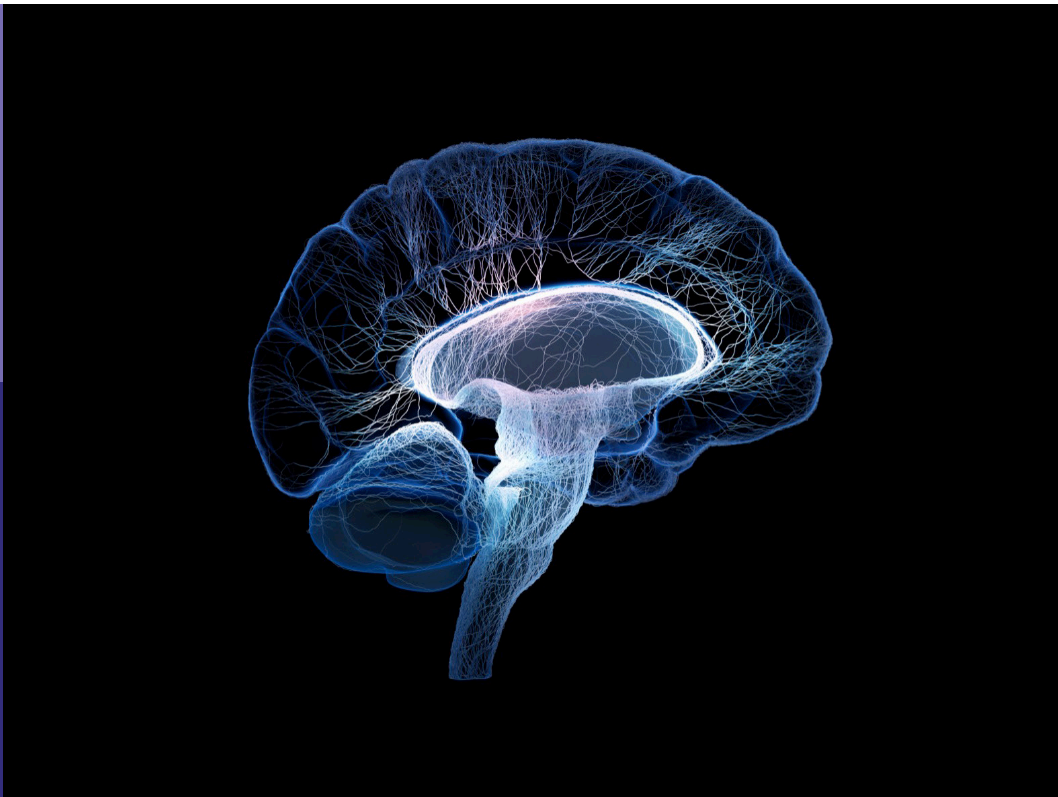
frontiers | Research Topics

Neuroscience and the media

Edited by
Celia Andreu-Sánchez, Miguel Ángel Martín-Pascual
and José M. Delgado-García

Published in
Frontiers in Neuroscience
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-8325-4040-4
DOI 10.3389/978-2-8325-4040-4

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Neuroscience and the media

Topic editors
Celia Andreu-Sánchez — Autonomous University of Barcelona, Spain
Miguel Ángel Martín-Pascual — Spanish Public Television, Spain
José M. Delgado-García — Universidad Pablo de Olavide, Spain

Citation
Andreu-Sánchez, C., Martín-Pascual, M. Á., Delgado-García, J. M., eds. (2023).
Neuroscience and the media. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-4040-4

Table of contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 17 November 2023
DOI 10.3389/fnins.2023.1327123

OPEN ACCESS

EDITED AND REVIEWED BY
Rufin VanRullen,
Centre National de la Recherche Scientifique (CNRS), France

*CORRESPONDENCE
Celia Andreu-Sánchez
✉ celia.andreu@uab.cat

RECEIVED 24 October 2023
ACCEPTED 01 November 2023
PUBLISHED 17 November 2023

CITATION
Andreu-Sánchez C, Martín-Pascual MÁ and Delgado-García JM (2023) Editorial: Neuroscience and the media.
Front. Neurosci. 17:1327123.
doi: 10.3389/fnins.2023.1327123

Editorial: Neuroscience and the media

Celia Andreu-Sánchez[1]*, Miguel Ángel Martín-Pascual[1,2] and José María Delgado-García[3]

[1]Neuro-Com Research Group, Department of Audiovisual Communication and Advertising, Universitat Autònoma de Barcelona, Barcelona, Spain, [2]Research and Development, Institute of Spanish Public Television (RTVE), Corporación Radio Televisión Española, Barcelona, Spain, [3]Division of Neurosciences, University Pablo de Olavide, Seville, Spain

KEYWORDS

neurocinematics, neuroscience, visual perception, media content, neurocommunication

