativity theory, which suggests that ambiguous emotional expressions are automatically processed as negative, followed by a cognitive switch for positively interpreted ambiguous faces [5,11,40,41].

Our study uncovered differences in valence perception related to negative bias during the pre-stimulus phase among healthy participants, alongside individual variations in biased behavior favoring negativity. This prompted the question of what factors may contribute to such biases in our data. The strong influence of bias in affective disorders, in which a high sensitivity for sadness was observed, highlights the impact of mood on biased interpretations of facial expressions [25,43]. Whereas the efficacy of mood interventions on negative biases in both clinical [42,45] and healthy populations [27] underscores the dynamic nature of affective processing. Interestingly, our findings indicate that the observed positive association of neural activity with negativity bias was mood independent, as evidenced by the non-significant associations of the negativity bias with the STAI-S/-T and BDI-2 scores. This suggests that the effects observed in our study may arise from trial-by-trial fluctuations in oscillatory states, directly impacting biased processing of subsequent ambiguous facial expressions. While mood undoubtedly contributes to bias modulation, our findings suggest that fluctuations between probably affective states may also exert a substantial influence on shaping interpretation biases. However, it remains unclear which factors, beyond mood, may influence these fluctuations. In the context of interpretation bias concerning ambiguous facial expressions in healthy participants, recent studies have shed light on the significant influence of various factors, including social connectedness, stress and emotional regulation abilities on the interpretation bias [26,27]. For instance, it is hypothesized that heightened regulatory capabilities may result in a tendency for more positively biased interpretations of ambiguous facial expressions, while individuals with lower regulatory skills may exhibit the opposite inclination [26,27,40]. Moreover, these regulatory abilities are likely correlated with a higher overall capacity to cope with stress, whereby the latter has also been associated with a greater negativity bias [27]. Even though both anxiety as well as depression tendencies did not correlate with the negativity bias in our study, stress and the ability to regulate ones emotions might have been a contributing factor to the neural effects of biased valence ratings observed in this study.

Further research should examine how pre-stimulus activity influences the subsequent processing of emotional stimuli, with a particular focus on frequencies and how they influence the effect. And also how pre-stimulus activity interacts with the interpretation bias should be studied in order to comprehensively explore the underlying mechanisms.

Conclusion

The processing of emotions is inherently complex. In our study, we identified both post- as well as pre-stimulus variations of neural oscillations between positively and negatively perceived valences of emotional facial expressions. We reliably replicated these findings using two separate sets of stimuli featuring surprised facial expressions. Our study demonstrates that in addition to stimulus-driven mechanisms, neural activity preceding the stimulus likely contributes to how the valence of subsequent emotional facial expressions is processed. This relationship likely influences early post-stimulus processes, but possibly also late post-stimulus processes. This is supported by our replication of the results using two separate stimulus sets in two experiments, one featuring artificial expressions and the other natural expressions. Furthermore, our successful decoding analyses on both experiments at subject- as well as at group-level indicates that the observed pre-stimulus oscillations are associated with perceived emotional valence. Additionally, our findings suggest that pre-stimulus neural dynamics are associated with a biased perception of emotional valence, indicating that biases may arise before stimulus onset and influence subsequent emotion processing. Our observations highlight that pre-stimulus oscillations are relevant for the processing of subsequent emotional facial expressions and that there is a necessity of further investigations on the role of preceding oscillations on the processing of emotions.

Data availability

The datasets generated and/or analyzed during this study are available on request from the corresponding author.

References

1. Kappas, A., Krumhuber, E. & Küster, D. Facial behavior. *Nonverbal Commun.* https://doi.org/10.1007/978-3-319-19650-3_301697 (2013).
2. Todorov, A., Pakrashi, M. & Oosterhof, N. N. Evaluating faces on trustworthiness after minimal time exposure. *Soc. Cogn.* **27**, 813–833 (2009).
3. Schyns, P. G., Petro, L. S. & Smith, M. L. Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: Behavioral and brain evidence. *PLoS ONE* **4**, e5625 (2009).
4. Ekman, P. Should we call it expression or communication? *Innov. Eur. J. Soc. Sci. Res.* **10**, 333–344 (1997).

5. Rellecke, J., Palazova, M., Sommer, W. & Schacht, A. On the automaticity of emotion processing in words and faces: Event-related brain potentials evidence from a superficial task. *Brain Cogn*. **77**, 23–32 (2011).
6. Schacht, A. & Sommer, W. Emotions in word and face processing: Early and late cortical responses. *Brain Cogn*. **69**, 538–550 (2009).
7. Achaibou, A., Pourtois, G., Schwartz, S. & Vuilleumier, P. Simultaneous recording of EEG and facial muscle reactions during spontaneous emotional mimicry. *Neuropsychologia* **46**, 1104–1113 (2008).
8. Eimer, M. & Holmes, A. An ERP study on the timecourse of emotional face. *Neurosci*. *Lett*. **496**, 147–151 (2011).
9. Léveillé, E. *et al*. Effects of emotion and emotional ambiguity on early visual event-related potentials. *Emotion* **23**, 787–804 (2022).
10. Utama, N. P., Takemoto, A., Nakamura, K. & Koike, Y. Single-trial EEG data to classify type and intensity of facial emotion from P100 and N170. In *Proc. Int. Jt. Conf. Neural Networks* 3156–3163. https:// doi. org/ 10. 1109/ IJCNN. 2009.5178 8 18 (2009).
11. Rellecke, J., Sommer, W. & Schacht, A. Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. *Biol*. *Psychol*. **90**, 23–32 (2012).
12. Zhang, M., Fu, Q., Chen, Y. H. & Fu, X. Emotional context modulates micro-expression processing as reflected in event-related potentials. *PsyCh J*. **7**, 13–24 (2018).
13. Batty, M. & Taylor, M. J. Early processing of the six basic facial emotional expressions. *Cogn*. *Brain Res*. **17**, 613–620 (2003).
14. Balconi, M. & Pozzoli, U. Face-selective processing and the effect of pleasant and unpleasant emotional expressions on ERP correlates. *Int*. *J*. *Psychophysiol*. **49**, 67–74 (2003).
15. Guex, R. *et al*. Temporal dynamics of amygdala response to emotion- and action-relevance. *Sci*. *Rep*. **10**, 1–16 (2020).
16. Phelps, E. A. Emotion and cognition: Insights from studies of the human amygdala. *Annu*. *Rev*. *Psychol*. **57**, 27–53 (2006).
17. Vuilleumier, P. How brains beware: Neural mechanisms of emotional attention. *Trends Cogn*. *Sci*. **9**, 585–594 (2005).
18. Duval, E. R., Moser, J. S., Huppert, J. D. & Simons, R. F. *What's in a Face ? The Late Positive Potential Reflects the Level of Facial*. https:// doi. org/ 10. 1027/ 0269- 8803/ a0000 83 (2013).
19. Kaminska, O. K. *et al*. Ambiguous at the second sight: Mixed facial expressions trigger late electrophysiological responses linked to lower social impressions. *Cogn*. *Affect*. *Behav*. *Neurosci*. **20**, 441–454 (2020).
20. Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A. & Ding, M. Neural substrate of the late positive potential in emotional processing. *J*. *Neurosci*. **32**, 14563–14572 (2012).
21. Manierka, M. S., Rezaei, R., Palacios, S., Haigh, S. M. & Hutsler, J. J. In the mood to be social: Affective state influences facial emotion recognition in healthy adults. *Emotion* **21**, 1576–1581 (2021).
22. Song, S. *et al*. Event-related alpha power in early stage of facial expression processing in social anxiety: Influence of language context. *Psychophysiology* **61**, 1–12. https:// doi. org/ 10. 1111/psyp .1445 5 (2023).
23. Park, G., Vasey, M. W., Kim, G., Hu, D. D. & Thayer, J. F. Trait anxiety is associated with negative interpretations when resolving valence ambiguity of surprised faces. *Front*. *Psychol*. **7**, 1–8 (2016).
24. Sussman, T. J., Szekely, A., Hajcak, G. & Mohanty, A. It's all in the anticipation: How perception of threat is enhanced in anxiety. *Emotion* **16**, 320–327 (2016).
25. Joormann, J. & Gotlib, I. H. Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *J*. *Abnorm*. *Psychol*. **115**, 705–714 (2006).
26. Neta, M. & Brock, R. L. Social connectedness and negative affect uniquely explain individual differences in response to emotional ambiguity. *Sci*. *Rep*. **11**, 1–12 (2021).
27. Harp, N. R., Freeman, J. B. & Neta, M. Mindfulness-based stress reduction triggers a long-term shift toward more positive appraisals of emotional ambiguity. *J*. *Exp*. *Psychol*. *Gen*. https:// doi. org/ 10. 1037/ xge00 01173 (2022).
28. Perri, R. L. *et al*. Getting ready for an emotion: Specific premotor brain activities for self-administered emotional pictures. *Front*. *Behav*. *Neurosci*. **8**, 1–9 (2014).
29. Sussman, T. J., Weinberg, A., Szekely, A., Hajcak, G. & Mohanty, A. Here comes trouble: Prestimulus brain activity predicts enhanced perception of threat. *Cereb*. *Cortex* **27**, 2695–2707 (2017).
30. Roehe, M. A. *et al*. Early alpha/beta oscillations reflect the formation of face-related expectations in the brain. *PLoS ONE* **16**, 1–18 (2021).
31. Taesler, P. & Rose, M. Multivariate prediction of pain perception based on pre-stimulus activity. *Sci*. *Rep*. **12**, 1–9. https://doi .org / 10. 1038/ s41598- 022- 07208-1 (2022).
32. Ploner, M., Lee, M. C., Wiech, K., Bingel, U. & Tracey, I. Prestimulus functional connectivity determines pain perception in humans. *Proc*. *Natl*. *Acad*. *Sci*. *U*.*S*.*A*. **107**, 355–360 (2010).
33. Hanslmayr, S. *et al*. Prestimulus oscillations predict visual perception performance between and within subjects. *Neuroimage* **37**, 1465–1473 (2007).
34. Salari, N., Büchel, C. & Rose, M. Functional dissociation of ongoing oscillatory brain states. *PLoS ONE* **7**, e38090 (2012).
35. Salari, N. & Rose, M. Dissociation of the functional relevance of different pre-stimulus oscillatory activity for memory formation. *Neuroimage* **125**, 1013–1021 (2016).
36. Winterling, S. L., Shields, S. M. & Rose, M. Reduced memory-related ongoing oscillatory activity in healthy older adults. *Neurobiol*. *Aging* **79**, 1–10 (2019).
37. Rassi, E., Wutz, A., Müller-Voggel, N. & Weisz, N. Prestimulus feedback connectivity biases the content of visual experiences. *Proc*. *Natl*. *Acad*. *Sci*. *U*.*S*.*A*. **116**, 16056–16061 (2019).
38. Neta, M. & Whalen, P. J. The primacy of negative interpretations when resolving the valence of ambiguous facial expressions. *Psychol*. *Sci*. **21**, 901–907 (2010).
39. Sun, S., Yu, H., Yu, R. & Wang, S. Functional connectivity between the amygdala and prefrontal cortex underlies processing of emotion ambiguity. *Transl*. *Psychiatry* **13**, 1–14 (2023).
40. Petro, N. M., Tong, T. T., Henley, D. J. & Neta, M. Individual differences in valence bias: FMRI evidence of the initial negativity hypothesis. *Soc*. *Cogn*. *Affect*. *Neurosci*. **13**, 687–698 (2018).
41. Neta, M., Berkebile, M. M. & Freeman, J. B. The dynamic process of ambiguous emotion perception. *Cogn*. *Emot*. **35**, 722–729 (2021).
42. Penton-Voak, I. S., Munafò, M. R. & Looi, C. Y. Biased facial-emotion perception in mental health disorders: A possible target for psychological intervention? *Curr*. *Dir*. *Psychol*. *Sci*. **26**, 294–301 (2017).
43. Li, M. *et al*. The neural correlates of the recognition of emotional intensity deficits in major depression: An ERP study. *Neuropsychiatr*. *Dis*. *Treat*. **19**, 117–131 (2023).
44. Li, J., Ma, H., Yang, H., Yu, H. & Zhang, N. Cognitive bias modification for adult's depression: A systematic review and metaanalysis. *Front*. *Psychol*. **13**, 968638 (2023).
45. Menne-Lothmann, C. *et al*. How to boost positive interpretations? A meta-analysis of the effectiveness of cognitive bias modification for interpretation. *PLoS ONE* **9**, e100925 (2014).
46. Rutter, L. A., Passell, E., Scheuer, L. & Germine, L. Depression severity is associated with impaired facial emotion processing in a large international sample. *J*. *Affect*. *Disord*. **275**, 175–179 (2020).
47. Kirsch, I. Placebo effect in the treatment of depression and anxiety. *Front*. *Psychiatry* **10**, 1–9 (2019).
48. Morton, D. L., El-Deredy, W. & Jones, A. K. P. Placebo analgesia: Cognition or perception. *Handb*. *Exp*. *Pharmacol*. **225**, 71 (2014).

49. Brassen, S., Gamer, M. & Bchel, C. Anterior cingulate activation is related to a positivity bias and emotional stability in successful aging. *Biol. Psychiatry* **70**, 131–137 (2011).
50. Neta, M. & Tong, T. T. Don't like what you see? Give it time: Longer reaction times associated with increased positive affect. *Emotion* **16**, 730–739 (2016).
51. Petro, N. M., Basyouni, R. & Neta, M. Positivity effect in aging: Evidence for the primacy of positive responses to emotional ambiguity. *Neurobiol. Aging* **106**, 232–240 (2021).
52. Baker, J., Gamer, M., Rauh, J. & Brassen, S. Placebo induced expectations of mood enhancement generate a positivity effect in emotional processing. *Sci. Rep.* **12**, 1–12 (2022).
53. Goeleven, E., De Raedt, R., Leyman, L. & Verschuere, B. The Karolinska directed emotional faces: A validation study. *Cogn. Emot.* **22**, 1094–1118 (2008).
54. Dyck, M. *et al.* Recognition profile of emotions in natural and virtual faces. *PLoS ONE* **3**, e3628 (2008).
55. Beck, A. T., Steer, R. A. & Brown, G. Beck depression inventory–II (BDI-II). *APA PsycTests* https:// doi. org/ 10. 1037/ t00742-000 (1996).
56. Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R. & Jacobs, G. A. *State-Trait Anxiety Inventory Self Evaluation Questionnaire, Form Y (STAI)* (British Psychological Society, 2003).
57. Lundqvist, D., Flykt, A. & Öhman, A. The Karolinska directed emotional faces—KDEF (CD ROM). *Stock. Karolinska Institute, Dep. Clin. Neurosci. Psychol. Sect.* (1998).
58. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
59. Cohen, M. X. *Analysing Neural Time Series Data—Theory and Practice* (MIT Press, 2014).
60. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
61. Jin, C. Y., Borst, J. P. & van Vugt, M. K. Predicting task-general mind-wandering with EEG. *Cogn. Affect. Behav. Neurosci.* **19**, 1059–1073 (2019).
62. Treder, M. S. MVPA-light: A classification and regression toolbox for multi-dimensional data. *Front. Neurosci.* **14**, 1–19 (2020).
63. Strube, A., Rose, M., Fazeli, S. & Büchel, C. Alpha-to-beta- and gamma-band activity reflect predictive coding in affective visual processing. *Sci. Rep.* **11**, 1–15 (2021).
64. Onoda, K. *et al.* Anticipation of affective images and event-related desynchronization (ERD) of alpha activity: An MEG study. *Brain Res.* **1151**, 134–141 (2007).
65. Scholz, S., Schneider, S. L. & Rose, M. Differential effects of ongoing EEG beta and theta power on memory formation. *PLoS ONE* **12**, 1–18. https:// doi. org/ 10. 1371/ journ al.pone . 01719 13 (2017).
66. Merkow, M. B., Burke, J. F., Stein, J. M. & Kahana, M. J. Prestimulus theta in the human hippocampus predicts subsequent recognition but not recall. *Hippocampus* **24**, 1562–1569 (2014).
67. Ostrowski, J. & Rose, M. Increases in pre-stimulus theta and alpha oscillations precede successful encoding of crossmodal associations. *Sci. Rep.* **14**, 1–15 (2024).
68. Pfurtscheller, G., Neuper, C. & Mohl, W. Event-related desynchronization (ERD) during visual processing. *Int. J. Psychophysiol.* **16**, 147–153 (1994).
69. Hanslmayr, S., Staudigl, T. & Fellner, M. Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Front. Hum. Neurosci.* **6**, 1–12 (2012).
70. Lange, J., Keil, J., Schnitzler, A., van Dijk, H. & Weisz, N. The role of alpha oscillations for illusory perception. *Behav. Brain Res.* **271**, 294–301 (2014).
71. Strube, A., Rose, M., Fazeli, S. & Büchel, C. The temporal and spectral characteristics of expectations and prediction errors in pain and thermoception. *Elife* **10**, 1–41 (2021).
72. Kragel, P. A., Knodt, A. R., Hariri, A. R. & LaBar, K. S. Decoding spontaneous emotional states in the human brain. *PLoS Biol.* **14**, e2000106 (2016).
73. Mühlberger, A. *et al.* Early cortical processing of natural and artificial emotional faces differs between lower and higher socially anxious persons. *J. Neural Transm.* **116**, 735–746 (2009).
74. Moser, E. *et al.* Amygdala activation at 3T in response to human and avatar facial expressions of emotions. *J. Neurosci. Methods* **161**, 126–133 (2007).
75. Basel, D., Aviram, T. & Lazarov, A. Lack of an attention bias away from relatively negative faces in dysphoria is not related to biased emotion identification. *Behav. Ther.* **53**, 182–195 (2022).
76. Suslow, T., Hußlack, A., Kersting, A. & Bodenschatz, C. M. Attentional biases to emotional information in clinical depression: A systematic and meta-analytic review of eye tracking findings. *J. Affect. Disord.* **274**, 632–642 (2020).
77. Adolphs, R. Cognitive neuroscience: Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* **4**, 165–178 (2003).
78. Vuilleumier, P. & Pourtois, G. Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia* **45**, 174–194 (2007).
79. Rozin, P. & Royzman, E. B. Negativity bias, negativity dominance, and contagion. *Personal. Soc. Psychol. Rev.* **5**, 296–320 (2001).
80. Kaffenberger, T., Brühl, A. B., Baumgartner, T., Jäncke, L. & Herwig, U. Negative bias of processing ambiguously cued emotional stimuli. *Neuroreport* **21**, 601–605 (2010).

Acknowledgements

The German Research Foundation (DFG; SFB/Transregio 169, Project B3 and TRR 289 Treatment Expectation— Project Number 422744262, Project A03) supported the present research. The authors thank Jan Ostrowski for the helpful discussion and comments on the manuscript, as well as Michele Frerichs and Lennart May for their help in data acquisition.

Author contributions

M.R. and C.J. designed the study. C.J. collected the data, C.J. and M.R. analyzed the data and C.J. wrote the manuscript under the supervision of M.R. All authors reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. The present research was supported by the German Research Foundation, DFG; SFB/Transregio 169, B3 (https://www.crossmodal-learning.org/) and TRR 289 Treatment Expectation—Project Number 422744262, Project A03.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-69433-0.
Correspondence and requests for materials should be addressed to M.R.
Reprints and permissions information is available at www.nature.com/reprints.