Editorial on the Research Topic
Neuroscience and the media

We watch media content constantly, and yet little is known about how the brain manages the perception of this type of content. Since the 1950s, when Gastaut et al. (1952), Cohen-Séat et al. (1954) and Gastaut and Bert (1954) started to analyze the brain activity of subjects viewing movies with different narratives, much interdisciplinary work has been done. Learning how media affect spectators can have an impact not only in the neurocinematics field of study (Hasson et al., 2008), but also in the understanding of human communication (Murch, 1995; Nakano et al., 2009; Andreu-Sánchez et al., 2018). With this Research Topic, we had the goal of increasing the knowledge about how the brain perceives and processes media content. We thought that this could be of great interest both for media creators who would benefit from a knowledge of perceptual patterns, and for scientists using media content as stimuli in their investigations, who would have more information when designing their stimuli and the variables in them; it would also be of great interest for clinical purposes, since learning the correlations of audio-visual content and brain behavior could inspire new insights.

The different manuscripts included within this topic show several interesting results.

One paper proposes finding physiological markers of viewers' perception and correlating them to short film ratings. The authors reveal some EEG (such as the positive correlation of beta/alpha with film ratings) and peripheral markers (such as facial muscles) that reflect viewers' rating and can predict them to a certain extent (Kosonogov et al.). Another work addresses neural correlates of continuity editing, and finds that the scale of the cuts in the editing affects brain activity. The authors find that edits with an increased scale lead to amplification of the event-related potential (ERP) deflection, while scale reduction leads to decreases, compared with edits keeping the scale across cuts (Sanz-Aznar et al.). In a related work, the impact that camera movements have on audiences is studied. On this regard, results are mixed: while movement made by cameras affects the viewers' sense of involvement, those same movements do not necessarily increase emotional responses in viewers (Yilmaz et al.).

The sound is also investigated to compare perception in viewers when listening to monophonic, stereo, and surround modes. The surround presentation mode shows higher event-related desynchronization (ERD) in alpha and low-beta in the centro-parietal area. The authors suggest that this may be related with an embodied simulation mechanism (Langiulli et al.). Another work proves that the presence of an unfamiliar person while listening to audios modulates the perception: a more homogeneous perception pattern is found when listening to audios when alone, and a more heterogeneous behavior is shown when the listening is done with another person present (Kauttonen et al.).

Advertising communication is also approached in this Research Topic. One work studies differential neural reward reactivity in response to food advertising in children. The authors carry out an experimental proposal using fMRI with children aged 9–12 years old with food and non-food dynamic and static ads, and find significantly higher responses in certain areas, such as the right and left hemispheres of the amygdala and insula for the dynamic food ad medium (Yeum et al. a, b). The authors conclude that the advertising medium has specific effects on neural response to food cues. Another study examines attitudes toward political advertising. Through three experiments, the authors look at differences in social vigilantism and the need for cognition. They found that higher levels of social vigilantism would be related to greater intentions to counterargue and better memory for attitude-incongruent information (Miller et al.).

Additionally, there is an investigation aimed at studying the functional effects of steady-state visual evoked potentials (SSVEP) elicited by rhythmic visual stimulation (RVS) on visuospatial selective attention, since those have been used as biomarkers in studies of neural processing based on the assumption that they would not affect cognition. Here the authors find that target discriminatory accuracy and reaction time vary significantly across the RVS frequency (Li et al.). Another study uses eye-tracking data to modify training of deep convolutional neural networks, and thus change the models' visual attention during object recognition in natural images. The authors present a novel approach to visual perception that can have an impact in neuroscience and in computer science studies (van Dyck et al.).

The topic also includes a paper that investigates the role of spontaneous theory of mind on the processing of dramatic irony scenes in films. Its authors suggest that exposure to undisclosed critical information in cinema enhances the frequency of spontaneous epistemic state inferences and integration into event models of exploitation (Cabañas et al.). Another work provides an overview of approaches to the study of the media while it introduces an organizing scheme that connects the causal path from media content to brain responses and to media effects. In that manuscript, the authors argue the need for creating a new substantive science at the intersection of media and neuroscience (Schmälzle and Huskey).

Finally, media professionalization is studied in a work that finds beta-band differences in primary motor cortex between media and non-media professionals when watching motor actions in movies, suggesting that media expertise could be related with beta-band activity in motor actions (Andreu-Sánchez et al.). This could be of interest in brain-computer interface training contexts.

Overall, this Research Topic includes several works aimed at learning more about how media can be studied from a neuroscientific perspective.

Author contributions

CA-S: Conceptualization, Project administration, Writing—original draft, Writing—review & editing. MM-P: Conceptualization, Project administration, Writing—original draft, Writing—review & editing. JD-G: Conceptualization, Project administration, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grant PID2021-122446NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe to JD-G.

Acknowledgments

We thank all the authors and, especially, the reviewers who participated in making this Research Topic a reality.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Andreu-Sánchez, C., Martín-Pascual, M. Á., Gruart, A., and Delgado-García, J. M. (2018). Chaotic and fast audiovisuals increase attentional scope but decrease conscious processing. *Neuroscience* 394, 83–97. doi: 10.1016/j.neuroscience.2018.10.025

Cohen-Séat, G., Gastaut, H., and Bert, J. (1954). Modification de l'E.E.G. pendant la projection cinématographique. *Rev. Int. Film.* 5, 7–25.

Gastaut, H., and Bert, J. (1954). EEG changes during cinematographic presentation (moving picture activation of the EEG). *Electroencephalogr. Clin. Neurophysiol.* 6, 433–444. doi: 10.1016/0013-4694(54)90058-9

Gastaut, H., Terzian, H., and Gastaut, Y. (1952). Etude d'une activité électroencéphalographique méconnue: le rythme rolandique en arceau. *Mars. Med.* 89, 296–310.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D. J. (2008). Neurocinematics: the neuroscience of film. *Projections* 2, 1–26. doi: 10.3167/proj.2008.020102

Murch, W. (1995). *In the Blink of an Eye: A Perspective on Film Editing*. 2nd ed. Beverly Hills, CA: Silman-James Press.

Nakano, T., Yamamoto, Y., Kitajo, K., Takahashi, T., and Kitazawa, S. (2009). Synchronization of spontaneous eyeblinks while viewing video stories. *Proc. Biol. Sci.* 276, 3635–3644. doi: 10.1098/rspb.2009.0828

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 13 September 2022
DOI 10.3389/fnins.2022.975639

OPEN ACCESS

EDITED BY
Paul Sajda,
Columbia University, United States

REVIEWED BY
Samuel Spevack,
Exponent, United States
Kaveri Thakoor,
Columbia University, United States

*CORRESPONDENCE
Leonard Elia van Dyck
leonard.vandyck@plus.ac.at

SPECIALTY SECTION
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

RECEIVED 22 June 2022
ACCEPTED 25 August 2022
PUBLISHED 13 September 2022

CITATION
van Dyck LE, Denzler SJ and
Gruber WR (2022) Guiding visual
attention in deep convolutional neural
networks based on human eye
movements.
Front. Neurosci. 16:975639.
doi: 10.3389/fnins.2022.975639

Guiding visual attention in deep convolutional neural networks based on human eye movements

Leonard Elia van Dyck[1,2]*, Sebastian Jochen Denzler[1] and Walter Roland Gruber[1,2]

[1]Department of Psychology, University of Salzburg, Salzburg, Austria, [2]Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, Austria

Deep Convolutional Neural Networks (DCNNs) were originally inspired by principles of biological vision, have evolved into best current computational models of object recognition, and consequently indicate strong architectural and functional parallelism with the ventral visual pathway throughout comparisons with neuroimaging and neural time series data. As recent advances in deep learning seem to decrease this similarity, computational neuroscience is challenged to reverse-engineer the biological plausibility to obtain useful models. While previous studies have shown that biologically inspired architectures are able to amplify the human-likeness of the models, in this study, we investigate a purely data-driven approach. We use human eye tracking data to directly modify training examples and thereby guide the models' visual attention during object recognition in natural images either toward or away from the focus of human fixations. We compare and validate different manipulation types (i.e., standard, human-like, and non-human-like attention) through GradCAM saliency maps against human participant eye tracking data. Our results demonstrate that the proposed guided focus manipulation works as intended in the negative direction and non-human-like models focus on significantly dissimilar image parts compared to humans. The observed effects were highly category-specific, enhanced by animacy and face presence, developed only after feedforward processing was completed, and indicated a strong influence on face detection. With this approach, however, no significantly increased human-likeness was found. Possible applications of overt visual attention in DCNNs and further implications for theories of face detection are discussed.

KEYWORDS

vision, attention, brain, deep neural networks, eye tracking, saliency map, object recognition, face detection

Introduction

With the groundbreaking discovery of simple and complex cells as well as their receptive field arrangements, Hubel and Wiesel (1959, 1962) not only laid the foundation for decades of subsequent findings in visual neuroscience but also ignited an idea in another field emergent at that time – computer vision. The *Neocognitron* was introduced by Fukushima and Miyake (1982), who translated the aforementioned findings to one of the earliest multi-layer artificial neural networks and thereby set the cornerstone for image and pattern understanding in deep learning. Nearly half a century later, computer and biological vision research cannot be imagined without the Neocognitron's well-known successors, commonly referred to as *Deep Convolutional Neural Networks* (DCNNs; LeCun et al., 2015), the state-of-the-art models of biological processes such as object recognition. View-invariant core object recognition, that is "the ability to rapidly recognize objects despite substantial appearance variation" (DiCarlo et al., 2012), is a key mechanism in visual information processing that allows humans and animals to perceive, understand, and act upon the surrounding environment.