7.2. Study 2: On the transition from implicit to explicit knowledge in enriched versus non-enriched environments

Jaap, C., Maack, M. C., Taesler, P., Steinicke, F., & Rose, M. (2022). Enriched environments enhance the development of explicit memory in an incidental learning task. *Scientific Reports*, 12(1).

scientific reports
www.nature.com/scientificreports

OPEN Enriched environments enhance the development of explicit memory in an incidental learning task

Carina Jaap[1,3], Marike C. Maack[1,3], Philipp Taesler[1], Frank Steinicke[2] & Michael Rose[1]✉

Learning, rendered in an implicit (unconscious) or explicit (conscious) way, is a crucial part of our daily life. Different factors, like attention or motivation, influence the transformation from implicit to explicit memory. Via virtual reality a lively and engaging surrounding can be created, whereby motivational processes are assumed to be a vital part of the transition from implicit to explicit memory. In the present study, we tested the impact of an enriched virtual reality compared to two conventional, non-enriched 2D-computer-screen based tasks on implicit to explicit memory transformation, using an audio-visual sequential association task. We hypothesized, that the immersive nature of the VR surrounding enhances the transfer from implicit to explicit memory. Notably, the overall amount of learned sequence pairs were not significantly different between experimental groups, but the degree of awareness was affected by the different settings. However, we observed an increased level of explicitly remembered pairs within the VR group compared to two screen-based groups. This finding clearly demonstrates that a near-natural experimental setting affects the transformation process from implicit to explicit memory.

---

[1] NeuroImage Nord, Department for Systems Neuroscience, University Medical Center Hamburg Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany. [2]Human-Computer Interaction, Department of Informatics, University of Hamburg, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany. [3]These authors contributed equally: Carina Jaap and Marike C. Maack. ✉email: rose@uke.uni-hamburg.de

Abbreviations

ITI	Intertrial interval
SEM	Standard error of the mean
VR	Virtual reality

Learning is influenced by multiple factors, like attention, and motivation, and is rendered in an implicit (unconscious) or explicit (conscious) way [1,2]. Implicit memory is usually encoded incidentally so that regularities in the environment are extracted without the actual intention to learn them, and thereby influence our behavior without awareness. Evidence suggests, that implicitly acquired knowledge can become explicit, allowing us to extract and use regularities from the environment without having learned them consciously[3–5]. This memory transformation, from the implicit to the explicit domain, represents a crucial mechanism as learning such regularities and complex rules facilitates the development of higher cognitive functions such as reasoning and language [6]. This way, implicit learning forms the basis to adapt to a complex and changing world and guide the decision making in our daily life. **The fluency hypothesis.** Whether implicit information is transformed to become explicit and thereby conscious, depends on a multitude of factors [6–8]. One prominent theory to explain the transformation from incidental learning to explicit memory is based on the ongoing evaluation of predictions [6,9]. In particular, implicit perceptual associations result in predictions for upcoming events and these predictions are assumed to be used to monitor the outcome of consecutive processing. To study this effect, a sequential task is often implemented [6,10]. During the presentation of a predictable sequence of stimuli, a discrepancy between the predicted and the actual processing speed can be detected by the participants. Hence, this detection process can trigger attention towards the cause of this discrepancy and due to the consecutive search processes, explicit memory is generated[4,11]. This is in line with the Unexpected-Event theory [11], which postulates that any metacognitive judgment, like unexpected fluency or accuracy, can be an unexpected event and trigger attributive processes. It can be assumed that this fluency of processing is enhanced in computer-screen-related tasks because, in more near-natural settings, the evaluation of stimuli and responses are slowed down by several factors. For example, in VR applications, the participants can explore the virtual environment to different as well as variable amounts of time. Additionally, the use of a controller as a response device slows down the response in contrast to classical response devices. These factors directly reduce the fluency of stimulus processing and should therefore affect the emergence of explicit memory.

Following this, our implemented 2D-computer-screen application probably creates more fluency of the responses, related to different factors, like smoother transition of trials compared to trials within the VR task design and therefore enable a rhythmical stimulus processing. However, the fluency in the computer-screen application is favored by the time needed for stimulus evaluation and response times due to different response devices. Following these assumptions, we hypothesized to find an enhanced emergence of explicit memory in a conventional 2D-computer-screen based environment in case that the fluency of stimulus processing is a crucial part of implicit to explicit memory formation. In the following we call this assumption, the *fluency hypothesis*. The contrary hypothesis, the *enriched environmental hypothesis*, can be formulated based on theories regarding learning within a more near-natural environment, assuming that rather realistic stimuli and context directly enhance explicit learning processes[12,13].

The enriched environmental hypothesis. The current cognitive research aspires to apply experimental designs in real-world settings (i.e. audio-visual processing while walking or performing daily activities [14,15]), as it was demonstrated that the context of learning (and retrieval) processes are highly relevant for their outcome[8,12,13]. Despite their advantages, these experimental settings are, however, extremely vulnerable to uncontrollable variables. To control external variables and provide a realistic setting, VR environments have become an attractive option. Previous applications demonstrated that these VR environments enable a sufficient, near-natural experience using interactive elements and multisensory stimulation, resulting in multiple levels of excitement and engagement [16,17], which support an improved multi-sensory integration[18]. Moreover, VR facilitates a more salient processing of stimuli by dynamic engagement of the sensorimotor system, which provokes more naturalistic behavioral and physiological responses than abstract stimuli[19,20]. It was previously shown that learning in VR promotes better performance in an enriched environment[21,22] and the literature demonstrated that these enriched stimuli are responsible for the motivational significance of stimuli [1,23,24]. This motivational effect can also enhance the detection of violated predictions due to an increased level of attention instead of performance fluency. Notably, compared to desktop PC settings, VR Head-Mounted Displays like the HTC Vive induce greater feelings of being present in the VR experimental surrounding, and higher motivation to interact with the environment[16,25]. So far, VR has already been broadly used in the research of explicit episodic and spatial learning (for a review see[22]). Here it has been shown that in contrast to basic computer setups, episodic memory performance is increased in VR settings[26,27], which is most likely caused by the near-natural or more specifically immersive VR environments. Therefore, the benefit of using VR studies for explicit memory supports our assumption that VR can enhance the transformation from implicit to explicit memory**.** It is an open question whether the emergence of explicit memory during incidental learning can also benefit from a more near-natural experimental setting or if this phenomenon is exclusive to task designs, in which the volunteers we instructed to memorize the stimuli. We expect, that the detection of unexpected events within ongoing prediction evaluations, needed for the transfer from implicit to explicit knowledge, benefits from the near-natural experimental application in VR reflected in increased explicit memory performance. This hypothesis, the *enriched environmental hypothesis*, is based on the existing evidence from the implicit learning domain and potential effects within a VR experience (i.e., rising motivational and attentional as well as engaging processes).

The present study. In the present study, we employed a sequential-association task, which has been introduced in previous studies[6,28,29]. The task was used as a between-subjects design contrasting a conventional 2D-computer-based presentation with an enriched near-natural VR application. As the evaluation of stimuli and responses are probably slowed in the near-natural VR task, we tested our design within two different 2D-computer-screen groups, differing only in the number of trials, and one VR group. One 2D-computer-screen group practiced the identical amount of trials as presented in the VR condition, which operated as control group for the duration of VR condition and to exclude an influence of experimental length. Previous studies already demonstrated this smaller amount of trial in a 2D-computer-screen application lead to a partly generation of explicit memory in an incidental learning task[4,6]. Both 2D-computer-screen condition were control conditions for our fluency hypothesis.

The use of stimuli from different modalities (crossmodal) has been shown to be beneficial in learning paradigms because memory formation seems to benefit if encoded elements are derived from separate modalities, facilitating their integration (i.e., visual, and auditory [30–32]). The crossmodal stimuli implemented in the conventional 2D-computer-screen based task were modified from a previous study and consisted of simple tones and plain squares as visual stimuli [6] (see Condition 2 for further details). The sequential regularities consisted of the presentation of alternating visual and auditory stimuli, building a fixed eight-digit sequence in 85% of the trials. Only in 15% of all trials this sequence was violated. The sequential task structure was unknown to the participants, and the content could only be learned incidentally.

Importantly, to assess the degree of explicit memory for the embedded sequences, an identical completion task[33] and a free recall test were conducted at the end of all experimental conditions and always outside the VR. Thus, the results from these tasks can be directly compared between the different learning settings. Both post-tests were combined with a confidence rating [6] to identify participants' explicit knowledge [5,8,34,35].

Condition	N taking part in total	N of discarded datasets	N of datasets in the final analysis
VR	51	2	49
PC-short	22	1	21
PC-long	29	2	27

Table 1. Overview of number of participants taking part in each condition before and after discarding relevant datasets.

Previous applications revealed that participants with explicit memory express their knowledge with high confidence. The correct responses under high confidence are an indicator of explicit memory and, hence, will be used to differentiate them from implicit memory [36,37]. To test for probable differences in memory formation, based on the latter mentioned evaluations of the implemented stimuli in VR and the 2D screen conditions, a stimulus value rating was performed as a last step of the post-experimental assessment.

Hypothesis. We hypothesized that both 2D-computer-screen groups perform similar, as it can be assumed that a *fluency* based unexpected event is detected in both variations of the experiments. Our modified version of the task for the VR contained images of different naturalistic landscapes and complex instrumental sounds. Furthermore, the enriched aspect of the VR was achieved by putting the participants directly into the naturalistic landscapes while performing the sequential task. Furthermore, we hypothesized that an enhanced explicit memory can be explained by two potential mechanisms. On the one hand, if *fluency* is an important factor in the transfer of implicit to explicit memory, we expected to find a boosted performance in explicit memory formation in the 2D-computer-screen application compared to the group performing the task in the VR condition. On the other hand, if the performance of explicit memory is greater within the VR condition, motivational and enriched environmental-based attention processes play a crucial role in naturalistic learning scenarios.

Study aim. The aim of the present study was a direct comparison of both the *fluency* and *enriched environmental hypothesis* regarding the mechanism of incidental perceptual learning processes concerning the transfer from implicit to explicit memory. Therefore, we have contrasted a simple 2D-computer-screen experiment and a complex and enriched VR experiment. The influence of both hypotheses can be compared between the different experimental surroundings by assessing the degree of implicit and explicit memory after learning within two tasks that were identical for all experimental conditions.

Material and methods

We tested a sequential association task within a VR environment as well as in two conventional 2D-computerscreen-based tasks. The conventional 2D-computer-screen based conditions will be called PC-short (260 trials) or condition 2 and PC-long (520 trials) or condition 3 in the following.

Participants. 102 (51 in cond. 1; 22 in cond. 2 and 29 in cond. 3) healthy participants with normal hearing and normal or corrected-to-normal vision took part in this study. Data of five participants had to be discarded as the participants did not complete the tutorial successfully (two in cond. 1, one in cond. 2 and two in cond. 3).
Data of 49 participants were part of the final analysis of the VR condition (29 Females, age M = 27.07 years). In the 2nd condition (PC-short), the datasets of 21 (11 Females, age M = 24.7 years) participants were included in the final analysis. In the 3rd condition (PC-long), the datasets of 27 (21 Females, age M = 27.6 years) participants were included in the final analysis (see Table 1). All experiment protocols were approved by the local Ethics Committee of the General Medical Council Hamburg (PV7022) and our methods were carried out in accordance with ethical guidelines and regulations. Before taking part in the experiments, all participants gave their written informed consent and were paid an expense allowance of 10 €/h.

Condition 1 (VR). *Apparatus.* Inside the VR. The visual stimuli were presented in a virtual surrounding, built with the Unity 3.0 engine, via a head-mounted display i.e., HTC Vive. The responses were tracked with a HTC Vive controller. The acoustic stimuli were presented via headphones. The volume was adjusted by the participants to a comfortable level. The tutorial, as well as the main experiment, took place within the VR environment.

Outside the VR. After the main experiment, a completion task, a free recall, and a stimulus value rating were performed on a computer screen (23″, ~ 1 m distance to the participant) using a standard computer mouse. The tones were presented via two loudspeakers (HD 201, Sennheiser, Germany) one on each side of the screen. The volume of the acoustic stimuli was adjusted by the participants to a comfortable level during the before mentioned post-tests.

Stimuli. The visual stimuli in the VR consisted of four different landscapes (ocean, desert, ice, and forest). The participants were teleported into a landscape and were able to get a 360° view of the landscape. The landscapes were designed with the

Unity engine. So, it was a fully immersive experience and not just a 360° image of the landscapes. The acoustic stimuli consisted of four complex instrumental tones, with a duration of 1000 ms each.

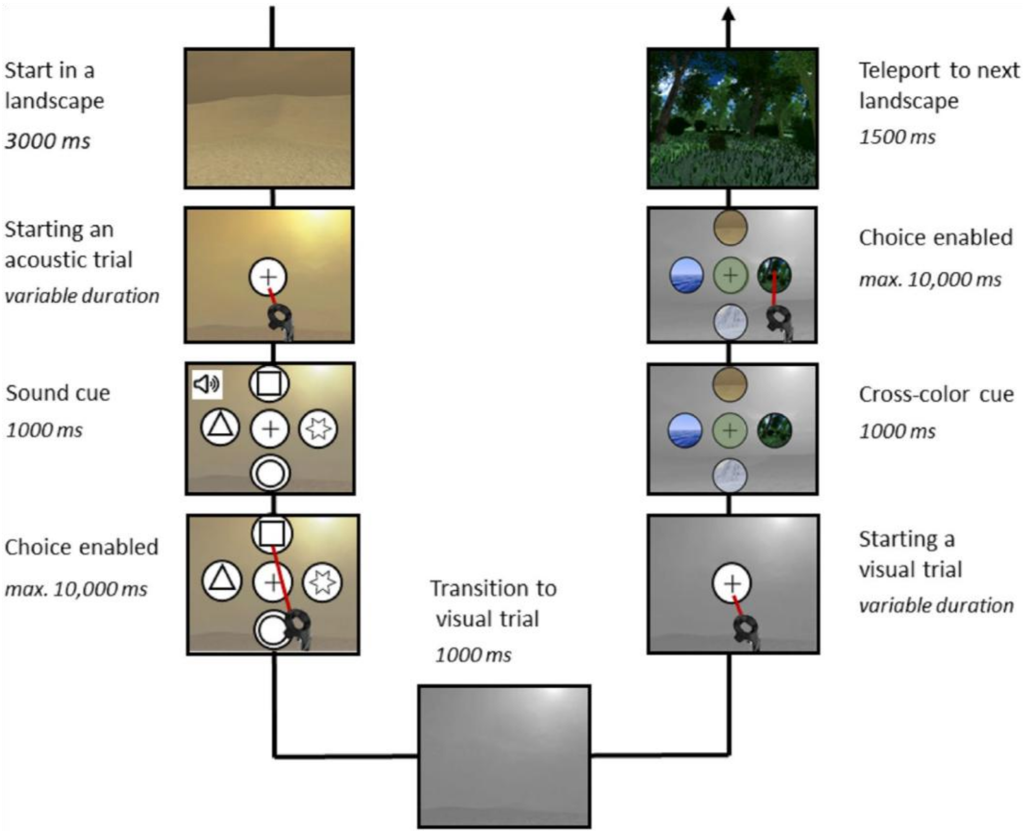


Figure 1. Timeline of a trial within the VR experiment (from upper left to the upper right). The participants were teleported to a virtual landscape where they had 3000 ms time for exploration. Afterwards, a white circle with a cross in the middle appeared in front of them, symbolizing the option to start a trial. The participants could autonomously start by pointing at the circle with the beam of their controller. Next, a sound was played for 1000 ms and four symbols occurred. The sound-symbol combinations were previously learned in the tutorial. Importantly, during the sound was played, participants could not respond. After the sound had ended, participants could choose one of the four symbols within 10,000 ms. After selecting the corresponding symbol, there was a transition from the acoustic to visual trial within 1000 ms during which the color of the current surrounding was desaturated. Afterwards, the visual trial started in the same way as the acoustic trial. Again, the participants could autonomously start the visual trial by pointing at the circle with the beam of their controller. Upon start, four visual stimuli symbolizing four landscapes appeared. Then the target circle in the middle changed from white to the dominant color of one of the four predefined landscapes within 1000 ms. During the color changing process, the participants could not give an answer. Next, participants had to select the icon representing the color of the target (i.e., if the target turned green, participants were expected to select the forest icon within 10,000 ms). After selecting a landscape icon, the participants were teleported to the next correct landscape, independent of their choice (i.e., if the participant chose the ocean landscape although the target was green, still the forest landscape was presented).

Each sound was paired with a symbol (e.g. the piano sound was resembled as a star). By matching each sound with a specific symbol, the participants could match the played sound with the respective symbol within the experiment (see Fig. 1).

Experimental design and procedure. Before entering the main experiment, each participant was introduced to the stimuli and the task instructions for the main experiment during a tutorial. First, the participants had to learn the correct combination of tones and their corresponding symbol in a familiarization task. For the visual trials, the participants learned to match a color with one of the four landscape icons within a familiarization task. Each color was chosen in accordance to the dominant color of the corresponding landscape (desert = yellow). The tutorial ended when less than 2 errors over the last 10 trials were generated by the participants for each trial type. In the main experiment the participants were instructed to answer as fast and precise as possible in each visual and acoustic trial. The VR condition consisted of 260 teleports which is the equivalent of 520 (50% acoustic and 50% visual) trials. To avoid motion sickness, the participants were teleported slowly into the next scenery. The participants were offered several breaks during the VR experiment.
For a detailed overview of the timeline of trials within a teleport, representing a set of an acoustic and visual stimulus presentation in the main experiment, see Fig. 1.

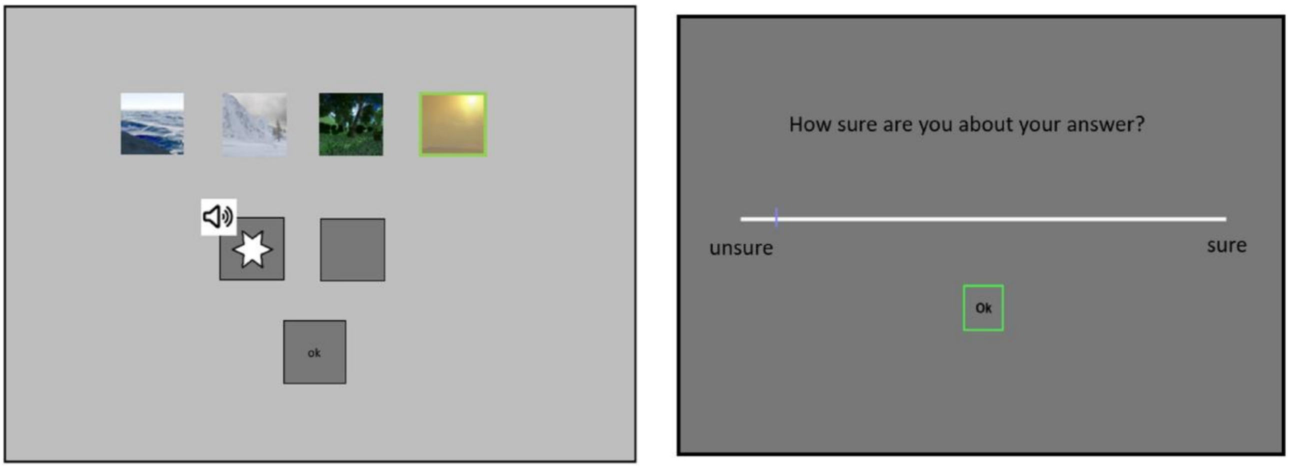


Figure 2. A completion task trial (left) and confidence rating (right) performed outside the VR. In the completion task, four stimuli were displayed above a given stimulus. The participants then had to match the correct visual or acoustic stimulus to the given one. Within the shown crossmodal completion task trial, the participants were asked to match the given acoustic stimulus with the surrounding they associate with it. The participants were asked to make a guess if necessary. After each trial in the completion task (as well as the free recall) the participants had to perform a confidence rating (left).

As we were interested in the learning behavior within the VR, the acoustic and visual trial presentations were part of a sequence consisting of eight digits formed by pairs of eight different stimuli. Within the sequence, each visual stimulus (one of four landscapes) was paired with a specific sound (one of four musical instruments) e.g. visual 1: Desert with acoustic 1: Piano sound. Therefore, pairs or even the whole sequence could be learned in principle. The sequence was interrupted by deviants (15%) with a maximum of three in a row. The sequential regularity of the trial presentations was not introduced to the participants. To avoid motor learning, the order of the visual response option was randomized in each trial [33].