Today, a lot of research is devoted to the architectural and functional similarities of the brain and DCNNs during visual object recognition (Yamins and DiCarlo, 2016; Cichy and Kaiser, 2019; Kietzmann et al., 2019; Richards et al., 2019; Storrs and Kriegeskorte, 2019). Evidence has been steadily accumulated throughout different modalities and measurements (i.e., primarily brain imaging and neural time series data). This remarkable body of literature suggests that especially the hierarchical organization of both biological and artificial cascades seems to be highly similar (Yamins et al., 2013, 2014; Cadieu et al., 2014; Khaligh-Razavi and Kriegeskorte, 2014; Güçlü and van Gerven, 2015; Cichy et al., 2016; Kheradpisheh et al., 2016; Horikawa and Kamitani, 2017; Greene and Hansen, 2018; Schrimpf et al., 2020). Neural activations of areas along the ventral visual pathway, which is known to be primarily specialized in object recognition (Ishai et al., 1999), and network layers can be mapped onto each other in an ascending spatial and temporal fashion. While earlier regions of the visual cortex (e.g., primary visual cortex V1 or secondary visual cortex V2) correlate with predominantly earlier model layers and encode mainly lower-level features such as edges, blobs, and color, later regions (e.g., visual area V4 or inferior temporal cortex ITC) coincide with later model layers in their functionality of grouping conceptually higher-level features in order to recognize parts and in turn entire objects (Marr, 1982; LeCun et al., 2015). Despite this striking parallelism, there is also a lot of research that revolves around the fundamental dissimilarities between the ventral visual pathway and its computational models. Several studies have demonstrated that in addition to feedforward propagation of neural activity, the processing of some visual inputs may require additional recurrent and feedback connections (Felleman and Van Essen, 1991; Lamme and Roelfsema, 2000; Muckli, 2010; Muckli et al., 2015), which are presumably only activated in primary visual areas after an initial end-to-end sweep of ∼150–200 ms (Thorpe et al., 1996; Kar and DiCarlo, 2020; Loke et al., 2022) and entirely missing in off-the-shelf DCNNs (Kar et al., 2019; Kietzmann et al., 2019). This is not surprising, given the fact that the cortex is a heavily specialized, hierarchical architecture with an abundance of differently directed neuronal projections (for review see Bastos et al., 2012).

Even though DCNNs have already reached human benchmark performance in object recognition tasks (Krizhevsky et al., 2012; He et al., 2016), recent advances in computer vision hold out the prospect of further milestones with increasingly deep and sophisticated models. However, from a neuroscientific perspective, this development introduces DCNNs that operate in increasingly dissimilar fashion compared to the postulated biological mechanisms (Nonaka et al., 2021), which makes them incomparable to human vision and less useful as explanatory models. In contrast, a key mission of vision science is to reverse-engineer object recognition to build better and, in turn, infer from more brain-like models.

Therefore, a question of both scientific and practical nature arises: *How can we develop DCNNs that use more human-like processes to recognize objects?* Theoretically, the most suitable answer to this demand would incorporate both an architecture similar to that of the ventral visual pathway and analogical data that the brain is using to solve this particular challenge. In a previous study, we provided evidence that a DCNN with a receptive field organization resembling that of the ventral visual pathway displays more human-like visual attention in object recognition (van Dyck et al., 2021). With eye tracking, we recorded fixations of human participants during an object categorization task with natural images and compared it to *Gradient-weighted Class Activation Mapping* saliency maps (Selvaraju et al., 2017; hereafter referred to as GradCAMs) of several models with different biological plausibility. The architectural modification yielded more human-like visual attention, as the biologically inspired model focused more on image parts that primarily attracted fixations. Our findings complemented the work by Mehrer et al. (2021), who introduced an image dataset called *Ecoset*, which organizes objects in entry-level categories commonly used by humans (Rosch, 1978). Interestingly, training standard DCNNs on this restructured entry-level category system (e.g., *bird* and *fish*) already improved the model's similarity to human fMRI activity significantly compared to a subordinate category structure (e.g., *great grey owl* and *coho salmon*) – the default in computer vision. This finding leads us to the assumption that a data-driven manipulation of training examples based on human eye movements should likewise increase the similarity between *in vivo* and *in silico* object recognition. Therefore, in this paper, we focus on increasing the human-likeness of visual attention in DCNNs through manipulated visual inputs, rather than previously reported modified architecture, to demonstrate that there are generally two approaches in the sense of nature (i.e., innate mechanisms or architecture) and nurture (i.e., experience or training data) to influence visual attention. To achieve this, we apply a fine-tuning approach to a widely used DCNN called AlexNet (Krizhevsky et al., 2012), in order to guide its focus during visual information processing toward features that are highly relevant to humans and thereby increase its human-likeness in object recognition. In this procedure, we use human observers' fixation density maps (hereafter referred to as eye tracking heatmaps) as a blueprint for a blurring manipulation applied to the training images accordingly. While image parts, that human observers fixated on, remain clearly visible and unchanged in information, parts that did not receive any overt visual attention are blurred and thereby reduced in informational content. Through this *Human-Spotlight* manipulation, the models should additionally be trained to "see" through the lens of and use features meaningful to human observers.

In this study, we aim to shed light on three sets of research questions. Hard and soft attention mechanisms in deep learning (for review see Niu et al., 2021) have been primarily investigated from an engineering standpoint to solve practical problems (Juefei-Xu et al., 2016). Here, we want to show that this approach could be a key tool to incorporate human-likeness in DCNNs and by doing so take a step forward toward explainable artificial intelligence (Adadi and Berrada, 2018). Therefore, the first section of this paper will revolve around the question (i) *whether guided focus based on human eye movements — the process of not only teaching a model how to correctly categorize but additionally where to look for information relevant to humans—increases the human-likeness of the models*. We will use human eye tracking heatmaps acquired in an object categorization paradigm (see **Figure 1A**) to blur images accordingly, which are then used for fine-tuning different DCNNs originating from a common pretrained architecture (see **Figure 1B**). We hypothesize that the model should learn manifest, accuracy-relevant as well as latent, similarity-relevant information and thereby increase in human-likeness. Furthermore, recent evidence from Jang et al. (2021) is important to take into account here, as it suggests that training several state-of-the-art DCNNs, including AlexNet, on images with added noise increases the models' similarity to human behavioral performance and neural representations under challenging conditions. They argue that this might be a result of the increased noise-robustness found in later layers of these models. Therefore, in addition to our *Human-Spotlight* experimental condition (hereafter HS), we incorporate an *Anti-Spotlight* control condition (hereafter AS; see **Figure 1C**), the inverse of the former manipulation, to show that it is possible to guide visual attention of DCNNs in either direction and that the observed effects are a result of the intended manipulation and not simply due to training on noisy data.

Another important question we want to address in the second section of this paper is (ii) *how the growing body of neural time series and neuroimaging literature on the temporal unfolding of feedforward and feedback visual processing can be tested in the realm of eye tracking*. In line with powerful frameworks such as predictive coding theory (Friston, 2005), feedback connections in the visual cortex are thought to drive primarily top-down modulations of the bottom-up sensory input based on internally generated predictions of the external world. This idea fits well with findings on manipulated images that seem to require especially recurrent/feedback processing after ∼150–200 ms of feedforward processing and therefore considerably challenge DCNNs (Tang et al., 2018; Kar et al., 2019; Rajaei et al., 2019). An illustrative example of this are occlusion manipulations, where a part of an image is masked. Based on general assumptions of predictive coding, this covert part should elicit a strong prediction error, as the bottom-up sensory input disagrees with and must be replaced by top-down predictions. As shown by Muckli et al. (2015) with high-resolution layer fMRI, in the case of occlusion, more feedback connections are activated in the cortical layers of V1, as this region seems to receive additional top-down attention. Thus, contrary to parts of our previous analyses (van Dyck et al., 2021), we have come to the general assumption that the initial feedforward sweep has to be completed for preattentive processing before systematic eye movements (i.e., saccades) can be planned and increasingly feedback, attentive processing of critical features (i.e., fixations) can set in Treisman (1985), Van der Stigchel et al. (2009). Hence, we hypothesize that the similarity between eye tracking heatmaps and GradCAMs should abruptly increase and peak after ∼150–200 ms postulated by neural time series data and decrease continuously, as processing shifts gradually from feedforward to feedback mechanisms and purely feedforward DCNNs should not be able to capture the latter.

The third section of this paper will follow up on (iii) *possible methodological advantages and disadvantages of forcing a model to use specific features.* As previous works reported performance decreases in noise-trained DCNNs when tested on natural, noise-free images (Geirhos et al., 2018; Jang et al., 2021), here we hypothesize that an increase in similarity causes a decrease in accuracy and vice versa. As the model generally optimizes its parameter weights, changing the availability of these features should, in most cases, lead to comparatively suboptimal feature distributions with noise added by the image manipulation. To investigate this trade-off for the HS and AS manipulation, we will fine-tune model versions with different stepwise ratios of natural and HS or AS images and describe how accuracy and similarity change as a function of the magnitude of our guided focus manipulation (see **Figure 1D**).