Post-experimental assessment of knowledge and stimulus value rating outside the VR. A completion task (see Fig. 2), a free recall, and a stimulus rating followed the main VR experiment. All post-experimental tasks were retrospectively performed outside the VR on a computer screen. In each trial, one of eight stimuli was given and had to be completed with a stimulus that matched the given stimulus (see Fig. 2). The trials were either crossmodal, with a given stimulus in one modality and four choices given from the opposite modality, or unimodal in which the given stimulus and the choices of answers were of the same modality. The completion task consisted of 64 trials with 50% crossmodal trials (25% visual and 25% acoustic matching). After each trial, the participants had to choose if they were sure or unsure about the given answer (see Fig. 2). With this rating, we later could separate the given answers into implicit (correct answer rated as unsure) and explicit (correct answer rated as sure) knowledge about the presented sequence.

The completion task was followed by a free recall. At this point, the participants were told that there was an order in which the stimuli were presented most of the time during the experiment. Within the free recall, the participants were asked to choose an order for the eight given stimuli. The chosen order should resemble a sequence the participants most likely were presented within the main experiment. The order for the chosen stimuli was not limited. After the participants logged their choice in, by clicking on an "ok" button, the participants were asked if they were sure or unsure about their chosen order.

At the end of the latter mentioned memory assessment tasks, the participants performed a stimulus value rating of the visual stimuli. We implemented a stimulus value rating to test for potential motivational significance of stimuli between the stimuli used in the visually enriched VR and the non-enriched screen based experiments, as this contextual influence can probably correlate with learning effects due to motivational and attentional factors[23]. Each visual stimulus was presented once and the participants could rate it on a continuous scale with "I dislike it" (negative rating) on the left, "Neutral" in the middle and "I like it" (positive rating) on the right side of the scale.

Condition 2 (PC-short). *Apparatus.* The visual stimuli were presented on a 23″ screen (SyncMaster P2370; Samsung). The distance between screen and participant was approximately 1 m. The tones were presented via two loudspeakers (HD 201, Sennheiser, Germany) one on each side of the screen. The volume was adjusted by the participants. For recording the answer of the participant, we used a standard keyboard and computer mouse. The psychtoolbox on Matlab was used to present the experiment.

The visual stimuli consisted of four colored and easily distinguishable squares. The color of each square was chosen analogue to a landscape within the VR condition (blue = ocean, green = forest, yellow = desert, white = icy landscape). Four simple tones (sine waves: 120 Hz, 286 Hz, 389 Hz and 527 Hz) with a duration of 1000 ms, were used as auditory stimuli [38]. Black circles of diameters ranging from 20 to 80% of the size of the visual stimulus were displayed as a visual response option for the acoustic stimuli. The circle size represented the frequency height e.g., the biggest circle represented the tone with the lowest frequency.

Experimental design and procedure. The participants were instructed to respond as quickly and correctly as possible to the target stimulus which was presented in the centre of the screen for the visual stimuli or as a tone to which the participants had to match one of four circles. The participants underwent a training before entering the main experiment (see condition 1). The response options were displayed above the target. The last visual target was still present during an acoustic trial to keep it analogue to the setting in the VR experiment in which the participants remained in a landscape during the acoustic trial (see Fig. 1). After the participants gave their response, the trial ended. Answers had to be given within 2500 ms per trial.

Responses had to be made with the index and middle finger of both hands on a regular keyboard. The enabled keys were "y", "x", ", ; " ,and ". : ".

Condition 2 or PC-short consisted of 260 trials. The trials were part of a sequence consisting of eight stimuli, four in each condition starting with a visual stimulus (Sequence: V1 (e.g. blue square) A1 (e.g. 286 Hz), V2A2 V3A3 V4A4; for more details, see Condition 1).

Post-experimental assessment of knowledge and stimulus value rating. A completion task, a free recall and a stimulus rating followed the main experiment. All post-experimental tasks were performed on a computer screen and were identical, with an exception for the used visual and acoustic stimuli, to the tasks performed by the experimental group 1 (see cond. 1, VR).

Condition 3 (PC-long). *Apparatus.* See cond. 2.
Stimuli. See cond. 2.

Experimental design and procedure. The procedure was the same as in condition 2 except that condition 3 consisted of 520 instead of 260 trials. See cond. 2 for further details.

Post-experimental assessment of knowledge and stimulus value rating. See cond. 1 and cond. 2.

Behavioral data analysis. The important parameters, which can be compared between all experimental conditions, are the amount of memory expressed as implicit or explicit memory. These parameters can be taken from the completion task, as well as the free recall. The latter tasks were identical for all three conditions, except for a change in both visual and acoustic stimuli in the VR condition. All correct answers were taken into account for the analysis of task performance in the completion task, as well as the free recall. Within the main experimental conditions, specific acoustic and visual stimuli formed pairs within a sequence in 85% of the trials. We counted an answer as correct if the participant was able to match a given stimulus with a stimulus of the other modality that was either the following or the previous stimulus within the sequence. As the last visual stimulus was present, when the acoustic stimulus was presented and vice versa, we assumed, that not only forward but also backward learning within the sequence was feasible. Therefore, we accepted an answer in both directions. The amount of implicit memory was calculated as the percentage of "unsure", correct, answers from the sum of all possible answers per participant. The amount of explicit memory was calculated as the percentage of "sure", correct, answers from the sum of all possible trials per participant. This method was used for the assessment of the type of acquired knowledge (implicit; explicit) in the different conditions in both the completion task as well as in the free recall.

To test for differences in the amount as well as quality of gained knowledge, we performed an ANOVA with the factors Condition (VR; PC-short; PC-long) and Learning-Type (implicit; explicit) with performance in each learning type as the dependent variable within the completion task as well as the free recall. Furthermore, we tested for probable differences in the quality of gained knowledge between the two conventional 2D-screenbased tasks and performed an ANOVA with the factors PC-Conditions (PC-short; PC-long) and Learning-Type (implicit; explicit) with performance in each learning type as the dependent variable within the completion task as well as the free recall. For the analysis of the stimulus value rating, all given answers were taken into account. We controlled for potential outliers, i.e. participants selecting only "sure" and while showing constant errors, before we went on with the further analysis. For each rating, where a value between 0 (unpleasant) and 1 (pleasant) was possible, the absolute distance to the neutral rating (0.5) was calculated. A mean over these adjusted stimulus ratings of the four visual stimuli was calculated for each participant in each condition and used for the further analysis. To test for a probable impact of the enriched stimuli, used in the VR, on the stimulus value rating, we performed a two-sided two-sample t-test over stimulus ratings in VR versus PC (PC-short & PC-long merged) conditions. Furthermore, we tested for probable correlations of stimulus value ratings on implicitly and explicitly gained knowledge in each group. The statistical analysis was performed in R (4.0.5) and Matlab (2020b).

Results

Completion task. Notably, we found no difference in the amounts of overall learned sequence pairs between conditions (see Table 2), but the degree of developed explicit memory was affected by the different settings in the three conditions.

The interaction effect of Conditions and Learning-Type ($F_{(2188)} = 27.3$, $p < 0.0001$; see Fig. 3) revealed more explicit memory in the VR group as compared to both PC versions. Overall, volunteers acquired more implicit than explicit memory (main effect of Learning-Type: $F_{(1188)} = 23.1$, $p < 0.0001$).

	Conditions					
Type	**VR [%]**	**SEM [%]**	**PC-short [%]**	**SEM [%]**	**PC-long [%]**	**SEM [%]**
Implicit	19.2	3.15	45.8	5.44	52	2.5
Explicit	36.7	4.2	14.4	4.4	4.62	1.5
Total learned	56	3.12	60.3	3.52	56.6	2.4

Table 2. Completion task performance over learning types in each condition. Mean and SEM in percent for the performance between and within conditions (VR = 49, PC-short = 21, PC-long = 27) for Learning-Type (implicit; explicit) and over all learned stimuli.

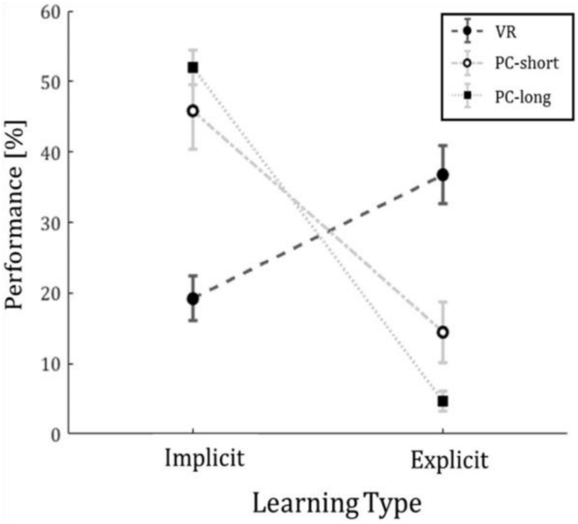


Figure 3. Completion task performance of learned pairs within the given sequence Type (implicit; explicit) and Condition (VR, PC-short, PC-long). The light grey, dotted lines represent the performance of participants of PC-short (N = 21) and PC-long (N = 27) and the dark grey, dashed line represents the performance of VR (N = 49) participants in the completion task. The performance is divided into implicitly learned sequence pairs (left) and explicitly learned sequence pairs (right). The mean of the performance is visualized as black circle (VR), hollow circle (PC-short) and black square (PC-long). The error bars represent the SEM.

Free recall. Notably, we found no difference in the amounts of overall learned sequence pairs between conditions (see Fig. 4) , but the degree of awareness was affected by the different settings between conditions.

In the free recall, the VR experimental group had a mean performance of 37.57% for implicit and 26.03% (SEM = 3.61%) for explicit memory. In contrast, the conventional, 2D screen based experimental groups resulted in an average performance of 46.73% implicit and 7.44% explicit memory (PC-short; SEM = 5.12%) and 65.06% implicit and 1.27% explicit memory (PC-long; SEM = 4.87%).

Overall, the volunteers acquired more implicit than explicit memory (main effect of Learning-Type: $F_{(1,188)} = 55.9$, $p < 0.0001$). The interaction effect of Conditions and Learning-Types ($F_{(2,188)} = 10.9$, $p < 0.0001$; see Fig. 4) revealed more explicit memory in the VR group as compared to both PC groups.

Gained knowledge in the short versus long version of the 2D-computer-screen based conditions. We tested for probable differences in the mean of gained explicit and implicit knowledge between a short and a long version of the 2D-computer-screen based sequential-association-task. Notably, we found no difference in the amounts of overall learned sequence pairs between the 2D-computer-screen based conditions. Overall, volunteers acquired more implicit than explicit memory in both the completion task (main effect of Learning-Type: $F_{(1,92)} = 82.94$, $p < 0.0001$; see Fig. 3) as well as the free recall (main effect of LearningType: $F_{(1,92)} = 265.2$, $p < 0.0001$; see Fig. 4). An interaction effect for the influence of Conditions PC-short versus PC-long on Learning-Type was significant for both completion task ($F_{(1,92)} =$

12.44, $p < 0.001$) and free recall ($F_{(1,92)} = 4.770$, $p < 0.0315$) (see Figs. 3 and 4). However, this effect is related to less explicit memory in the PC versions.

Stimulus value rating. We tested for probable differences in the mean of stimulus value ratings between an enriched visual environment in the VR task and non-enriched visual stimuli in both PC tasks.

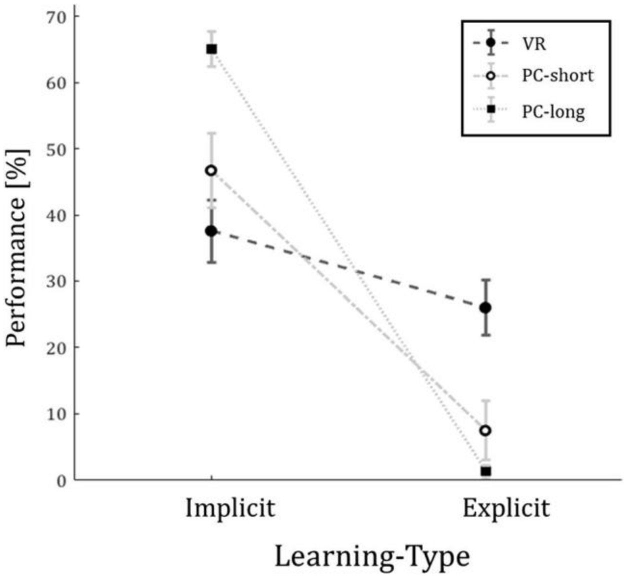


Figure 4. Free recall performance of learned pairs within the given sequence divided by Learning-Type (implicit; explicit) and Condition (VR, PC-short, PC-long). The light grey, dotted lines represent the performance of participants of PC-short (N = 21) and PC-long (N = 27) and the dark grey, dashed line represents the performance of VR (N = 49) participants in the free recall. The performance is divided into implicitly learned sequence pairs (left) and explicitly learned sequence pairs (right). The mean of the performance is visualized as black circle (VR), hollow circle (PC-short) and black square (PC-long). The error bars represent the SEM.

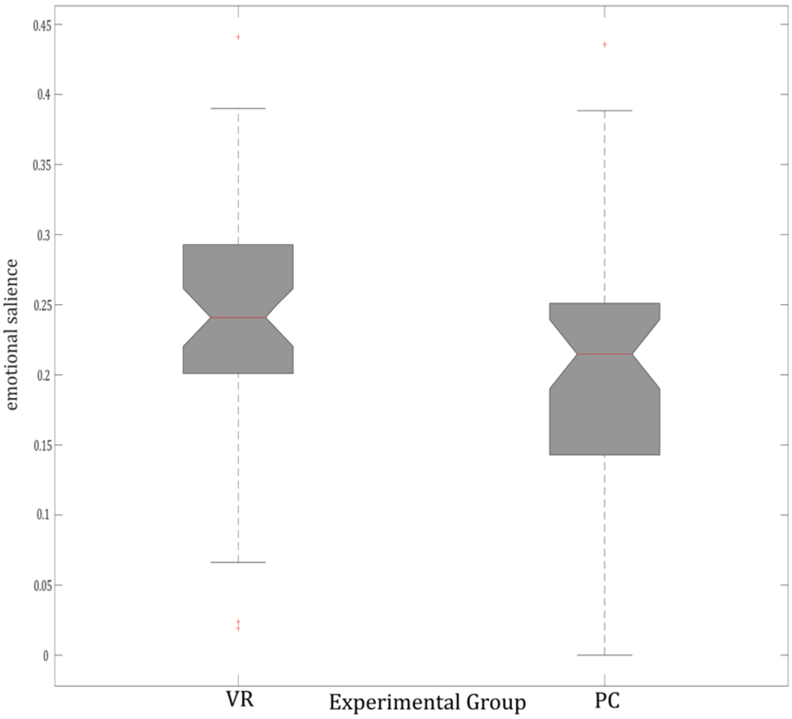


Figure 5. The stimulus value rating for the VR condition and the two merged computer-screen conditions. The Boxplot shows for the VR condition a median of 0.24 (25-percentiles of 0.2 und 75-percentiles of 0.29) and 0.21 (25-percentiles of 0.14 and 75-percentiles of 0.25) for the computer-screen conditions.

We could not find a statistically significant difference in stimulus value ratings ($t_{(95)}$ = 1.82, p = 0.071) between conditions of VR (Mean = 0.24; SEM = 0.012) compared to both conventional 2D-computer-screen based conditions taken together (Mean = 0.21; SEM = 0.015; see Fig. 5). Furthermore, we tested for associations of stimulus values and the amount of implicit and explicit knowledge separately for each task. We found no correlation of the stimulus value ratings and the performance of implicit memory, as well as no correlation between the stimulus value ratings and the performance of explicit memory (all $p > 0.05$).

We additionally calculated an ANCOVA for both post-experimental assessments thereby the stimulus value rating functions as the covariate to test whether the interaction effect is affected by the stimulus value. We gained comparable significant effects to the previous performed ANOVA. The ANCOVA for the completion task resulted in an interaction effect between Condition and Learning-Type (F (2188) = 2.82, $p < 0.0001$) and a main effect for Learning-Type: F (1188) = 12.21, $p < 0.0001$). The ANCOVA for the free recall resulted in an interaction effect between Condition and Learning-Type (F (2188) = 31.79, $p < 0.0001$) and a main effect for Learning-Type: F (1188) = 73.45, $p < 0.0001$).

Discussion

Although incidental learning is a fundamental process occurring in everyday life, its underlying mechanism is commonly investigated in artificial laboratory environments. In particular, the potential benefits for the generation of explicit memory during learning in naturalistic contexts remain elusive. In this study, we aimed on closing this gap by implementing a near-natural sequential association task in an enriched environment presented in VR. We compared the generated knowledge in the enriched task environment with the learning behavior we got when participants perform a sequential association task in a classical and non-enriched environment. Our results revealed that participants successfully acquired a comparable amount of memory in both enriched and non-enriched learning environments.

Enhanced explicit memory formation in an enriched task design. Although the amount of learned items was similar across all applications, two memory tests (completion and free recall task) revealed that the extent of later expressed explicit memory was enhanced in the enriched environment compared to the conventional non-enriched environment. These results support the view that an enriched setting is an essential factor that can explain an increased explicit memory in the VR application of the incidental sequential association task. Therefore, the *enriched environmental hypothesis* is supported by our results. We could not find a significant difference in the visual stimulus value rating used in the VR condition compared to the conventional 2D-computer-screen based applications, indicating that the stimulus alone outside the VR cannot explain the difference in the explicit memory generation. Therefore, the enhanced development of explicit memory can be related to various advantages of the VR application[19,21,22,39]. One possible factor includes that VR applications facilitate the transition of information by its general characteristics in mobility and natural movements like the upright body position, and not only arms but especially head movement freedom. That way, VR applications might reflect a motivating environment with the potential to increase the feeling of an embodiment, like being physically present in the scenario, by which encoding efficiency and the transition into explicit memory are increased[25,40]. Closely related, evidence from Smith & Mulligan (2021)[25] includes the concept of immersion, which represents the degree of natural features in the VR portrayed by the virtual environment's sensory and interactive properties (see[41,42]). Previous research showed that immersion strongly correlates with participants' attentional engagement level[43,44]. Hence, the VR application not only provides a less vulnerable setting to attention decreases (i.e., due to mind wandering), it might even enhance attentional processes[16,25,45]. These advantages are a potential explanation for the enhanced explicit memory in the VR condition and constitute to the higher enriched environmental aspects that are increased in more real-life settings. Besides the environment posing an essential factor, nearnatural perception is also increased when stimulus material is enriched by context-related components, enabling superior memory formation[46–48]. These factors of the *enriched environmental hypothesis* have a clear impact on the transition from implicit to explicit memory.

The role of motivation in learning and why an enriched environment facilitates explicit memory formation by boosting motivational and attentional processes. In neuroscience, recent evidence has specifically connected the ventral striatum as a key player in the transition mechanism of implicit to explicit memory formation in incidental learning. The ventral striatum is mainly associated to motivational and reward processes [49] releasing dopamine in rewarding situations, consequently enhancing the generation of explicit memory. In a study by Clos et al. (2018)[10], which also included a sequential task, the dopaminergic level was pharmacologically modulated in human adults. It was shown that an increase in dopamine was directly connected to an increased transfer of information. This increase in information transfer was linked to the enhanced formation of explicit memory in an incidental sequential task. This way, successful predictions during task processing reflect an achievement within our neural system rewards. These rewarding processes encourage insight into hidden regularities and the emergence of explicit memory. The role of the dopaminergic system in the transfer from implicit to explicit memory may be related to the increased explicit memory rate in the present study as a consequence of the more rewarding and motivating experimental setting in an enriched task environment like we introduced to the participants in the VR surrounding.

The fluency hypothesis versus the enriched environmental hypothesis. To control for the feeling of task fluency (see the introduction for further details), we implemented two simple conventional 2D-computer-screen based versions, which differed in experiment length. The increased amount of implicit memory in the 2D-computer-screen condition indicated that the content is still learned, but is transferred to explicit memory only to a reduced degree. By directly comparing the effect of the *fluency* factor and the *enriched environmental* factor in the present experiment, the impact of the motivational and enriched experimental environment seems to have a larger effect on the emergence of explicit memory.

Conclusion. To summarize, the increased transition from implicit to explicit memory during incidental learning in the VR setting is related to an interaction of enhanced task processing, rewarding processes and attentional as well as motivational factors. These observations in the VR application have strong therapeutic implications for the rehabilitation of patients, who suffer from learning and memory impairments according to neurological diseases. We show that the transition from implicit to explicit memory is considerably influenced by the enriched environment, which includes context-enriched stimuli. In this work, we have obtained compelling evidence that the transition from implicit to explicit memory is primarily influenced by environment engaging processes like attention, motivation, and presence.

Limitations of this study. Within this study, we gained evidence that the environment used for studying incidental learning in humans influences the outcome of the type of generated knowledge to a great extent. As this study is one of the first attempts in elucidating the complex processes such an enriched and near-natural study design, here presented in VR, can have on incidental learning, we cannot clearly state which of the factors mentioned within our discussion plays a key role that led to the enhanced transfer from implicit to explicit memory formation. Yet, the findings emphasize the relevance for investigating learning and memory processes in more near-natural scenarios.

Data availability

The datasets generated and/or analyzed during this study are available on request from the corresponding author.

References

1. Anderson, B. A. The attention habit: How reward learning shapes attentional selection. *Ann. N. Y. Acad. Sci.* **1369**(1), 24–39. https://doi.org/10.1111/nyas.12957 (2016).
2. Astleitner, H. & Wiesner, C. An integrated model of multimedia effects on learning. *J. Educ. Multimed. Hypermedia* **11**(2), 177–191 (2002).
3. Cleeremans, A., Jiménez, L. & French, R. M. Implicit learning and consciousness: A graded, dynamic perspective. *Implicit Learn Conscious* https://doi.org/10.4324/9780203989401-9 (2002).
4. Esser, S. & Haider, H. Action–effects enhance explicit sequential learning. *Psychol. Res.* **82**(6), 1113–1129. https://doi.org/10.1007/s00426-017-0883-5 (2018).
5. Haider, H., Eberhardt, K., Kunde, A. & Rose, M. Implicit visual learning and the expression of learning. *Conscious Cogn.* **22**(1), 82–98. https://doi.org/10.1016/j.concog.2012.11.003 (2013).
6. Rose, M., Haider, H. & Büchel, C. The emergence of explicit memory during learning. *Cereb. Cortex.* **20**(12), 2787–2797. https://doi.org/10.1093/cercor/bhq025 (2010).
7. Esser, S., Lustig, C. & Haider, H. What triggers explicit awareness in implicit sequence learning? Implications from theories of consciousness. *Psychol. Res.* **86**, 1442–1457. https://doi.org/10.1007/s00426-021-01594-3 (2021).
8. Goujon, A., Didierjean, A. & Poulet, S. The emergence of explicit knowledge from implicit learning. *Mem. Cogn.* **42**(2), 225–236. https://doi.org/10.3758/s13421-013-0355-0 (2014).
9. Rünger, D. & Frensch, P. A. How incidental sequence learning creates reportable knowledge: The role of unexpected events. *J. Exp. Psychol. Learn. Mem. Cogn.* **34**(5), 1011–1026. https://doi.org/10.1037/a0012942 (2008).
10. Clos, M., Sommer, T., Schneider, S. L. & Rose, M. Enhanced transformation of incidentally learned knowledge into explicit memory by dopaminergic modulation. *PLoS ONE* **13**(6), 1–15. https://doi.org/10.1371/journal.pone.0199013 (2018).
11. Lustig, C., Esser, S. & Haider, H. The interplay between unexpected events and behavior in the development of explicit knowledge in implicit sequence learning. *Psychol. Res.* https://doi.org/10.1007/s00426-021-01630-2 (2021).
12. Schwabe, L. & Wolf, O. T. The context counts: Congruent learning and testing environments prevent memory retrieval impairment following stress. *Cogn. Affect. Behav. Neurosci.* **9**(3), 229–236. https://doi.org/10.3758/CABN.9.3.229 (2009).
13. Smith, S. M. & Vela, E. Environmental context-dependent memory. *Psychon. Bull. Rev.* **8**(2), 203–220 (2001).
14. Hölle, D., Meekes, J. & Bleichner, M. G. Mobile ear-EEG to study auditory attention in everyday life: Auditory attention in everyday life. *Behav. Res. Methods* **53**(5), 2025–2036. https://doi.org/10.3758/s13428-021-01538-0 (2021).
15. Piñeyro Salvidegoitia, M. *et al.* Out and about: Subsequent memory effect captured in a natural outdoor environment with smartphone EEG. *Psychophysiology* **56**(5), 1–15. https://doi.org/10.1111/psyp.13331 (2019).
16. Lüddecke, R. & Felnhofer, A. Virtual reality biofeedback in health: A scoping review. *Appl. Psychophysiol. Biofeedback* **47**(1), 1–15. https://doi.org/10.1007/s10484-021-09529-9 (2021).
17. Maarsingh, B. M., Bos, J., Van Tuijn, C. F. J. & Renard, S. B. Changing stress Mindset through Stressjam: A virtual reality game using biofeedback. *Games Health J.* **8**(5), 326–331. https://doi.org/10.1089/g4h.2018.0145 (2019).
18. Talsma, D., Senkowski, D., Soto-Faraco, S. & Woldorff, M. G. The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* **14**(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008 (2010).
19. Bohil, C. J., Alicea, B. & Biocca, F. A. Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* https://doi.org/10.1038/NRN3122 (2011).
20. Schmidt, S., Bruder, G. & Steinicke, F. Effects of virtual agent and object representation on experiencing exhibited artifacts. *Comput. Graph.* **83**, 1–10. https://doi.org/10.1016/j.cag.2019.06.002 (2019).
21. Schomaker, J. Unexplored territory: Beneficial effects of novelty on memory. *Neurobiol. Learn. Mem.* **2019**(161), 46–50. https://doi.org/10.1016/j.nlm.2019.03.005 (2018).
22. Smith, S. A. Virtual reality in episodic memory research: A review. *Psychon. Bull. Rev.* **26**(4), 1213–1237. https://doi.org/10.3758/s13423-019-01605-w (2019).
23. Chelazzi, L., Perlato, A., Santandrea, E. & Della, L. C. Rewards teach visual selective attention. *Vis. Res.* **85**, 58–72. https://doi.org/10.1016/j.visres.2012.12.005 (2013).
24. Sanz, L. R. D., Vuilleumier, P. & Bourgeois, A. Cross-modal integration during value-driven attentional capture. *Neuropsychologia* **120**, 105–112. https://doi.org/10.1016/j.neuropsychologia.2018.10.014 (2018).
25. Smith, S. A. & Mulligan, N. W. Immersion, presence, and episodic memory in virtual reality environments. *Memory* **29**(8), 983–1005. https://doi.org/10.1080/09658211.2021.1953535 (2021).
26. Dehn, L. B. *et al.* Training in a comprehensive everyday-like virtual reality environment compared to computerized cognitive training for patients with depression. *Comput. Hum. Behav.* **79**, 40–52. https://doi.org/10.1016/j.chb.2017.10.019 (2018).
27. Ruddle, R. A., Volkova, E., Mohler, B. & Bülthoff, H. H. The effect of landmark and body-based sensory information on route knowledge. *Mem. Cogn.* **39**(4), 686–699. https://doi.org/10.3758/s13421-010-0054-z (2011).
28. Kemény, F. & Lukács, Á. Sequence in a sequence: Learning of auditory but not visual patterns within a multimodal sequence. *Acta Psychol. (Amst.)* https://doi.org/10.1016/j.actpsy.2019.102905 (2019).
29. Zhao, F., Gaschler, R., Nöhring, D. O., Röttger, E. & Haider, H. Sequential modulation of across-task congruency in the serial reaction time task. *Acta Psychol. (Amst.)* **205**, 103043. https://doi.org/10.1016/j.actpsy.2020.103043 (2020).
30. Bavelier, D. & Neville, H. J. Cross-modal plasticity: Where and how?. *Nat. Rev. Neurosci.* **3**(6), 443–452. https://doi.org/10.1038/nrn848 (2002).
31. Butler, A. J. & James, K. H. Cross-modal versus within-modal recall: Differences in behavioral and brain responses. *Behav. Brain Res.* **224**(2), 387–396. https://doi.org/10.1016/J.BBR.2011.06.017 (2011).
32. Shams, L. & Seitz, A. R. Benefits of multisensory learning. *Trends Cogn. Sci.* **12**(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006 (2008).
33. Rose, M., Haider, H., Salari, N. & Büchel, C. Functional dissociation of Hippocampal mechanism during implicit learning based on the domain of associations. *J. Neurosci.* **31**(39), 13739–13745. https://doi.org/10.1523/JNEUROSCI.3020-11.2011 (2011).
34. Haider, H., Eberhardt, K., Esser, S. & Rose, M. Implicit visual learning: How the task set modulates learning by determining the stimulus-response binding. *Conscious Cogn.* **26**(1), 145–161. https://doi.org/10.1016/j.concog.2014.03.005 (2014).
35. Jablonowski, J. & Rose, M. The functional dissociation of posterior parietal regions during multimodal memory formation. 1–17. https://doi.org/10.1002/hbm.25861 (2022).
36. Esser, S. & Haider, H. The emergence of explicit knowledge in a serial reaction time task: The role of experienced fluency and strength of representation. *Front. Psychol.* **8**(502), 1–18. https://doi.org/10.3389/fpsyg.2017.00502 (2017).
37. Dienes, Z. & Seth, A. Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Conscious Cogn.* **19**(2), 674–681. https://doi.org/10.1016/j.concog.2009.09.009 (2010).
38. Conway, C. M. & Christiansen, M. H. Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychol. Sci.* **17**(10), 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x (2006).

39. Plancher, G., Barra, J., Orriols, E. & Piolino, P. The influence of action on episodic memory: A virtual reality study. *Q. J. Exp. Psychol.* **66**(5), 895–909. https://doi.org/10.1080/17470218.2012.722657 (2013).
40. Kilteni, K., Groten, R. & Slater, M. The sense of embodiment in virtual reality. *Presence Teleoper. Virtual Environ.* **21**(4), 373–387. https://doi.org/10.1162/PRES_a_00124 (2012).
41. Fuchs, P., Moreau, G. & Guitton, P. *Virtual Reality: Concepts and Technologies* (CRC Press, Florida, 2011).
42. Slater, M. A note on presence terminology. *Presence Connect.* **3**(3), 1–5 (2003).
43. Darken, R. P., Bernatovich, D., Lawson, J. P. & Peterson, B. Quantitative measures of presence in virtual environments: The roles of attention and spatial comprehension. *Cyberpsychol. Behav. Impact Internet Multimed. Virtual Real Behav. Soc.* **2**(4), 337–347. https://doi.org/10.1089/cpb.1999.2.337 (1999).
44. Lessiter, J., Freeman, J., Keogh, E. & Davidoff, J. A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence Teleoper. Virtual Environ.* **10**(3), 282–297. https://doi.org/10.1162/105474601300343612 (2001).
45. Dehaene, S. & Changeux, J. P. Experimental and theoretical approaches to conscious processing. *Neuron* **70**(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018 (2011).
46. Eggermont, J. J. Multisensory processing. *Hear Loss* https://doi.org/10.1016/b978-0-12-805398-0.00003-7 (2017).
47. Schmid, C., Büchel, C. & Rose, M. The neural basis of visual dominance in the context of audio-visual object processing. *Neuroimage* **55**(1), 304–311. https://doi.org/10.1016/j.neuroimage.2010.11.051 (2011).
48. Van Atteveldt, N., Murray, M. M., Thut, G. & Schroeder, C. E. Multisensory integration: Flexible use of general operations. *Neuron* **81**(6), 1240–1253. https://doi.org/10.1016/j.neuron.2014.02.044 (2014).
49. Pool, E. R. *et al.* Differential contributions of ventral striatum Subregions to the motivational and hedonic components of the affective processing of reward. *J. Neurosci.* **42**(13), 2716–2728. https://doi.org/10.1523/JNEUROSCI.1124-21.2022 (2022).

Acknowledgements

The present research was supported by the German Research Foundation (DFG; SFB/Transregio 169, Project B3, C8 and DFG RO 2653/9-1). We thank Theresia Vordermark and Celia Kumpmann for helping measuring the data.

Author contributions

M.R. and F.S. designed the study. C.J. and P.T recorded the data, C.J., M.M. and M.R. analyzed the data and C.J.,
M.M. and M.R. wrote the manuscript. All authors reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.R.

7.3. Study 3: Crossmodal Correspondence Effect

Jaap, C., & Rose, M. (2024). Dissociable neuronal mechanism for different crossmodal correspondence effects in humans. *Acta Neurobiologiae Experimentalis*, 136–152.

RESEARCH PAPER

Acta Neurobiol Exp 2024, 84: 136–152
DOI: 10.55782/ane-2024-2439

Dissociable neuronal mechanism for different crossmodal correspondence effects in humans

Carina Jaap* and Michael Rose

NeuroImage Nord, Department for Systems Neuroscience, University Medical Center Hamburg Eppendorf, Hamburg, Germany
**Email: c.jaap@uke.de*

Crossmodal correspondences (CMCs) refer to associations between seemingly arbitrary stimulus features in different sensory modalities. Pitch‑ size correspondences refer to the strong association of e.g., small objects with high pitches. Pitch‑ elevation correspondences refer to the strong association of e.g., visuospatial elevated objects with high pitches. We used functional magnetic resonance imaging (fMRI) to study the neural components, which underlie the CMCs in pitch‑ size and spatial pitch‑ elevation. This study focuses on answering the question of whether or not different CMCs are driven by similar neural mechanisms. The comparison of congruent against incongruent trials allows the estimation of CMC effects across different CMCs. The analysis of the measured neural activity in different CMCs strongly pointed toward different mechanisms which are involved in the processing of pitch‑ size and pitch‑ elevation correspondences. Differential, whole brain effects were observed within the superior parietal lobule (SPL), cerebellum and Heschls' gyrus (HG). Further, the angular gyrus (AnG), the intraparietal sulcus (IPS) and anterior cingulate cortex (ACC) were engaged in processing the CMCs but showed different effects for processing congruent compared to incongruent stimulus presentations. Within pitch‑ size significant effects in the AnG and ACC were found for congruent stimulus presentations whereas for pitch‑ elevation, significant effects in the ACC and IPS were found for incongruent stimulus presentations. In summary, the present results indicated differential neural processing in different simple audio‑ visual CMCs.

Key words: fMRI, crossmodal correspondence, multisensory, audio‑ visual, spatial, attention, perception

INTRODUCTION

Crossmodal correspondences (CMCs) refer to almost universally experienced (implicit) associations between stimulus features in different sensory modalities. A well-studied example of correspondence effect was found for pitch and visual elevation. When an object that is visually elevated in space is paired with a high-pitched tone, a stronger association of these features is observed compared to pairing the same object with a low-pitched tone (Ben-Artzi & Marks, 1995; Chiou & Rich, 2012; Evans, 2020; Evans & Treisman, 2010; Jamal et al., 2017; McCormick et al., 2018; Melara & Brien, 1987). A second pitch-based CMC is pitch and size. Presenting e.g., a small object together with a high-pitched tone resulted in a successful crossmodal correspondence in a study by Evans and Treisman (2010) and numerous other studies (Bien et al., 2012; Bonetti & Costa, 2018; Gallace & Spence, 2006; Parise & Spence, 2012).

There are several theories regarding the origin of pitch-size and pitch-elevation CMCs resulting in the assumption of different or common information-processing mechanisms for the different CMC effects. Despite numerous behavioral studies on crossmodal correspondences (Chiou & Rich, 2012; Evans & Treisman, 2010; Evans, 2020; Parise et al., 2014; Spence, 2011, 2020; Uno & Yokosawa, 2022a, 2022b), only a few studies used neuroimaging to address the neural basis of different pitch-based CMCs (McCormick et al., 2018; Sadaghiani et al., 2009).

A prominent theory about the origin of pitch-elevation correspondences is based on language processing (Parise et al., 2014; Spence, 2011, 2020; Spence & Sathian, 2020). In most cultures, the words ‘high’ and ‘low’ can describe both, the height of a pitch and the position of an object in space. This linguistic link is not described for pitch and size correspondences. Even though a language-driven cause for pitch-elevation associations is plausible (Ben-Artzi & Marks, 1999), a growing number of studies on the CMC effect between pitch and spatial-elevation raise the question if other variables than language probably cause the strong CMC of these seemingly arbitrary stimulus pairs (McCormick et al., 2018; Parise et al., 2014; Parkinson et al., 2012).

A second theory on CMCs declares that the correspondences between pitch and elevation as well as pitch and size probably arise from regularities in our natural environment that are stored in memory (Parise et al., 2014; Spence, 2011; 2020; Spence & Sathian, 2020). For example, larger bodies usually resonate lower pitches and smaller objects tend to resonate higher pitches (Parise et al., 2014). We are confronted with this regularity frequently in our daily lives. Children typically have a higher-pitched voice than adults (Lee et al., 1999) and small animals tend to make higher-pitched noises than larger animals (Bowling et al., 2017). We also tend to perceive higher-pitched tones from objects elevated in space than when on the ground (Parise et al., 2014). Following this assumption, CMCs probably have their roots in statistical regularities, i.e. naturally learned rules and assumptions from our environment (Parise et al., 2014; Spence, 2011; 2020; Spence & Sathian, 2020). If both CMCs have their origin in similar mechanisms, great activations within comparable brain regions will be measured in both pitch-size and pitch-elevation CMCs.

The third and last theory we are going to address is the theory of perceived intensity, which is also called a theory of magnitude (Spence, 2011). This theory declares that the CMC effect probably evolved from a correspondence in intensity or magnitude in the underlying neuronal structure of corresponding stimulus pairs (Spence, 2011; Spence & Sathian, 2020). The main idea underlying the magnitude in CMCs is a shared polar dimension of the stimulus pairs perceived as congruent. According to this notion, a high-pitched tone and a small visual stimulus would be situated on the same side of their respective polar dimension. Compared to incongruent stimulus pairs, congruent pairs would share ‘more’ in terms of intensity or magnitude (Chang & Cho, 2015). A common neural activation in terms of magnitude was found for e.g., numbers by Piazza et al. (2007) and sizes with luminance by Pinel and colleagues (2004). If pitch and size and pitch and elevation correspondences have their origin in similar coded neuronal responses, we hypothesize to find greater activations within the intraparietal sulcus (IPS) for congruent trials as a common effect in both CMCs (Humphreys & Ralph, 2015; Piazza et al., 2007; Pinel et al., 2004).

The CMC is in behavioral studies often measured *via* the reaction time (RT) differences between congruent and incongruent stimulus pairs. Thus being significant, these differences are rather small in absolute values (Chiou & Rich, 2012; Evans & Treisman, 2010). The study by Evans and Treisman, performed in 2010, included eight subjects in their pitch-size visual experiment, in which the absolute difference between congruent and incongruent trials was 14.4 ms. Their pitch-elevation visual paradigm included twelve participants and the absolute difference between congruent and incongruent RTs was 18.6 ms. Within their fMRI paradigm, McCormick and colleagues (2018) did not find significant RT differences between congruent and incongruent stimulus presentation in the pitch-elevation CMC, what may be related to the overall small size of the effect. They validated their findings outside the scanner with a behavioral task (McCormick et al., 2018). Based on these previous findings, performing a behavioral test outside the scanner appears to be an appropriate measure to validate CMC effects studied with fMRI (Koten et al., 2013; McCormick et al., 2018). In our study, we implemented a congruence classification task outside the scanner to measure the behavioral CMC effect in addition to the typically measured RTs.