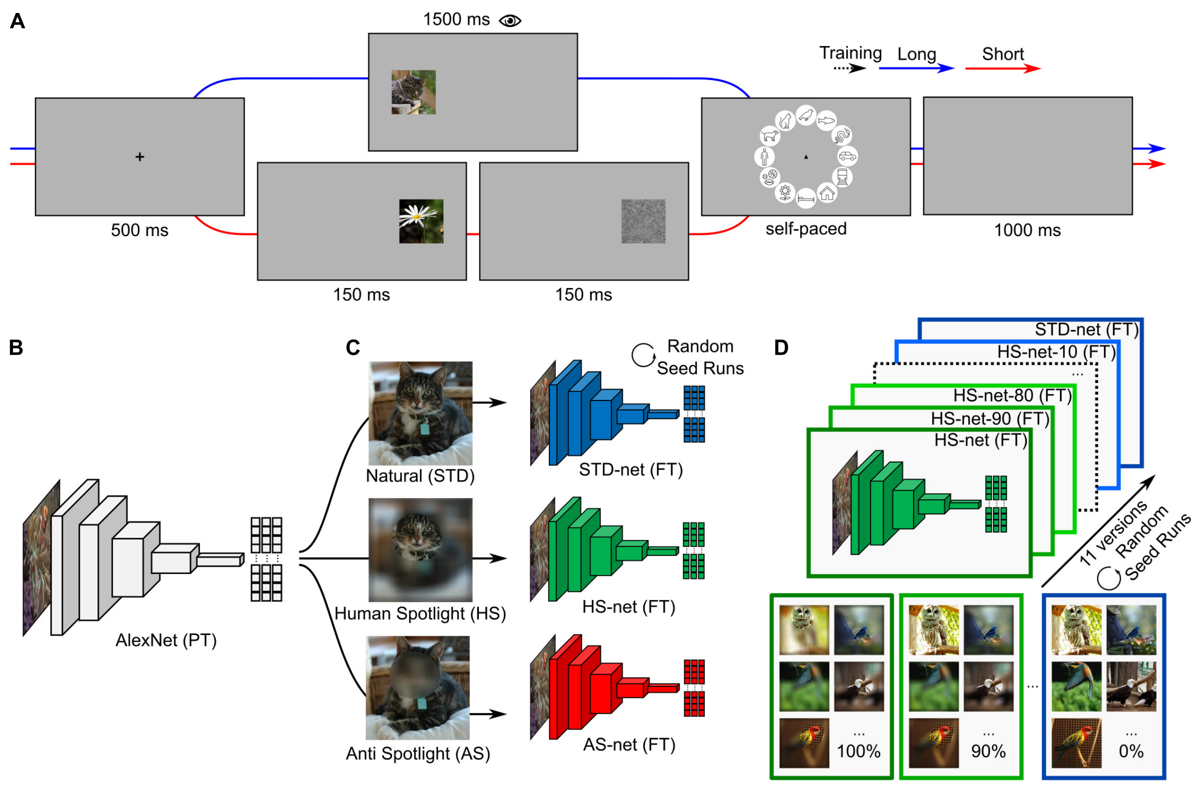


FIGURE 1
Conceptual overview. **(A)** Experimental design of the eye tracking paradigm. The experiment consists of two blocks. Trials in the first block consist of a fixation cross (500 ms), followed by an image presented on either the left or the right side for a long duration (1,500 ms), a circular categorization screen with category icons (self-paced with a time limit of 5,000 ms), and a blank interstimulus interval (1,000 ms). Trials in the second block are similar except that an image was presented only briefly (150 ms) and followed by a visual backward mask (1/f pink noise; 150 ms). **(B)** Pretrained AlexNet (PT) consists of an image input, five convolutional, and three fully connected layers. **(C)** AlexNet versions are fine-tuned (FT) on identical images of twelve categories with different data-driven blurring manipulations (STD-nets, standard; HS-nets, Human-Spotlight; AS-nets, Anti-Spotlight). **(D)** HS-nets and AS-nets with different manipulation ratios in the training dataset are fine-tuned ranging from all HS and AS images (100%) to all STD images (0%) in steps of 10%. The same procedure, here visualized for HS-nets, was applied to AS-nets with the corresponding manipulation type. Source: images of cat, birds, and nature obtained from the Ecoset image dataset (https://dx.doi.org/10.24433/CO.4784989.v1).

Materials and methods

General procedure

For the comparison of object recognition in humans and DCNNs, we attempted to create an experimental setup considering literature on fair and insightful human-machine comparisons (Firestone, 2020; Funke et al., 2021). Therefore, we adopted various measures to align the tasks accordingly. A few examples that aim to restrict certain one-sided advantages are the consistent use of entry-level categories, a forced-choice categorization, and an image presentation of 150 ms followed by a visual backward mask for the evaluated categorization data. First, the eye tracking experiment was conducted. Here, data from block one (i.e., long image presentation of 1,500 ms) were used to compute heatmaps and to enable a comparison of several time windows across the course of feedforward and recurrent processing, while data from block two (i.e., short image presentation of 150 ms) were devoted to behavioral measures such as accuracy and error patterns based on feedforward visual processing only. The eye tracking data from one half of the images were used to create image versions with different guided focus manipulations. These images were then used to fine-tune three different AlexNet manipulation types (hereafter *standard AlexNet* = STD-net, *Human-Spotlight AlexNet* = HS-net, *Anti-Spotlight AlexNet* = AS-net) with otherwise identical training procedures. The only difference here were the differently blurred versions of the same fine-tuning examples. The obtained models were then tested on the other half of the images. Both eye tracking heatmaps and GradCAMs were also used from incorrectly classified images (while GradCAMs were always extracted for the true category label) to avoid comparisons contingent on subjective responses (VanRullen, 2011).