Even though pitch-based correspondences are almost universally experienced and well-studied (Ben-Artzi and Marks, 1995; Bien et al., 2012; Chiou & Rich, 2012; Evans, 2020; Evans and Treisman, 2010; Gallace & Spence, 2006; Jamal et al., 2017; Marks, 1987; Spence, 2011; 2020; Zeljko et al., 2019), the evidence for the underlying neural mechanisms is still lacking. A study that used functional magnetic resonance imaging (fMRI) to examine pitch and elevation congruencies showed a probable involvement of the right angular gyrus (AnG) as well as the mid-IPS for corresponding stimulus presentations (McCormick et al., 2018).

The main focus in our study was to examine the neural basis of the processing of a pitch-elevation CMC and compare this to a pitch-size CMC while both CMCs are always in the focus. The estimation of the CMC effect can be achieved by the calculation of the difference of congruent > incongruent (C > I) presentations. The calculated difference then allows a direct comparison of the neural substrates of the CMC effect between the different CMCs. This comparison can be used to test common or different neural correlates of different CMCs focusing on the CMC effect, thus directly testing the different theoretical assumptions about the origin of the CMC effect.

If we find a common effect within the IPS for congruent > incongruent presentations, a magnitude driven CMC is likely to cause this effect. An effect within the left inferior frontal gyrus (IFG) is favorable for a CMC driven by language, which we hypothesize to find most likely for congruent > incongruent pitch-elevation presentations. However, if CMCs are based on statistical representations of our environment, we will most likely find an effect within areas common for attention and memory retrieval like the anterior cingulate cortex (ACC) or the AnG.

Although we are mainly interested in congruency effects, it cannot be excluded that effects for incongruent stimulus presentations are also part of the processing of the stimuli in our tested CMCs. Stronger effects for incongruent stimuli could be due to for example response conflict or a shift of attention (Chiou & Rich, 2012; Spence & Sathian, 2020).

Besides the question about the neural mechanisms between two different CMCs, we were interested in a probable modulation of the effect within one CMC by stimulus contrast. It has been hypothesized that the CMC effect depends on the ability to form a unique correspondence between stimulus pairs (Chiou & Rich, 2012). Therefore, we additionally measured a variant of the pitch-size CMC, in which we reduced the difference, i.e. the contrast between the stimuli to probably also reduce the CMC effect (Chiou & Rich, 2012). If the CMC effect is modulated by the contrast of the stimulus pairs, we hypothesized to find a reduced neural effect for the variant of the pitch-size CMC with a reduced difference between the stimulus pairs.

METHODS

Participants

Thirty-three mentally and physically healthy participants (21 females, age M=24.8 years, SD=3.8 years) with normal hearing and normal or corrected-to-normal vision took part in this experiment. The participants were recruited through a local online job platform. Four participants had to be excluded from the final analysis (two due to technical issues and two due to excessive movement in the scanner (>5 mm)). Therefore, the final sample size was 29 participants. All experiment protocols were approved by the Ethics Committee of the General Medical Council Hamburg (PV7022) and all our methods were carried out in accordance with relevant ethical guidelines and regulations. All participants gave their written informed consent and were paid an expense allowance of 10 €/h.

Apparatus

Inside the scanner

The stimuli were presented using Presentation® software (Version 22.01, Neurobehavioral Systems, Inc., Berkeley, CA) running on Windows 7. A mirror placed on the head coil with ~ 12 cm distance to the participant’s face was used to reflect the stimulus presentation from a 40’’ LCD screen with a refresh rate of 60 Hz. The auditory stimuli were presented using MR compatible in-ear head phones (MR confon). Participant responses were tracked using two MR compatible button boxes.

Outside the scanner

For a tutorial as well as the congruence classification task, Psychopy (Version 3.2.4) software running on a 15’ hp laptop with Windows 10 was used to present the stimuli. A button box with two active buttons (one for each hand) was used to track the participants’ responses. Auditory stimuli were presented *via* loudspeaker on both sides of the screen.

Stimuli

Black squares on a grey background were used as visual stimuli (Fig. 1A, B) and instrument tones (edited with Audacity® recording and editing software version 2.4.1) were used as acoustic stimuli. For pitch-size, squares (0° 39’ .26’’ & 3° 55’ .03’’) were presented with the sound of a piccolo flute (1225 Hz, D#/Eb6) or a double bass (73 Hz, D2). For pitch-size variant with reduced difference, squares (1° 18’ .30’’ & 2° 36’ .79’’) were presented either with a violin (588 Hz, D5) or a bassoon (149 Hz, D3). In between trials, a 0° 29’ .50’’, white fixation cross was presented in the middle of the screen. Small squares presented together with high-pitched tones and bigger squares presented together with low-pitched tones will be referred to as pitch-size congruent condition in the following (Fig. 1 A).
For measuring the pitch-elevation CMC, a black square (1° 8’ .54’’) was presented either above or below a 0° 29’ .50’’, white fixation cross. The cross was presented in the center of the screen and the distance of the squares to the center was 3° 14’ .18’’. The auditory stimuli used were the same as in the variant of pitch-size with reduced difference. Squares above the fixation cross presented together with higher pitched tones and squares presented below the fixation cross together with lower tones will be referred to as congruent trials in the following (Fig. 1 A).

Experimental design and procedure

We tested two different CMC types within our study, pitch-size and pitch-elevation, as well as a variant of the pitch-size CMC with reduced difference between the stimulus contrasts. A tutorial was performed by the participants before entering the scanner. Within the tutorial, the two distinct CMCs as well as the pitch-size variant with reduced difference were introduced separately to the participants. Each tutorial part for the two distinct CMCs and the pitch-size variant with reduced difference consisted of twelve trials (8 congruent; 4 incongruent). Within the tutorial, the participants were not introduced to the concept of congruence and incongruence hidden behind the stimulus pairings. The purpose of the tutorial was to familiarize the participants with the stimulus pairs and the focus was always on the visual stimuli.

After the tutorial, the participants were placed in the scanner. Before the experiment started, the volume of the acoustic stimuli were adjusted while the participants were exposed to the scanner noise. With the latter procedure we ensured a comfortable but valid presentation of the acoustic stimuli in the scanner.

We used an event-related design with jittered inter trial intervals (ITI) to present the stimuli in the scanner. In the main experiment, each participant saw all CMCs, the two distinct CMCs and the variant of pitch-size with reduced difference (Fig. 1A, B), in separate runs. The duration of a run was ~ 10 minutes. The order in which the two distinct CMCs or the variant of pitch-size with reduced difference were presented was counterbalanced between participants.

Each of the three runs, in which one of the two distinct CMCs (Fig. 1A) or the variant of pitch-size with reduced difference (Fig. 1B) was presented, consisted of 96 trials with 48 repetitions of each condition (congruent; incongruent) and 24 presentations of each stimulus pair (e.g., small square and high pitch) (Chiou & Rich, 2012). The 96 trials were presented in a pseudo-randomized order and this order was also randomized between participants. In each trial, a visual stimulus was presented simultaneously with a sound (Fig. 1C). The participants were instructed to respond to the different visual stimuli as fast and precise as possible. For small as well as elevated stimuli, the correct button press was performed with the left index finger. For large and low presented stimuli, the button press was performed with the right index finger. The audio-visual presentation lasted for 1000 ms followed by 500 ms of extended key response time. The inter-trial interval was jittered between 2000 – 8000 ms with a mean of ~ 5000 ms (Fig. 1C). Instructions were prompted on a screen in the scanner before each new run started. The participants had the opportunity to take a short break between the runs, however, they had to stay in the scanner during the short break.

A stimulus congruence classification task was performed outside the scanner following the main experiment. Within the stimulus congruence classification, the participants were instructed to classify if the audio-visual stimulus presentations match each other or not (Fig. 1D). The congruence classifications was separately performed for each distinct CMC and variant of pitch-size with reduced difference, whereby each condition (congruent; incongruent) was presented six times in a random order. The participants were instructed to classify the presented audio-visual pairs by clicking on the respective side of a scale with a computer mouse (Fig. 1D). Thereby only the ends of the scale could be clicked, no gradual adjustment was possible. The participants were instructed to classify intuitively if the audio-visual stimuli were matching or not. No feedback on the chosen pair was given. We conducted this final task to test whether the participants correctly matched the congruent and incongruent stimulus pairs in accordance with the CMC theory (Fig. 1A, B).

Behavioral data analysis

The focus of the analysis of the behavioral data was the stimulus congruence classification performed outside the scanner. All statistical tests on the behavioral data were performed in JASP (Version 0.16.1).

All congruence classifications were taken into account for the further analysis. The classifications were then divided into trials in which participants chose ‘matching’ and trials in which participants chose ‘not matching’ separately for each condition and each CMC, i.e., the two distinct CMCs (Table 1) and the variant of pitch-size with reduced difference (Table 2). We were interested in whether participants would classify our congruent stimulus pairs as matching and our incongruent stimulus pairs as not matching, i.e. whether participants show the expected classification of pairs in accordance with the CMC theory. We also wanted to know whether these classifications are dependent on the tested CMCs. Therefore, we conducted a repeated measures ANOVA to test the effect of the within-subject factors 'distinct CMCs (pitch-size & pitch-elevation),' as well as 'Classifications of congruent stimuli (congruent stimuli rated as matching & congruent stimuli rated as not matching)' on stimulus classifications. We also conducted a second repeated measures ANOVA to test the effect of the within subject factors 'pitch-size and pitch-size variant with reduced difference', as well as 'Classifications of congruent stimuli (congruent stimuli rated as matching & congruent stimuli rated as not matching)' on stimulus classifications. We also conducted these repeated measures ANOVAs for the stimulus classification performed on incongruent stimuli.

For the sake of completeness, we analyzed the RTs of the in-scanner task. Only RTs from correct trials and trials with RTs below 1000 ms were taken into account in the further analysis. We chose this threshold, as the stimulus presentation ended after 1000 ms. Furthermore, we wanted to avoid including decisions formed by e.g., complex cognitive processing or inattentiveness. We performed two repeated measures ANOVAs, one for the two distinct CMCs and another for the pitch-size CMC and its variant with reduced difference, to test whether there are differences between the RTs of the trial conditions (congruent trials & incongruent trials) and whether these differences are different between the CMCs. One repeated measures ANOVA was conducted to compare the within subject factors 'distinct CMCs (pitch-size & pitch-elevation)', as well as 'Condition (congruent trials & incongruent trials)' on RTs and the other repeated measures ANOVA was conducted to compare the effect of the within subject factors 'pitch-size and its variant with reduced difference', as well as 'Condition (congruent trials & incongruent trials)' on RTs.

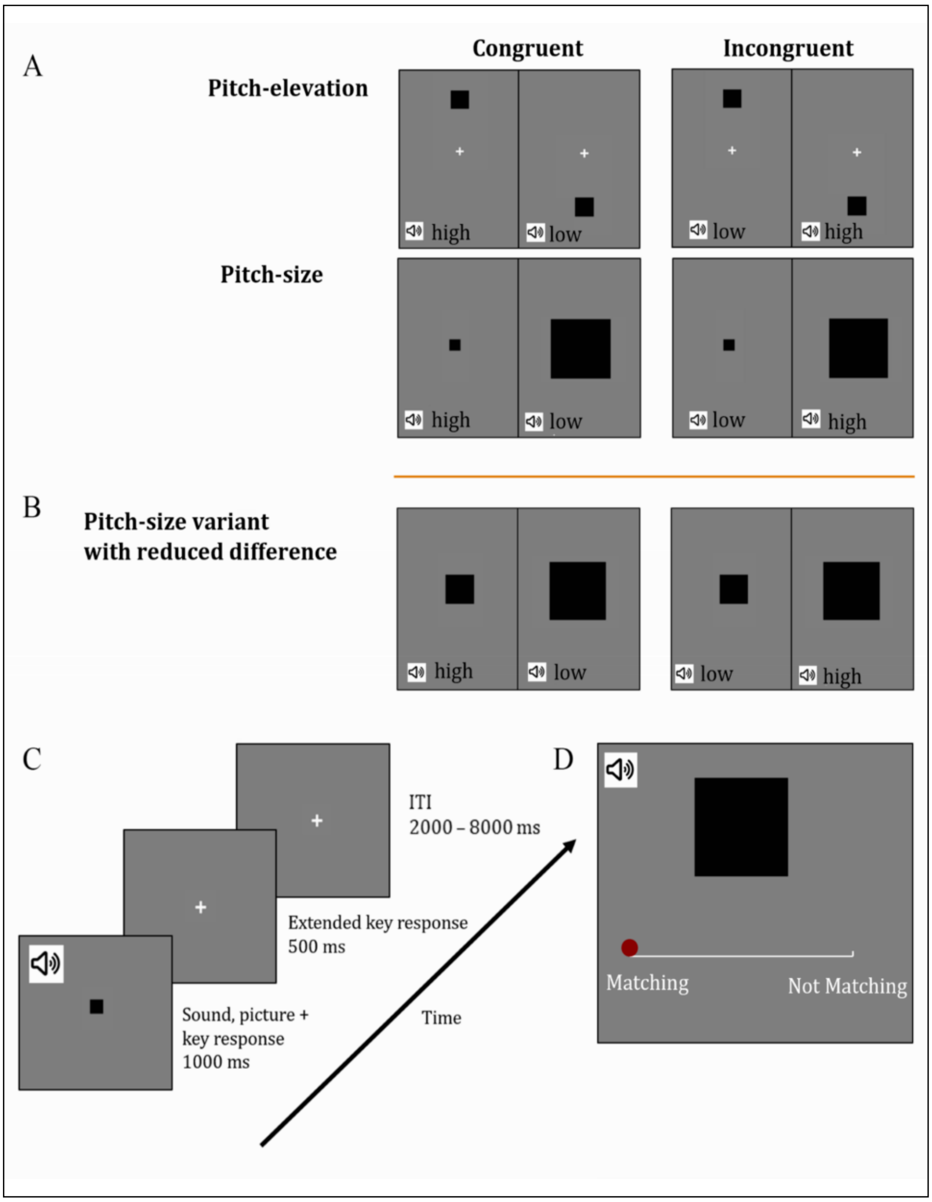


Fig. 1. (A) Overview of the congruent (left) and incongruent (right) stimulus pairs for each of the two distinct CMCs (from top to bottom: Pitch-elevation, pitch-size). (B) Overview of the congruent (left) and incongruent (right) stimulus pairs for the pitch-size variant with reduced difference between the visual and acoustic stimulus pairs. (C) Schematic sequence of events in a trial (here pitch-size). (D) Example of a trial within the post experimental test for the strength of congruence outside the scanner. Participants were asked to classify simultaneous sound and square presentations as matching or not matching without knowing the purpose of the main experiment. Only the ends of the scale could be clicked, no gradual adjustment was possible.

Functional data acquisition

Imaging was performed on a 3-T scanner (Siemens Trio) using a 64-channel head coil. A standard gradient echo-planar imaging (EPI) T2*-sensitive sequence with 54 axial slices (2 mm thickness with .5 mm gap, voxel size=2 × 2 × 2 mm; repetition time (TR)=1640 ms and echo time (TE)=29 ms, multiband factor=2, Flip angle=70°) was acquired for functional imaging. A high resolution (1 × 1 × 1 mm voxel size) T1 weighted, three dimensional, defaced MPRAGE image (TR=7.1 ms, TE=2.98 ms, FA=9°, inversion time=1100 ms) was additionally acquired for each participant. The experiment started after the scanner reached magnetic stabilization.

Functional data analysis

Preprocessing and statistical analysis of the fMRI data were carried out in SPM12 (http://www.fil.ion. ucl.ac.uk/spm/) on Matlab version R2020a. Image preprocessing steps included a correction for the magnetic field distortion by unwarping the images using a fieldmap, as well as motion correction with registration on the first EPI, correcting for between subject anatomical differences by normalizing images on EPI with the EPI template provided by SPM12 and smoothing the normalized images with a 6 mm (full widths half maximum; FWHM) Gaussian kernel. We did not correct for RT differences between congruent and incongruent conditions as the expected RT difference was below 100 ms (Chiou & Rich, 2012; Evans & Treisman, 2010).

The hemodynamic response for each condition (congruent; incongruent) was modelled as an event-related design (for further information see Experimental design and procedure). The six contrast images (main effects) per participant, calculated from onsets of each condition, were entered into a flexible factorial group level analysis and all statistical comparisons were estimated on the group level.

Functional data analysis of the distinct CMCs

To test for differences between the processing of the conditions in the distinct pitch-size and pitch-elevation CMCs, we estimated two interaction contrasts at the second level. To test for enhanced neural effects of congruent trials selectively in the pitch-size CMC an interaction contrast was estimated (Pitch-size (C>I) > pitch-elevation (C<I)). To test for enhanced neural effects of congruent trials selectively in the pitch-elevation CMC, another interaction contrast was estimated (Pitch-size (C<I) < pitch-elevation (C>I)). We also tested for differences in neural effects between the conditions within the distinct CMCs. Therefore, we estimated contrasts that tested for enhanced neural effects of congruent stimuli within pitch-size (C>I) and pitch-elevation (C>I), as well as contrasts that tested for enhanced neural effects of incongruent stimuli within pitch-size (C<I) and pitch-elevation (C<I) (Table 3). To test for common neural effects of congruent trials within the two distinct CMCs, a global conjunction was estimated (Pitch-size (C>I) & pitch-elevation (C>I)). Statistically significant whole brain fMRI effects were family wise error corrected (FWE, *p*<0.05).

Functional data analysis of pitch-size and its variant with reduced difference

To test for common neural effects of congruent trials within pitch-size and its variant with reduced difference, a global conjunction was estimated (Pitch-size (C>I) & pitch-size variant with reduced difference
(C >I)).

To test for enhanced neural effects of congruent trials selectively in the pitch-size CMC, an interaction contrast was estimated (Pitch-size (C>I) > pitch-size variant with reduced difference (C<I)). Statistically significant whole brain fMRI effects were family wise error corrected (FWE, *p*<0.05).

Regions of interest

Main region of interests related to previous studies and their corresponding coordinates used in this study are the ACC [TAL: x=+/- 6, y=30, z=38] (Roelofs et al., 2006; coordinates were converted into MNI space using the implemented mni2tal tool from Yale Bioimage Suite Web (Version: 1.2.0 (2020/08/25)), the AnG [MNI: x=+/- 48, y=-64, z=34] to test for a statistical or environmental driven CMC (Humphreys & Ralph, 2015) and the IPS for a magnitude driven CMC [MNI: x=+/- 43, y=-42, z=48] (Humphreys & Ralph, 2015). We tested for language related effects within the triangular part [MNI: x=-39, y=26, z=13] of the left IFG as well as the orbital part [MNI: x=-30, y=35, z=-14] of the left IFG (Liuzzi et al., 2017). We reported significant fMRI effects of our ROIs using a sphere with a radius of 10 mm which was small volume corrected with FWE ($p<0.05$).

RESULTS

Behavioral results

Behavioral results of the stimulus congruence classification task

We tested separately whether the congruent and incongruent trials were overall classified as matching compared to not matching with a classification task. We also tested whether there is a difference in classifications of congruence or incongruence depending on the CMC. We tested this dependence between the distinct CMCs and between pitch-size and its variant with reduced difference. The results of the classification task, which was performed outside the scanner, showed that the congruent stimulus pairs were overall classified as matching and incongruent stimulus presentations were overall classified as not matching, i.e., the participants' classification aligns with the CMC theory (Spence, 2011). A dependence of the classification strength was observed for the classification of congruent trials between pitch-size and pitch-elevation with more congruent trials rated as matching in pitch-size compared to pitch-elevation (details in the following sections).