Human observers—Eye tracking experiment

In the eye tracking experiment, a total of 49 participants (25 female, 22 male, 2 other) with an age ranging from 18 to 62 years (M = 24.60, SD = 8.07) were tested. Participants had normal or corrected-to-normal vision, no problems with color vision, and no other eye diseases. One participant had to be excluded from the analyses due to inaccurate eye tracking calibration. The experimental procedure was admitted by the University of Salzburg ethics committee, in line with the declaration of Helsinki, and agreed to by participants via written consent prior to the experiment. Psychology students were compensated for taking part in the study with accredited participation hours.

The experimental paradigm (see **Figure 1A**) consisted of two blocks with 360 trials each. During trials of the first block with long image presentation, participants were presented a fixation cross for 500 ms, followed by an image positioned at ∼13 degrees of visual angle on either the left or right side for 1,500 ms. The original image was scaled by factor two (454-by-454 px and 11.52 degrees of visual angle wide), while the original resolution (227-by-227 px) was kept. After that, a categorization screen with circularly arranged icons of the twelve entry-level categories was displayed for a maximum duration of 5,000 ms. Participants were instructed to choose a category as fast as possible with a mouse click. The origin of the mouse cursor was set to the center of the circle in each trial. The order of category icons was fixed with roughly animate categories on the top and inanimate categories on the bottom half to aid memorability of icon locations. A blank screen with an interstimulus interval of 1,000 ms terminated the trial. Trials in the second block with short image presentation were identical except for a brief image presentation of only 150 ms, which was followed by a visual backward mask (1/f pink noise) for the same duration. All participants had to complete twelve practice trials before starting the actual experiment. At the end of the practice trials, open questions could be discussed and clarified. Training trials were excluded from analyses. The image datasets of the first and second block with 360 images each were interchanged across participants, the order of the presented images within blocks was fully randomized for each participant, and the assignment of image datasets to blocks as well as the presented image position were randomized across different versions of the experiment. With this design, it was possible to simultaneously obtain eye tracking data during longer time periods of image presentation (i.e., 1,500 ms), to test the similarity between eye tracking heatmaps and GradCAMs across time, as well as categorizations based on brief image presentations (i.e., 150 ms), to compare categorizations as a result of assumed feedforward visual processing for every image across two participants. The experiment took approximately one hour to complete and was divided by short breaks in-between and within blocks. The experiment was programmed with ExperimentBuilder (SR Research Ltd., Mississauga, ON, Canada) and conducted in a controlled laboratory setting. Participants were seated in front of a screen (1920-by-1080 px, 60 Hz) and had to place their head onto a chin rest at a distance of 60 cm. An EyeLink 1000 (SR Research Ltd., Mississauga, ON, Canada) desktop mount was used to track the participants' right eye at a sampling rate of 1,000 Hz. Recorded eye movements were preprocessed and analyzed in MATLAB (Version 2022a, The MathWorks, Inc., Natick, MA, United States) using the Edf2Mat Toolbox (Etter and Biedermann, 2018). First, all timepoints with x- and y-coordinates within the presented image coordinates were grouped into time windows of 50 ms before heatmaps were compiled for these time windows of shifting onset. In this way, the first 1,000 ms within the image boundaries were analyzed, allowing a maximum of 20 discrete time windows. After that, the obtained heatmaps were downscaled to original image size (227-by-227 px) and a Gaussian filter with a standard deviation of 20 px was applied. Lastly, heatmaps of individual participants were averaged and normalized per image.

Deep convolutional neural networks—Fine-tuning and saliency maps

The investigated models were trained using MATLAB and Deep Learning Toolbox (The MathWorks Inc, 2022, The MathWorks, Inc., Natick, MA, United States) using a conventional transfer learning approach. Here, an implementation of AlexNet (see **Figure 1B**; Krizhevsky et al., 2012), consisting of an image input, five convolutional, as well as three fully connected layers, and pretrained on >1 million images of 1,000 categories from the ImageNet database (Deng et al., 2009), was fine-tuned on the images from the publicly available Ecoset dataset (Mehrer et al., 2021), which were also presented in the first block of the eye tracking experiment. Unmanipulated, STD-images were used for fine-tuning STD-net, while manipulated HS-images and AS-images were used for HS-nets and AS-nets, respectively (see **Figure 1C**). During the fine-tuning procedure, the last fully connected, softmax, and classification layer were replaced while earlier layers were frozen and the models retrained with the Stochastic Gradient Descent with Momentum (SGDM) optimizer, a mini-batch size of 42, an initial learn rate of 10^{-4}, and a weight learn rate factor of 20 in mentioned layers. All models were trained with varying image versions but identical settings for a maximum of 30 epochs and a validation patience of five epochs. No data augmentation was applied throughout the process. Saliency maps were computed based on the models last convolutional layer using the GradCAM attribution algorithm (Selvaraju et al., 2017) for the true class label. The last convolutional layer marks the final processing stage of the encoding procedure, contains information used for subsequent classification, and in turn corresponds best to the theoretical concept of eye movements as a final processing stage and behavioral result of visual perception. In this way, the saliency map (227-by-227 px) represents the normalized contribution of a specific pixel to the class probability score of the true image category. Unlike eye tracking heatmaps, saliency maps do not vary across time for static images. The previously reported grid-phenomena of saliency map activations (van Dyck et al., 2021) should not influence the comparability in this study, as all models share identical architecture.