Stimulus congruence classification results of the two distinct CMCs

To test whether congruent stimulus pairs were significantly classified as matching by the participants (Table 1) and to test whether this classification different between the distinct CMCs, a repeated measures ANOVA with the distinct CMCs and Classifications of congruent stimuli (congruent trials rated as matching & congruent trials rated as not matching) as within subject factors was performed. For the distinct CMCs, this repeated-measures ANOVA showed a reliable effect for the factor Classifications of congruent stimuli ($F_{(1,28)}=133.8$, $p<0.001$, $hp2=0.83$; Table 1). This means that congruent pairs were significantly classified as matching (M=86.2%, SEM=3.9%; Table 1). An interaction of distinct CMCs × Classifications of congruent stimuli was statistically significant ($F_{(1,28)}=7.398$, $p=0.011$, $hp2=0.209$) and a *post-hoc* test revealed that significantly more congruent pairs were classified as matching within pitch-size compared to pitch-elevation ($pholm=0.022$). This means that a significant difference was observed for the number of congruent trials rated as matching between the pitch-size and pitch-elevation CMCs, i.e., a higher congruence classification was observed for congruent stimuli in pitch-size compared to pitch-elevation (Table 1). Furthermore, significantly more congruent pairs were classified as matching within pitch-size ($pholm<.001$) as well as within pitch-elevation ($pholm<.001$). We observed no effect for distinct CMCs ($F_{(1,28)}=-6.010e^{-14}$, $p=1.0$, $hp2=-2.146e^{-15}$).

To test whether incongruent stimulus pairs were significantly classified as not matching by the participants and to test whether this classification is different between the distinct CMCs this classification is different between the distinct CMCs, a repeated measures ANOVA with the distinct CMCs and Classi-

Table 1. Congruence classification of the two distinct CMCs for congruent and incongruent stimulus presentations. Mean in percent for stimulus

conditions (congruent, incongruent) within each distinct CMC (pitch- size, pitch- elevation) classified as matching, i.e. congruent or not matching, i.e. incongruent. The maximum possible percentage for each presented condition (congruent; incongruent) in each distinct CMC was 100% for classifications of matching and not matching taken together.

Condition presented	Pitch- size classified as		Pitch- elevation classified as	
	matching	not matching	matching	not matching
congruent	92	8	80.5	19.5
incongruent	23.6	76.4	31	69

fications of incongruent stimuli (incongruent trials rated as matching & incongruent trials rated as not matching) as within subject factors was performed. For the distinct CMCs, this repeated-measures ANOVA showed a reliable effect for the factor Classifications of incongruent stimuli (F$_{(1,28)}$=22.34, *p*<.001, *hp2*=0.44). This means that incongruent pairs were significantly classified as not matching (M=72.7%, SEM=5.6%; Table 1). No effect for distinct CMCs (F$_{(1,28)}$=-4.695e^{-14}, *p*=1.0, *hp2=-1.677e-15*) as well as the interaction of distinct CMCs × Classifications of incongruent stimuli (F$_{(1,28)}$=1.714, *p*=0.2, *hp2=0*.058) was observed. This means that no significant difference was observed between pitch-size and pitch-elevation for the classification of incongruent stimuli.

Stimulus congruence classification results of the pitch- size CMC and its variant with reduced difference

To test whether congruent stimulus pairs were significantly classified as matching by the participants (Table 2) and to test whether this classification is different between pitch-size CMC and its variant with reduced difference, a repeated measures ANOVA with the within subject factors pitch-size and its variant with reduced difference and Classifications of congruent stimuli (congruent trials rated as matching & congruent trials rated as not matching) was performed. This repeated-measures ANOVA showed a reliable effect for the factor Classifications of congruent stimuli (F$_{(1,28)}$=216.87, *p*<.001, *hp2*=0.89; Table 2). This means that congruent pairs were significantly classified as matching (M=87.4%, SEM=3.35%; Table 2). An interaction of Pitch-size and its variant with reduced difference × Classifications of congruent stimuli was statistically significant (F$_{(1,28)}$=5.072, *p*=0.033, *hp2*=0.152) and a *post-hoc* test revealed that significantly more congruent pairs were classified as matching within pitch-size (*pholm*<0.001) as well as within the pitch-size variant with reduced difference (*pholm*<0.001). However, there was no significant difference observed between congruent stimuli classified as matching for pitch-size and its variant with reduced difference (*pholm*=0.066). No main effect was observed for the stimulus classifications of pitch-size and its variant with reduced difference (F$_{(1,28)}$=1.699e^{-13}.0, *p*=1.0, *hp2*=6.068e^{-15}). This means that no significant difference was observed between pitch-size and its variant with reduced difference for the classification of congruent stimuli.

To test whether incongruent stimulus pairs were significantly classified as not matching by the participants (Table 2) and to test whether this classification is different between the pitch-size CMC and its variant with reduced difference, a repeated measures ANOVA with pitch-size and its variant with reduced difference and Classifications of incongruent stimuli (incongruent trials rated as matching & incongruent trials rated as not matching) as within subject factors was performed. This repeated measures ANOVA showed an effect for Classifications of incongruent stimuli (F$_{(1,28)}$=33.3, *p*<.001, *hp2*=0.54). This means that incongruent pairs were significantly classified as not matching (M=76.7%, SEM=5.4%; Table 2). No effect was observed for pitch-size CMC and its variant with reduced difference ($F_{(1,28)}$=-5.634e^{-15}, *p*=1.0, *hp2*=-2.012^{-16}) as well as for the interaction of pitch-size CMC and its variant with reduced difference × Classifications of incongruent stimuli ($F_{(1,28)}$=0.01, *p*=0.92, *hp2*=3.55e^{-4}). This means that no significant difference was observed

Table 2. Congruence classification of pitch-size and its variant with reduced difference for congruent and incongruent stimulus presentations. Mean in percent for stimulus conditions (congruent, incongruent) within pitch-size and pitch-size with reduced difference classified as matching, i.e. congruent or not matching, i.e. incongruent. The maximum possible percentage for each presented condition (congruent; incongruent) in each CMC was 100% for classifications of matching and not matching taken together.

Condition presented	Pitch-size classified as		Pitch-size variant classified as	
	matching	not matching	matching	not matching
congruent	92	8	82.8	17.2
incongruent	23.6	76.4	23	77

We discarded 2.25% of trials due to false or missing responses and 0.9% of trials due to slow button presses (over 1000 ms).

Error rates and reaction time data of the in scanner task for the two distinct CMCs

The error rates for pitch-size were 3.27% and for pitch-elevation 3.3%. To test whether there is a difference in RTs depending on Conditions (congruent trials & incongruent trials) and the tested distinct CMCs (pitch-size & pitch-elevation) we performed a repeated-measures ANOVA with the distinct CMCs and the two conditions as within subject factors on RTs. This ANOVA showed a main effect for CMC Types ($F_{(1,28)}$=12.11, *p*=0.002, *hp2*=0.302). The RTs in the pitch-elevation CMC were significantly lower (456.3 ms; SEM=12.1 ms) compared the RTs in Pitch-size (492 ms; SEM=9.5 ms). No main effect was observed for Condition ($F_{(1,28)}$=0.044, *p*=0.836, *hp2*=0.002) as well as for the interaction of distinct CMCs × Condition ($F_{(1,28)}$=0.055, *p*=0.817, *hp2*=0.002).

between pitch-size and its variant with reduced difference for the classification of incongruent stimuli.

Behavioral results of the in scanner-task recorded reaction time data

In summary, a significant difference in RTs was observed between the distinct CMCs, indicating a probable difference in processing the stimulus pairs dependent on the tested CMC type.

Error rates and reaction time data of the in scanner task for the pitch-size CMC and its variant

The error rates for pitch-size were 3.27% and for pitch-size with reduced difference were 2.95%. To test whether there is a difference in RTs depending on Conditions (congruent trials & incongruent trials) and the tested CMCs pitch-size and its variant with reduced difference, we performed a repeated-measures ANOVA with the pitch-size and its variant with reduced difference and the two conditions as within subject factors on RTs. This repeated-measures ANOVA showed a main effect for pitch-size and its variant with reduced difference ($F_{(1,28)}$=26.4, *p*<.001, *hp2=0.*485). RTs in the pitch-size CMC (492 ms; SEM=9.5 ms) were significantly lower compared to the pitch-size variant with reduced difference (526.5 ms; SEM=9.8 ms). The main effect for Condition ($F_{(1,28)}$=1.025, *p*=0.32, *hp2*=0.04) as well as for the interaction of pitch-size and its variant with reduced difference × Condition ($F_{(1,28)}$=1.25, *p*=0.27, *hp2=0.*04) were not significant. In summary, a significant difference in RTs was observed between pitch-size and its variant with reduced difference, indicating a probable difference in processing the stimulus pairs dependent on the tested pitch-size CMC.

Functional data

We were interested in finding the neural components of different CMC types as well as between pitch-size and its variant with reduced difference to evaluate different and common neural mechanisms underlying pitch-based CMC effects. To test for congruence effects in dependence of the CMCs pitch-size and pitch-elevation, interaction analyses were performed. To test whether there were common effects for congruent trials between distinct CMCs as well as between pitch-size and its variant pitch-size with reduced difference, global conjunction analyses (k>0) were performed.

Differences between pitch-size and pitch-elevation crossmodal correspondences

We did not find common congruence (C>I) effects for pitch-size and pitch-elevation CMCs when a global conjunction was performed. Albeit clear differences between the distinct CMCs were observed when we tested for enhanced neural effects of congruent trials selectively in the pitch-size CMC. The computed interaction for pitch-size (C>I) > pitch-elevation (C<I) revealed whole brain family wise error (FWE; *p*<0.05) corrected activations within the left superior parietal lobule (SPL; MNI coordinates: x=-26, y=-48, z=62; T=5.39, p=0.021), the left Heschls' gyrus (HG; MNI coordinates: x=-32, y=-24, z=6; T=5.33, p=0.026) and the left cerebellum (MNI coordinates: x=-22, y=-58, z=-30; T=5.22, p=0.041; Fig. 2). We observed greater neural effects for congruent trials (C>I) in pitch-size and greater neural effects for incongruent trials (C<I) in the pitch-elevation CMCs in these regions (see Fig. 2 as well as Table 3 and Table 4).

Additionally, small volume, FWE (*p*<0.05) corrected effects in our ROIs 10 mm spheres were observed within the right AnG (MNI coordinates: x=40, y=-66, z=32; T=3.7, *p*=0.03), left ACC (MNI coordinates: x=-2, y=22, z=34; T=4.13, p=0.008), right ACC (MNI coordinates: x=8, y=28, z=40; T=4.56, p=0.002) and nearly significant in the left IPS (MNI coordinates: x=-42, y=-36, z=46; T=3.51, p=0.05; Fig. 3 and Table 3). No activation of the IFG survived with the small volume correction.

The effect within the right AnG was dominated by a positive CMC effect within pitch-size (C>I; T=3.94, p=0.014 corr.). The measured effects within the right ACC were significantly greater for congruent trials in pitch-size (C>I; T=3.97, p=0.013 corr., Table 4) and for incongruent trials in pitch-elevation (C<I; T=3.61, p=0.038 corr., Table 4). The elicited effect in the IPS was driven by incongruent trials within the presentation of pitch-elevation (C<I; T=4.06, *p*=0.01 corr.; see Fig. 3). For the sake of completeness, we have included a table presenting the statistically significant results of the small volume corrected regions of interest for both pitch-size and pitch-elevation CMCs of both conditions (C>I, C<I; Table 4).

To evaluate a possible influence of RT differences between the pitch-size and pitch-elevation CMCs on the neural activity, we replicated the interaction analysis for pitch-size (C>I) > pitch-elevation (C<I) by including individual median RT data for each condition as a parametric modulator in the group level analysis. The results showed no substantial changes in activity neither in the whole brain nor the ROIs. All reported coordinates in the interaction analysis remained the same and the *p*-value remained at *p*<0.05 for the SPL (T=5.36), cerebellum (T=5.20), HG (T=5.29), right ACC (T=5.57), left ACC (T=4.11) and AnG (T=3.69). The neural activity found in the left IPS (T=3.46) stayed at a nearly significant level (*p*=0.058). No activation of the IFG survived with the small volume correction.

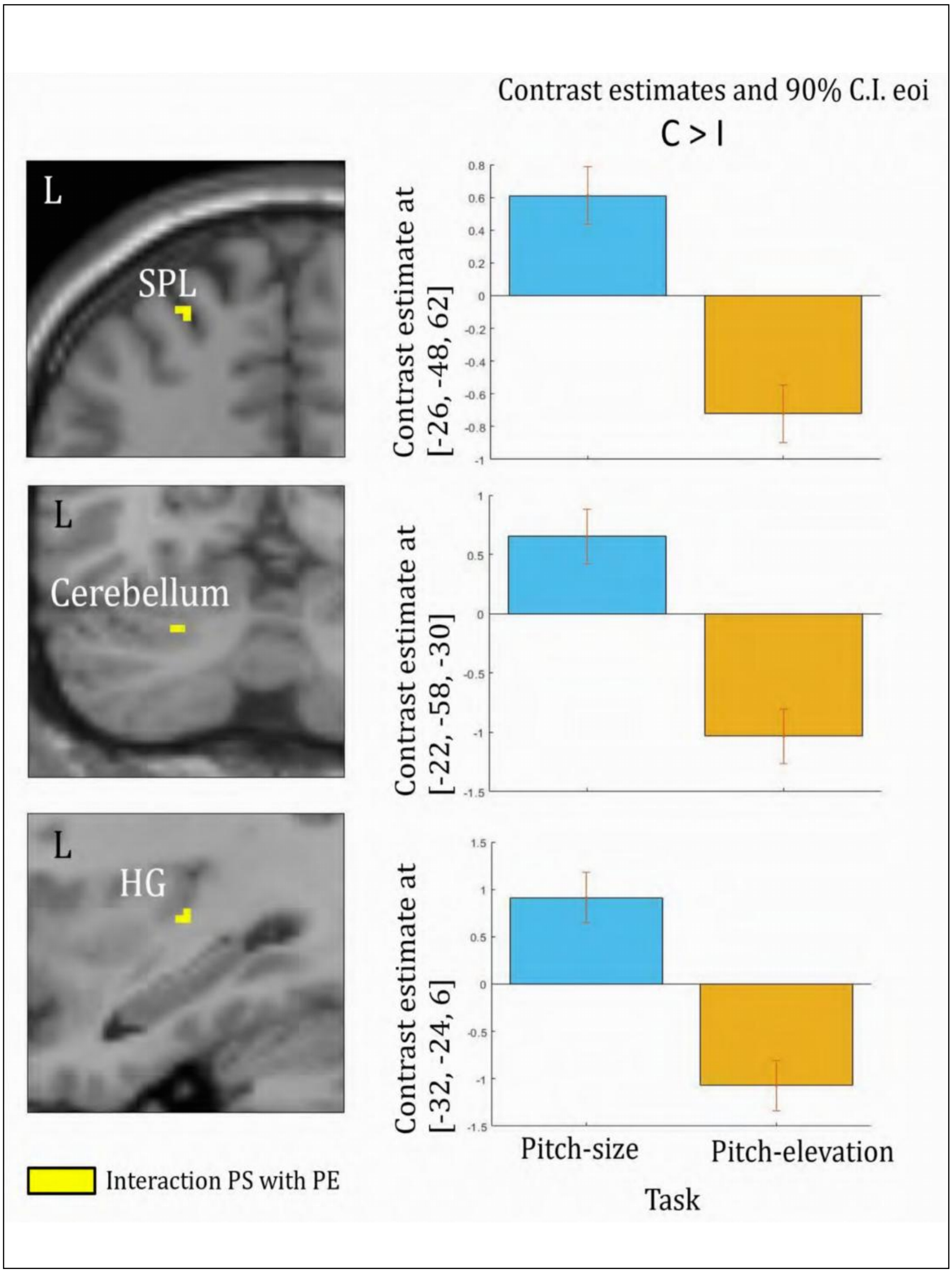


Fig. 2. Effects of the interaction for pitch- size with pitch- elevation (left), which tested for enhanced neural effects of congruent trials selectively in the pitch- size CMC. Contrast estimates for the corresponding region are displayed on the right (blue = Pitch- size; orange = Pitch- elevation). Each bar resembles the activation difference (C>I) within the specific region. Whole- brain FWE (peak- level; *p*<0.05) corrected effects were observed in the SPL, cerebellum and HG. Error bars indicate the observed standard error within each contrast. (The FWE corrected analysis with *p*<0.05 was used to create the images above).

Table 3. fMRI effects for an interaction between pitch-size and pitch-elevation. T Interaction effects within a) whole brain FWE corrected (*p*<0.05) peak-level effects and b) small volume, FWE corrected (*p*<0.05*)* peak-level effects within the defined ROIs. Statistically significant effects are shown and therefore no results for the interaction testing for enhanced neural effects of congruent trials selectively in pitch-elevation (pitch-size (C<I) < pitch-elevation (C>I)) are included in this table.

Pitch-size (C>I) > Pitch-elevation (C<I)

			MNI coordinates		
Hem	Regions	T-value	x	y	z
a)					
L	Superior parietal lobule	5.39	-26	-48	62
L	Heschels' gyrus	5.33	-32	-24	6
L	Cerebellum	5.22	-22	-58	-30
b)					
R	Anterior cingulate cortex	4.56	8	28	40
L	Anterior cingulate cortex	4.13	-2	22	34
R	Angular gyrus	3.70	40	-66	32
L	Intraparietal sulcus	3.51	-42	-36	46

Table 4. Condition specific fMRI findings for pitch- size and pitch- elevation. Findings of the small volume, FWE (*p<0.05*) corrected ROIs peak- level effects for pitch- size and pitch- elevation by condition (C>I; C<I). Statistically significant effects are shown and therefore no results for pitch- size C<I and pitch- elevation C>I are included in the table.

Pitch- size

Hem	Regions	T- value	MNI coordinates		
			x	y	z
C>I					
R	Anterior cingulate cortex	3.97	8	36	32
R	Angular gyrus	3.94	40	- 68	32

Pitch- elevation

Hem	Regions	T- value	MNI coordinates		
			x	y	z
C<I					
R	Anterior cingulate cortex	3.61	6	20	38
R	Anterior cingulate cortex	3.60	4	30	40
L	Intraparietal sulcus	4.06	- 42	- 36	44

These results confirm that the interaction effects between the pitch-size and pitch-elevation CMCs are independent of the observed RT differences.

Common effects for pitch-size and pitch-size with reduced difference correspondences

We were interested in to what extent a reduced difference between the visual and acoustic stimuli could modulate a CMC effect for pitch and size. No common effect was found between pitch-size and its variant with reduced difference. This means that the global conjunction over pitch-size (C>I) & pitch-size variant with reduced difference (C>I) did not lead to significant effects within the whole brain or our ROIs. We therefore concluded that there was likely a difference in neural effects between pitch-size and its variant with reduced difference. Subsequently, we examined whether there were any differences in our ROIs between pitch-size and its variant with reduced difference by conducting an interaction analysis that tested for enhanced neural effects of congruent trials selectively in pitch-size (pitch-size (C>I) > pitch-size variant with reduced difference (C<I)). A differential effect in the ACC (MNI co-ordinates: x=-2, y=30, z=26) with a T value of 3.75 was observed. However, it did not survive corrections for multiple comparisons.

DISCUSSION

In this study, the neural correlates between different CMCs were compared to evaluate a possible common or different processing of CMCs and further to obtain a possible clue regarding the theories on the origins of CMCs. Our results are in favor of different processes underlying different pitch-based CMCs as well as in favor of the theory that these CMCs are probably driven by statistical regularities from our environment. The participants significantly classified the congruent stimulus pairs as matching compared to not matching and the incongruent stimulus pairs as not matching compared to matching in both distinct CMCs as well as the variant of pitch-size with reduced difference in the stimulus congruence classification, which was performed outside the scanner. This means that the participants classified the stimulus pairs in accordance with the CMC theory (Fig. 1A, B).

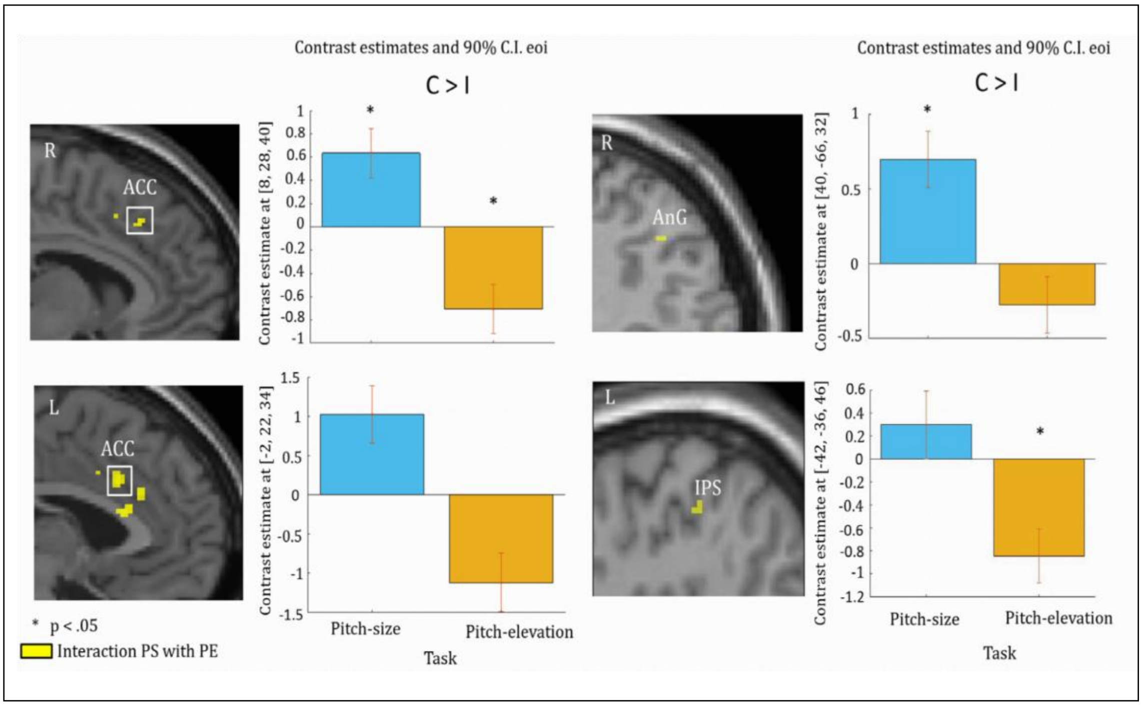


Figure 3. Effects of the interaction for pitch-size with pitch-elevation (left), which tested for enhanced neural effects of congruent trials selectively in the pitch-size CMC. Contrast estimates for the corresponding region are displayed on the right (blue = Pitch-size; orange = Pitch-elevation). Each bar resembles the activation difference (C > I) within the specific region. FWE corrected small volume effects (peak-level; $p < .05$) were observed in the ACC, AnG and IPS. Error bars indicate the observed standard error within each contrast. (For display purposes, a T value of > 4 was chosen to create the images above.)

Common and different regions are involved in processing different crossmodal correspondences

Overall, no area was observed to show common effects for congruence across the distinct CMCs within this study. This finding points towards probable differences in the underlying information processing of different CMCs. Differential effects were observed in the SPL, cerebellum, HG, ACC, AnG and IPS. In all regions, the effects were different between the conditions of the two CMCs. For the pitch-size CMC, a greater activation was observed for congruent > incongruent stimulus presentations while enhanced activity was seen for congruent<incongruent stimulus pairs in the pitch-elevation CMC within identical areas. However, the stimulus congruence classification task showed clear behavioral effects, supporting the assumption that congruent stimulus presentations are overall perceived as belonging together. This finding is consistent with the possibility that the neural effects are related to the perceived congruency between the visual and acoustic stimuli in the different CMCs. Therefore, the disparate neural effects can be interpreted as correlates of different neural processes underlying the different CMCs.

The processing and integration of multimodal information is assumed to start already in the primary sensory system (Baier et al., 2006; Crottaz-Herbette & Menon, 2006; Werner & Noppeney, 2010). This likely explains our observation of effects within the HG in both CMCs. The HG, which is part of the primary acoustic system, demonstrated a modulation of acoustic processing based on the congruency of the visual stimulus in CMCs. Despite the fact that our visual stimuli were clear and informative on their own, we found that processing of the visual stimulus was clearly influenced by the acoustic stimulus (Gallace & Spence, 2006; McDonald et al., 2000; Misselhorn et al., 2016; Werner & Noppeney, 2010). This is particularly interesting since the acoustic stimulus is irrelevant in all tested CMCs. On the one hand acoustic processing was previously reported as suppressed in visually dominating tasks (Johnson & Zatorre, 2005; Schmid et al., 2011) on the other hand, the influential nature of irrelevant acoustic stimuli in visual information processing was also reported by other studies (McDonald et al., 2000; Regenbogen et al., 2018; Tonelli et al., 2017). Thus, our results are a likely indication that in CMCs, differential processing of congruent and incongruent stimuli already occurs in the early sensory systems. To our surprise, we found differential processing of congruent and incongruent stimulus combinations in the early sensory system for the different CMCs. Greater effects in the HG were found for congruent stimulus presentations in pitch-size and for incongruent stimulus presentations in pitch-elevation (Fig. 2).

The ACC was also involved in both CMCs, which probably reflects the different functions of this area in the dynamical interplay between top-down and bottom-up multisensory processing (Benedict et al., 2002; Crottaz-Herbette & Menon, 2006; Downar et al., 2002; Laurienti et al., 2003). In particular, the role of the AnG in multisensory integration is well established (Bonnici et al., 2016; Cabeza et al., 2012; Hölig et al., 2017). Further, the AnG is assumed to be crucial for the direction of attention in relation to memory (Cabeza et al., 2012; Humphreys & Ralph, 2015; Jablonowski & Rose, 2022; Regenbogen et al., 2018). The involvement of this area can be interpreted as recruitment of memory-based attention in the pitch-size CMC. We speculate that pitch-size congruent trials were mainly driven by bottom-up processes, as the AnG is part of the ventral parietal cortex (VPC) (Cabeza et al., 2012; Humphreys & Ralph, 2015; Kim, 2010). According to the attention to memory (AtoM) model postulated by Cabeza and colleagues, pitch-size congruencies can be assumed to be driven by a detection of cues as part of a memory based retrieval (Cabeza et al., 2011; 2012). According to Cabeza and colleagues, the VPC *'[…] mediates the bottom-up capture of attention by salient memory contents (bottom-up AtoM)'* (Cabeza et al., 2012, p. 6).

In contrast to pitch and size correspondences, top-down processes seem to be involved in pitch-elevation CMCs. The IPS was postulated to be part of top-down attentional processes as this region is part of the dorsal parietal cortex (Cabeza et al., 2011; 2012; Regenbogen et al., 2018). Furthermore, the IPS showed greater activations dependent on task difficulty (Regenbogen et al., 2018). As the information about the location of a visual stimulus can be assumed to be biased by the misleading acoustic stimulus within incongruent pitch-elevation trials, top-down controlled attention shifts were probably enforced within incongruent stimulus presentations in the pitch-elevation CMC (Chiou & Rich, 2012; Regenbogen et al., 2018; Salmi et al., 2009). On a theoretical level, the fMRI results supported the theory about statistical assumptions extracted from our environment as a general basis for different CMCs (Maimon et al., 2021; Pisanski et al., 2017; Spence, 2011, 2020; Spence & Sathian, 2020; Zeljko et al., 2019). This would explain the facilitation of memory driven processes, which seem to best fit the measured neural effects. As shown in other studies before, there seems to be a strong connection between pitches with certain sizes or pitches and the spatial location of objects (Ben-Artzi & Marks, 1995; Bolam et al., 2022; Evans & Treisman, 2010; Jamal et al., 2017; Maimon et al., 2021; Pisanski et al., 2017; Tonelli et al., 2017). These connections are probably based on experiences from our daily lives (Bowling et al., 2017; Parise et al., 2014; Pisanski et al., 2017). However, this common basis result in differential neural processing in the two distinct CMCs. We can only speculate at this point that our task design allowed for an easier detection of congruent pitch-size presentations. The spatially differentiating incongruent pitch-elevation pairs on the other hand probably resulted in a confounded response detection processing (Bruns et al., 2014; Chiou & Rich, 2012; Maimon et al., 2021). Differences in processing the stimulus combinations in the two distinct CMCs may explain the lack of significant effects observed for the incongruent pitch-size and congruent pitch-elevation presentations, as well as the interaction testing for enhanced neural effects of congruent trials selectively in the pitch-elevation CMC. However, further studies are required to investigate these differences and clarify the findings.

Findings on language and magnitude driven effects in crossmodal correspondences

No statistically significant activation in our language specific ROIs was found within this study. Based on the findings from this study as well as previous research, the theories stating that pitch-elevation correspondences are driven by language, are not supported (McCormick et al., 2018; Parkinson et al., 2012). However, it is important to note that we cannot rule out any entanglements of language with other factors influencing the mechanisms underlying pitch and elevation correspondences.

When examining the theory of magnitude-driven correspondences, the effect within the left, instead of the proposed right, IPS exhibited nearly significant results in the interaction between pitch-size and pitch-elevation. Surprisingly, the effect showed significant results in relation to incongruent presentations when assessing the ROI in the main effect of pitch and elevation, contradicting our hypothesis. Hence, our data does not strongly support the existence of a shared foundation for correspondence effects driven by magnitude within or across different types of CMCs.

According to the theory of magnitude/intensity, an e.g., small visual stimulus would be located on the same side of a polar scale as a high-pitched tone on its respective scale. Thus, leading to an intrinsic matching of congruent pairs as they would be perceived as ‘more’ intense or ‘greater’ in magnitude when combined (Chang & Cho, 2015). Hence, incongruent pairs of audio-visual stimuli would be on their opposite sides of the polar scales and therefore perceived as less intense or lower in magnitude. While our findings in investigating magnitude and language effects in our chosen ROIs may be open to debate, they are in line with the relatively weak effects reported by McCormick et al. (2018) in their study on pitch-elevation correspondence.

Linking this study to a former fMRI study on pitch- elevation CMCs

It is important to note that McCormick et al. (2018) used a working memory task in their study on pitch-elevation correspondences, which might has obscured direct correspondence effects leading to no significant effects in a congruent versus incongruent fMRI contrast. Nevertheless, notable effects were observed when analyzing consecutive congruent versus consecutive incongruent trials. Even though we have to interpret these findings with caution, in regard of the magnitude and language hypothesis, both of our findings share some commonalities in terms of null or weak effects. On the other hand, we found similar effects in the right AnG, however probably originating in the pitch-size correspondence in our study. Rather than directly comparing or linking our outcomes to the study by McCormick et al. (2018), our aim was to start from a similar theoretical framework and employ similar ROIs to explore the congruence effect in different CMCs with a non-working memory related task design.

No effects of pitch and size and its variant with reduced difference

We hypothesized to find a modulation of effect strength related to the reduced difference of the contrast of pitches and sizes in our pitch-size CMCs. This hypothesis was not supported by the measured effects within this study. We did not observe common effects of congruence between pitch-size and its variant with reduced difference within our ROIs. According to the findings of a study by Chiou & Rich (2012), the mapping of acoustic and visual stimuli seems to be relative to the assignments of low and high pitches to e.g., small and large squares, however the ability to measure a CMC effect is dependent on the ability to form distinct pairs which clearly stand out from each other in the context they are presented in (Chiou & Rich, 2012; Zeljko et al., 2019). The context in which the stimuli are presented in is thought to support the formation of congruent mappings of stimuli (Chiou & Rich, 2012; Zeljko et al., 2019). The findings in our congruence classification task led to the assumption that a reduced difference of the contrast of our implemented stimuli did not significantly weaken the mentioned ability to associate the expected stimulus pairs as congruent or incongruent. We observed a differential activity in the ACC of pitch-size compared to its variant. This effect was observed when we performed an interaction analysis that tested for enhanced neural effects of congruent trials selectively in the pitch-size CMC. However, this activity did not survive multiple comparison corrections. Nevertheless, this observation suggests that differences in processing congruency may be involved in pitch-size compared to its variant. The slowed down RTs in the pitch-size variant with reduced difference compared to the pitch-size pairs with a great difference in stimulus contrasts probably hints towards a slowed down decision process. It remains speculative whether a lack in neural modulatory effects was caused by the great similarity of the two pitch-size CMCs, as well as the higher similarity of stimuli within the variant of the pitch-size with reduced difference.

Limitations of this study

In the present study are some limitations we want to address. Instead of performing the CMC presentations in an upright position like in classical behavioral measurements, the participants lay on their back during the main experiment in our study. The perceived upright should not have inferred with the pitch-elevation matching of congruence or not matching of incongruence as the presented visual and acoustic stimuli were aligned with the body position (Harris et al., 2015), nevertheless we cannot exclude a probable influential factor of the body position on the perception of stimulus positions within the pitch-elevation part of the experiment in the scanner. While we decided to use natural stimuli instead of pure tones, which allowed for a more natural representation of crossmodal correspondences, it is worth acknowledging the potential influence of timbre on our findings. We matched the instruments and their produced tones, i.e., higher tones are played by smaller instruments and lower tones are played by larger instruments within our distinct CMCs and the pitch-size variant with reduced difference. This alignment potentially amplified the correspondence effect observed in the pitch-size CMC. Previous studies, such as Evans and Treisman (2010), successfully utilized piano and violin sounds in their indirect crossmodal correspondence tasks without encountering issues related to the instruments used. However, it is worth noting that while there are studies exploring crossmodal correspondences involving timbre and other perceptual dimensions (Adeli et al., 2014; Qi et al., 2020; Wallmark & Allen, 2020), to our knowledge, none have specifically investigated the role of timbre in pitch-size or pitch-elevation correspondences using the stimuli employed in our study. While we believe our results provide valuable insights into the mechanisms hidden behind different crossmodal correspondences, it would be beneficial to replicate our study with an alternative experimental setup to ascertain the robustness of our conclusions. It would be for example from great interest if the same results we find in this study can be replicated when utilizing the same task for e.g., different crossmodal correspondences. Regarding the null effects we observed in our study, it is important to emphasize that the absence of significant findings does not necessarily imply the absence of an effect. Further investigation with a different paradigm may provide a better understanding of the effect. Although CMCs have been successfully studied using an implicit task design in behavioral studies (Evans, 2020; Evans & Treisman, 2010; Parise & Spence, 2012, Chiou & Rich, 2012), exploring differences between the processing of CMCs in implicit and explicit task designs may be of interest for future fMRI studies. For example, a recent study on sound-symbolic CMCs (Barany et al., 2023) showed that significant congruence effects occur in an explicit design, whereas an implicit task design led to significant incongruence effects in sound-symbolic CMCs (Mccormick et al., 2021; Peiffer-Smadja & Cohen, 2019). In future studies of pitch-based CMCs, it would also be worthwhile to test other stimuli, increase the number of trials, or expand the sample size to gain a more comprehensive understanding of the neural mechanisms involved in CMC effects.

CONCLUSION

This study aimed to elucidate the neural processing of different pitch-based correspondences. The results within our study argue in favor of the idea that different mechanisms drive the integration of stimulus features within different audio-visual CMCs. The strong matching of pitches with their congruent spatial location probably led to greater top-down attentional driven processes during incongruent stimulus presentations. On the other hand, attention as well as memory retrieval seem to be crucial for pitch-size correspondences. Our results support the findings of previous studies assuming both top-down and bottom-up processes are involved in forming the CMC effect in particular and multimodal integration in general (Getz & Kubovy, 2018; Salmi et al., 2009).

ACKNOWLEDGMENTS

The present research was supported by the German Research Foundation (DFG; SFB/Transregio 169, Project B3). The authors would like to thank Marike Maack for her helpful comments on the manuscript.

M.R. and C.J. designed the study. C.J. collected the data, C.J. and M.R. analyzed the data and C.J. wrote the manuscript under supervision of M.R. All authors reviewed the manuscript.

The datasets generated and/or analyzed during this study are available on request from the corresponding author.

REFERENCES

Adeli, M., Rouat, J., & Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00352

Baier, B., Kleinschmidt, A., Müller, N. G., & Muller, N. G. (2006). Cross- modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. *Journal of Neuroscience*, *26*(47), 12260–12265.

Barany, D. A., Lacey, S., Matthews, K. L., Nygaard, L. C., & Sathian, K. (2023). Neural basis of sound- symbolic pseudoword- shape correspondences. *Neuropsychologia*, *188*, 108657. https://doi.org/10.1016/j.neuropsychologia.2023.108657

Ben- Artzi, E., & Marks, L. E. (1995). Visual- auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, *57*(8), 1151–1162.

Ben- Artzi, E., & Marks, L. E. (1999). Processing linguistic and perceptual dimensions of speech: Interactions in speeded classification. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 579–595.

Benedict, R. H. B., Shucard, D. W., Maria, M. P. S., Shucard, J. L., Abara, J. P., Coad, M. Lou, Wack, D., Sawusch, J., & Lockwood, A. (2002). Covert Auditory Attention Generates Activation in the Rostral/Dorsal Anterior Cingulate Cortex. *Journal of Cognitive Neuroscience*, *14*(4), 637–645.

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size. Crossmodal binding in pitch- size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage*, *59*(1), 663–672. https://doi. org/10.1016/j.neuroimage.2011.06.095

Bolam, J., Boyle, S. C., Ince, R. A. A., & Delis, I. (2022). Neurocomputational mechanisms underlying cross- modal associations and their influence on perceptual decisions. *NeuroImage*, *247*, 1–32.

Bonetti, L., & Costa, M. (2018). Pitch- verticality and pitch- size cross- modal interactions. *Psychology of Music*, *46*(3), 340–356.

Bonnici, H. M., Richter, F. R., Yazar, Y., & Simons, J. S. (2016). Multimodal feature integration in the angular gyrus during episodic and semantic retrieval. *The Journal of Neuroscience,* 36(20), 5462–5471.

Bowling, D. L., Garcia, M., Dunn, J. C., Ruprecht, R., Stewart, A., & Frommolt, K. (2017). Body size and vocalization in primates and carnivores. *Scientific Reports*, *7*, 41070.

Bruns, P., Maiworm, M., & Röder, B. (2014). Reward expectation influences audiovisual spatial integration. *Attention, Perception & Psychophysics*, *76*(6), 1815–1827. https://doi.org/10.3758/s13414- 014- 0699- y

Cabeza, R., Mazuz, Y. S., Stokes, J., Kragel, J. E., Woldorff, M. G., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2011). Overlapping parietal activity in memory and perception: evidence for the attention to memory model. *Journal of cognitive neuroscience*, *23*(11), 3209–3217. https://doi. org/10.1162/jocn_a_00065

Cabeza, R., Ciaramelli, E., & Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends Cogn Sci.*, *16*(6), 338–352. https://doi.org/10.1016/ j.tics.2012.04.008.

Chang, S., & Cho, Y. S. (2015). Polarity correspondence effect between loudness and lateralized response set. *Frontiers in Psychology*, *6*, 683. https://doi.org/10.3389/fpsyg.2015.00683

Chiou, R., & Rich, A. N. (2012). Cross- modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, *41*(3) , 339–353. https://doi.org/10.1068/p 7161

Crottaz- Herbette, S; Menon, V. (2006). Where and when the anterior cingulate cortex modulates attentional response : Combined fMRI and ERP Evidence. *Journal of Cognitive Neuroscience*, *18*(5) , 766–780.

Downar, J., Crawley, A. P., Mikulis, D. J., Davis, K. D., Crawley, A. P., & Mikulis, D. J. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *Journal of Psychology*, *87*, 615–620.

Evans, K. K. (2020). The role of selective attention in cross- modal interactions between auditory and visual features. *Cognition*, *196*, 104119. https://doi.org/10.1016/j.cognition.2019.104119

Evans, K. K., & Treisman, A. (2010). Natural cross- modal mappings between visual and auditory features. *Journal of Vision*, *10*, 1–12. https://doi. org/10.1167/10.1.6

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception and Psychophysics*, *68*(7) , 1191–1203. https://doi.org/10.3758/BF 03193720

Getz, L. M., & Kubovy, M. (2018). Questioning the automaticity of audiovisual correspondences. *Cognition*, *175*, 101–108. https://doi.org/10.1016/ j.cognition.2018.02.015

Harris, L. R., Carnevale, M. J., D'Amour, S., Fraser, L. E., Harrar, V., Hoover, A. E. N., Mander, C., & Pritchett, L. M. (2015). How our body influences our perception of the world. *Frontiers in Psychology*, *6*, 1–10. https://doi.org/10.3389/fpsyg.2015.00819

Hölig, C., Föcker, J., Best, A., Röder, B., & Büchel, C. (2017). Activation in the angular gyrus and in the pSTS is modulated by face primes during voice recognition. *Human Brain Mapping*, *2565*, 2553–2565. https://doi. org/10.1002/hbm.23540

Humphreys, G. F., & Ralph, M. A. L. (2015). Fusion and fission of cognitive functions in the human parietal cortex. *Cerebral Cortex*, *25*, 3547–3560. https://doi.org/10.1093/cercor/bhu198

Jablonowski, J., & Rose, M. (2022). The functional dissociation of posterior parietal regions during multimodal memory formation. *Human Brain Mapping, 43*(11), 3469- 3485 . https://doi.org/10.1002/hbm. 25861

Jamal, Y., Lacey, S., Nygaard, L., & Sathian, K. (2017). Interactions between auditory elevation, auditory pitch and visual elevation during multisensory perception. *Multisensory Research*, *30*(3–5), 287–306. https://doi. org/10.1163/22134808- 00002553

Johnson, J. A., & Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events : Behavioral and neural correlates. *Cerebral Cortex*, *15*, 1609–1620. https://doi.org/10.1093/cercor/bhi039

Kim, H. (2010). NeuroImage Dissociating the roles of the default- mode, dorsal, and ventral networks in episodic memory retrieval. *NeuroImage*, *50*(4) , 1648–1657. https://doi.org/10.1016/j.neuroimage. 2010.01.051

Koten, J. W., Langner, R., Wood, G., & Willmes, K. (2013). Are reaction times obtained during fMRI scanning reliable and valid measures of behavior? *Experimental Brain Research*, *227*(1), 93–100. https://doi.org/10.1007/ s00221- 013- 3488- 2

Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E., & Burdette, J. H. (2003). Cross- modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Human Brain Mapping*, *19*(4), 213–223 . https://doi.org/10.1002/hbm. 10112

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3) , 1455–1468. https:// doi.org/10.1121/1.426686

Liuzzi, A. G., Bruffaerts, R., Peeters, R., Adamczuk, K., Keuleers, E., De Deyne, S., Storms, G., Dupont, P., & Vandenberghe, R. (2017). Cross- modal representation of spoken and written word meaning in left pars triangularis. *NeuroImage*, *150*, 292–307. https://doi.org/10.1016/ j.neuroimage.2017.02.032

Maimon, N. B., Lamy, D., & Eitan, Z. (2021). Space oddity: musical syntax is mapped onto visual space. *Scientific Reports*, *11*(1), 1–17. https://doi. org/10.1038/s41598- 021- 01393- 1

Marks, L. E. (1987). On Cross- Modal Similarity: Auditory- visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384–394. https://doi. org/10.1037/0096- 1523.13.3.384

McCormick, K., Lacey, S., Stilla, R., Nygaard, L. C., & Sathian, K. (2018). Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia*, *112*, 19–30. https://doi. org/10.1016/j.neuropsychologia.2018.02.029

Mccormick, K., Lacey, S., Stilla, R., & Nygaard, L. C. (2021). Neural Basis of the Sound- Symbolic Crossmodal Correspondence Between Auditory Pseudowords and Visual Shapes. https://doi. org/10.1163/22134808- bja10060

McDonald, John J; Teder- Sälejärvi, Wolfgang A., Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*, 906–908. https://doi.org/10.1038/35038085

Melara, R. D., & Brien, T. P. O. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology*, *116*(4), 323–336.

Misselhorn, J., Daume, J., Engel, A. K., & Friese, U. (2016). A matter of attention : Crossmodal congruence enhances and impairs performance in a novel trimodal matching paradigm. *Neuropsychologia*, *88*, 113–122. https://doi.org/10.1016/j.neuropsychologia.2015.07.022

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, *220*(3–4), 319–333. https://doi.org/10.1007/ s00221- 012- 3140- 6

Parise, C. V, Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(16), 6104–6108. https:// doi.org/10.1073/pnas.1322705111

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception*, *41*(7), 854–861. https://doi.org/10.1068/p7225

Peiffer- Smadja, N., & Cohen, L. (2019). The cerebral bases of the bouba- kiki effect. *NeuroImage*, 186, 679–689. https://doi.org/10.1016/j.neuroimage.2018.11.033

Piazza, M., Pinel, P., Bihan, D. Le, & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*, 293–305. https://doi.org/10.1016/ j.neuron.2006.11.022

Pinel, P., Piazza, M., Bihan, D. Le, Dehaene, S., Unit, I., & Ge, P. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, *41*, 983–993.

Pisanski, K., Isenstein, S. G. E., Montano, K. J., O'Connor, J. J. M., & Feinberg, D. R. (2017). Low is large: spatial location and pitch interact in voice- based body size estimation. *Attention, Perception, and Psychophysics*, *79*(4), 1239–1251. https://doi.org/10.3758/s13414- 016- 1273- 6

Qi, Y., Huang, F., Li, Z., & Wan, X. (2020). Crossmodal correspondences in the sounds of Chinese instruments. *Perception*, *49*(1) , 81–97. https://doi. org/10.1177/0301006619888992

Regenbogen, C., Seubert, J., Johansson, E., Finkelmeyer, A., Andersson, P., & Lundström, J. N. (2018). The intraparietal sulcus governs multisensory integration of audiovisual information based on task difficulty. *Human Brain Mapping*, *39*(3), 1313–1326. https://doi.org/10.1002/ hbm.23918

Roelofs, A., van Turennout, M., & Coles, M. G. (2006). Anterior cingulate cortex activity can be independent of response conflict in Stroop- like tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(37), 13884–13889. https://doi.org/10.1073/ pnas.0606265103

Sadaghiani, S., Maier, J. X., & Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *Journal of Neuroscience*, *29*(20), 6490–6499. https://doi.org/10.1523/JNEUROSCI.5437- 08.2009

Salmi, J., Rinne, T., Koistinen, S., Salonen, O., & Alho, K. (2009). Brain networks of bottom- up triggered and top- down controlled shifting of auditory attention. *Brain Research*, *1286*, 155–164. https://doi.org/10.1016/ j.brainres.2009.06.083

Schmid, C., Büchel, C., & Rose, M. (2011). The neural basis of visual dominance in the context of audio- visual object processing. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2010.11.051

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, and Psychophysics* (Vol. 73, Issue 4, pp. 971–995). https:// doi.org/10.3758/s13414- 010- 0073- 7

Spence, C. (2020). Simple and complex crossmodal correspondences involving audition. *Acoustical Science and Technology*, *41*(1) , 6–12. https:// doi.org/10.1250/ast.41.6

Spence, C., & Sathian, K. (2020). Audiovisual crossmodal correspondences: behavioral consequences and neural underpinnings. In Sathian, K and Ramachandran, VS (Ed.), *Multisensory Perception: From Laboratory to Clinic* (pp. 239–258). https://doi.org/10.1016/B978- 0- 12- 812492- 5.00011- 5

Tonelli, A., Cuturi, L. F., & Gori, M. (2017). The influence of auditory information on visual size adaptation. *Frontiers in Neuroscience*, *11*(594) . https:// doi.org/10.3389/fnins.2017.00594

Uno, K., & Yokosawa, K. (2022a). Cross- modal correspondence between auditory pitch and visual elevation modulates audiovisual temporal recalibration. *Scientific Reports*, *12*(1), 1–10. https://doi.org/10.1038/ s41598- 022- 25614- 3

Uno, K., & Yokosawa, K. (2022b). Pitch- elevation and pitch- size cross- modal correspondences do not affect temporal ventriloquism. *Attention, Perception, and Psychophysics*, 84, 1052–1063. https://doi.org/10.3758/ s13414- 022- 02455- w

Wallmark, Z., & Allen, S. E. (2020). Preschoolers' crossmodal mappings of timbre. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/ s13414- 020- 02015- 0

Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *The Journal of Neuroscience*, *30*(7) , 2662–2675. https://doi.org/10.1523/JNEUROSCI.5091-09.2010

Zeljko, M., Kritikos, A., & Grove, P. M. (2019). Lightness/pitch and elevation/pitch crossmodal correspondences are low-level sensory effects. *Attention, Perception, and Psychophysics*, *81*(5), 1609–1623. https://doi.org/10.3758/s13414-019-01668-w

8. Summary/Zusammenfassung

Context defines and shapes the perception of sensory cues, memories of past events and the meaning of words. The classical view of context frames it as a rich set of information surrounding an event. Thereby a context entails a variety of different aspects. These aspects may be internal, such as physiological or emotional states, or external, for example spatial and temporal cues. Given that context is inherently broad, encompassing a wide range of different aspects, it seems fair to propose that there are certainly cognitive processes that are influenced by contextual factors that have been overlooked or are underrepresented in research so far. However, exploring the effects of such contextual factors on cognitive processes could help fill gaps in the literature and enhance our understanding of these processes. Therefore, the objective of this cumulative dissertation is to examine the potential influence of certain contextual aspects on these processes in three studies. In study 1, the probable influence of environmental contextual factors on the transition from implicit to explicit memory during an incidental sequential-association task was studied using virtual reality. Virtual reality was implemented to create an enriched and immersive experiment environment. In Study 2, the potential role of oscillatory activity (internal contextual factors) on subsequent facial emotion processing and negativity bias was studied using electroencephalography. In study 3, the neural processes of distinct crossmodal correspondence effects were studied using functional magnetic resonance imaging. In this last study, the role of cultural and environmental factors as drivers for the crossmodal correspondence effect was explored.

The first study revealed that differences in oscillatory activity, specifically in the alpha band, regarding latter choices of valences exist already previous to stimulus presentation. Furthermore, a positive correlation of negativity bias with this difference in pre-stimulus oscillatory activity was observed. This finding emphasizes the necessity of considering pre-stimulus neural activity, an internal contextual factor, to achieve a more comprehensive understanding of the underlying neural mechanisms.

The second study focused on the role of environmental contextual aspects on the transfer of implicit to explicit memory in a sequential learning task. Hereby, a greater amount of explicit memory when the task was performed in an enriched environment was observed, which underpins the crucial role of environmental contextual factors on cognitive processes such as learning.

In the third and final study, investigating the origin of distinct pitch-based crossmodal correspondences, the environmental aspect, especially the process of learning that certain stimulus combinations likely belong together, seems to be crucial in the process of acquiring such correspondences. On the other hand, we found that a cultural aspect, specifically language, might not be the main driver for these correspondences. Furthermore, differences in neural activity were observed for distinct crossmodal correspondences.

The results of this dissertation have shown that various contextual aspects can have a non-negligible influence on cognitive processes. The results of this dissertation enrich the existing knowledge in the different areas and show that the influence of contextual aspects on cognitive processes should be considered in future studies.

Kontext definiert und formt die Wahrnehmung von sensorischen Hinweisen, Erinnerungen an vergangene Ereignisse und die Bedeutung von Worten. Nach der klassischen Auffassung beinhaltet ein Kontext ein umfangreiches Set an Informationen, die ein Ereignis umgeben. Dabei umfasst ein Kontext eine Vielzahl unterschiedlicher Aspekte. Diese Aspekte können interner Natur sein, wie z. B. physiologische oder emotionale Zustände, oder externer, zum Beispiel räumlicher und zeitlicher Anhaltspunkte. In Anbetracht der Tatsache, dass ein Kontext von Natur aus breit gefächert ist und eine Vielzahl unterschiedlicher Aspekte umfasst, kann man davon ausgehen, dass es sicherlich kognitive Prozesse gibt, die von kontextuellen Faktoren beeinflusst werden, die bisher in der Forschung übersehen wurden oder unterrepräsentiert sind. Die Erforschung von solchen Kontextfaktoren auf bestimmte kognitive Prozesse könnte jedoch dazu beitragen, Lücken in der Literatur zu schließen und unser Verständnis für diese Prozesse zu verbessern. Ziel dieser kumulativen Dissertation ist es daher, den möglichen Einfluss verschiedener Kontextfaktoren auf kognitive Prozesse in drei Studien zu untersuchen. In Studie 1 wurde der mögliche Einfluss von räumlichen Kontextfaktoren auf den Übergang von implizitem zum expliziten Wissen während einer inzidentellen sequentiellen Assoziationsaufgabe mit Hilfe von virtueller Realität untersucht. Dabei wurde eine virtuelle Realität genutzt, um eine angereicherte und immersive Experimentalumgebung zu schaffen. In Studie 2 wurde die mögliche Rolle oszillatorischer Aktivität (interne Kontextfaktoren) auf die nachfolgende Verarbeitung von Gesichtsemotionen und den Negativitäts-Bias mittels Elektroenzephalographie untersucht. In Studie 3 wurden mit Hilfe der funktionellen Magnetresonanztomographie die neuronalen Prozesse verschiedener modalitätsübergreifender Korrespondenzeffekte untersucht. In dieser letzten Studie wurde die Rolle kultureller und umweltbedingter Faktoren als treibende Kräfte des modalitätsübergreifenden Korrespondenzeffekts untersucht.

In der ersten Studie konnte gezeigt werden, dass Unterschiede in der oszillatorischen Aktivität, insbesondere im Alpha-Band, in Abhängigkeit von der Wahl der Valenz bereits vor der Reizdarbietung bestehen. Darüber hinaus wurde eine positive Korrelation des Negativitätsbias mit diesem Unterschied in der oszillatorischen Aktivität vor dem Stimulus beobachtet. Dieses Ergebnis unterstreicht die Notwendigkeit, die neuronale Aktivität vor dem Stimulus, welche ein interner Kontextfaktor ist, zu berücksichtigen, um ein umfassenderes Verständnis der zugrunde liegenden neuronalen Mechanismen zu erlangen.

Die zweite Studie konzentrierte sich auf die Rolle von kontextuellen Aspekten der Umgebung beim Übergang von implizitem auf expliziten Wissen in einer sequentiellen Lernaufgabe. Hierbei wurde eine größere Menge an explizitem Gedächtnis beobachtet, wenn die Aufgabe in einer angereicherten Umgebung durchgeführt wurde, was die entscheidende Rolle von kontextuellen Umgebungsfaktoren für kognitive Prozesse wie das Lernen unterstreicht.

In der dritten und letzten Studie, in der der Ursprung ausgeprägter tonhöhenbasierter modalitätsübergreifender Korrespondenzen untersucht wurde, scheint der Umweltaspekt, insbesondere der Prozess des Lernens, dass bestimmte Reizkombinationen wahrscheinlich zusammengehören, entscheidend für den Prozess des Erwerbs solcher Korrespondenzen zu sein. Andererseits haben wir herausgefunden, dass ein kultureller Aspekt, insbesondere die Sprache, möglicherweise nicht der Hauptfaktor für diese Korrespondenzen ist. Darüber hinaus wurden Unterschiede in der neuralen Verarbeitung verschiedener modalitätsübergreifender Korrespondenzen gefunden.

Die Ergebnisse dieser Dissertation haben gezeigt, dass verschiedene kontextuelle Aspekte einen nicht zu vernachlässigenden Einfluss auf kognitive Prozesse haben können. Die Ergebnisse dieser Dissertation bereichern das bestehende Wissen in den verschiedenen Bereichen und zeigen, dass der Einfluss kontextueller Aspekten auf kognitive Prozesse in zukünftigen Studien berücksichtig werden sollte.

A Declaration of own contribution

Jaap, C., Rose, M. Relevance of pre-stimulus oscillatory activity for the perceived valence of emotional facial expressions. Sci Rep 14, 19263 (2024). https://doi.org/10.1038/s41598-024-69433-0

Carina Jaap designed, programmed and conducted the experiment, analyzed the data and wrote the manuscript.

Jaap, C., Maack, M.C., Taesler, P. et al. Enriched environments enhance the development of explicit memory in an incidental learning task. Sci Rep 12, 18717 (2022). https://doi.org/10.1038/s41598-022-23226-5

Carina Jaap conducted the PC-long experiment, analyzed the data and wrote the manuscript with in collaboration with Marike Maack.

Jaap, C., & Rose, M. Dissociable neuronal mechanism for different crossmodal correspondence effects in humans. Acta Neurobiologiae Experimentalis, 136–152 (2024). https://doi.org/10.55782/ane-2024-2439

Carina Jaap designed and programmed the experiment, conducted the experiment, analyzed the data and wrote the manuscript.

B Acknowledgements

Almost five years went into this thesis. It was a journey in which I was fortunately not alone. I am grateful for the support I got from all sides starting with colleagues from which many of them turned into friends along the way.

Especially, I want to thank my supervisor, PD Dr. Michael Rose, for his support, good advice and the inspiring conversations we shared. My gratitude also goes to the members of my audit committee. First, PD Dr. Patrick Bruns who was also part of my thesis committee. I highly appreciate the conversations, interest in my work as well as the encouraging words provided when I faced challenges in publishing. I also want to thank Prof. Dr. med. Andreas K. Engel for taking interest in this topic and evaluating the present work. I also want to thank Prof. Dr. Helen Blank for showing interest in my work and giving valuable feedback during my thesis committee meetings.

Furthermore, I am grateful for the positive, diverse and open environment provided by the institute of Prof. Dr. med. Christian Büchel. The vital and great environment provided by the institute is what often helped me finding my way back on track in times when publishing my work was not easy or when facing doubts of whether or not I shall continue with my PhD. Plus, there was always someone joining for a game night, lunch at the park, visit at the cinema and so much more. Finishing my PhD also means that I am leaving behind a ‘work family’ that will be missed and cannot be replaced.

From all the people I first and foremost want to thank my group members, Marike Maack, Jan Ostrowski, Maren Wolf and Christoph Wittkamp. The many hours we spend discussing and helping each other out in the problems we faced during our work were so valuable. However, what especially will not be forgotten are the laughs we shared and the silly moments that made even a lengthy and draining day at work more delightful. There are so many people at the institute that would be worth to be noted but special thanks go to Annika Garlichs, Fabian Schneider, Carina Ufer and Janika Becker. I am deeply grateful for the laughter, the meaningful conversations and the abundance of knowledge we shared in the office, in the corridors, over lunch and beyond. Of course there are so many more people involved in my journey. Too many, to name them all, but I am grateful for the support I received from each and every one of you.

I am also deeply grateful for the constant support I have received from my family. Not only have they supported me throughout my PhD journey, but also throughout the seven years I have spent studying to get to where I am now.

Last but not least I want to express my gratitude to my husband. He did not really understand what I was actually doing most of the time, but always lend me an ear when I needed to go over the topics and issues I was working on. No matter whether it was during the day or in the middle of the night, he patiently listened and was my favorite rubber duck.

C Curriculum Vitae

Entfällt aus datenschutzrechtlichen Gründen.

D Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: ....