Additionally, to investigate the proposed accuracy versus similarity trade-off of our guided focus approach, HS-nets with different ratios of HS- and STD-images within categories were fine-tuned (see **Figure 1D**). These models ranged from versions fine-tuned on only manipulated images (HS-net-100) to versions trained on only STD-images (HS-net-0) with steps of 10% in-between. This resulted in eleven different versions, where for example three images per category were unmanipulated in the HS-net-90, half of the images were unmanipulated in the HS-net-50, and all images were manipulated and unmanipulated in the HS-net-100 (i.e., HS-net) and HS-net-0 (i.e., STD-net), respectively.

To increase the statistical power of both main analyses and overcome the general weakness of commonly used fixed-value comparisons, we estimate the models' accuracy and similarity by repeating this process with random initialization seeds (e.g., Mehrer et al., 2021) as well as shuffled training and validation splits for the number of compared human participants ($N = 23$) and ten times per ratio in the accuracy versus similarity trade-off analysis. Hence, we train and test the model throughout different runs and average all obtained values to improve signal-to-noise ratio and further reduce the influence of individual unmanipulated or manipulated images.

Image dataset

Data for fine-tuning the DCNNs consisted of images from the Ecoset testing set (Mehrer et al., 2021). In the HS condition, images which were displayed to human participants for 1,500 ms during the first experimental block were modified in the way that image parts with fixation densities below the first tertile were blurred and above remained unchanged to create a blurred out surrounding of the focal points of fixations. The inverse procedure was used in the AS condition. For blurring, a Gaussian filter ($W = 30$, $SD = 7$) was applied to the image followed by another Gaussian filter as an edge taper ($W = 35$, $SD = 9$). For the eye tracking experiment and DCNN testing, a dataset of 360 images in total, consisting of twelve entry-level categories, with six animate categories (namely *human, dog, cat, bird, fish,* and *snake*), six inanimate categories (namely *car, train, house, bed, flower,* and *ball*), and 30 images each was used. All images were randomly drawn and preprocessed in identical procedure. Images were cropped to the largest central square and resized to fit AlexNet's image input layer (227-by-227 px). A manual visual inspection ensured that images did not include multiple categories of interest, overlaid text, and image effects or did not show the object anymore due to cropping. All violations were removed. Furthermore, to identify face presence as well as object and face regions of interest (ROI), two independent raters manually detected and segmented the images of the testing dataset with in-house built MATLAB scripts. The obtained object ROIs ($N = 360$) and face ROIs ($N = 133$) were averaged, normalized, and a Gaussian filter was applied ($W = 10$, $SD = 5$, see **Supplementary Figures**). To investigate the detection of face ROIs, a face detection index was computed through the proportion of visual attention within compared to outside of the circumscribed region. This resulted in high scores in cases where visual attention was focused specifically on face regions and lower scores when attention was distributed across the image.

Results

Performance

To investigate object recognition performance, categorization data of human participants acquired under short image presentation of 150 ms were compared to the testing results of DCNN runs. A Welch-ANOVA, due to violated homoscedasticity (indicated by significant Levene's tests), revealed significant differences between the groups [$F(3,48.44) = 147.78$, $p < 0.001$, $\eta^2 = 0.77$; see **Figure 2**]. Bonferroni-corrected pairwise t-tests revealed significant differences between human participants ($M = 89.80$, $SD = 3.18$) and standardly fine-tuned STD-nets [$M = 88.27$, $SD = 1.06$; $t(30.91) = 2.31$, $p = 0.028$, $d = 0.64$]. Nevertheless, this significant but marginal accuracy difference demonstrates that DCNNs should principally be able to reach human benchmark performance with the fine-tuning and experimental procedures at hand. As expected, both HS-nets ($M = 82.22\%$, $SD = 1.92$) and AS-nets ($M = 81.11$, $SD = 1.57$) were significantly outperformed by human participants [HS-net: $t(41.81) = 10.21$, $p < 0.001$, $d = 28.21$; AS-net: $t(37.28) = 12.33$, $p < 0.001$, $d = 28.21$] and STD-nets [HS-nets: $t(33.89) = 13.26$, $p < 0.001$, $d = 3.91$; AS-nets: $t(38.29) = 18.25$, $p < 0.001$, $d = 5.38$] due to noise added by both guided focus manipulation types. However, contrary to our expectations, AS-nets did not significantly exceed HS-nets even though in the AS manipulation images were blurred to a lesser extent ($M = 23.62$, $SD = 0.06$) compared to the HS manipulation ($M = 76.38$, $SD = 0.06$). Instead, the opposite was observed, as HS-nets even significantly outperformed AS-nets [$t(42.26) = 2.15$, $p = 0.038$, $d = 0.63$]. Moreover, further exploratory analyses of accuracies across categories

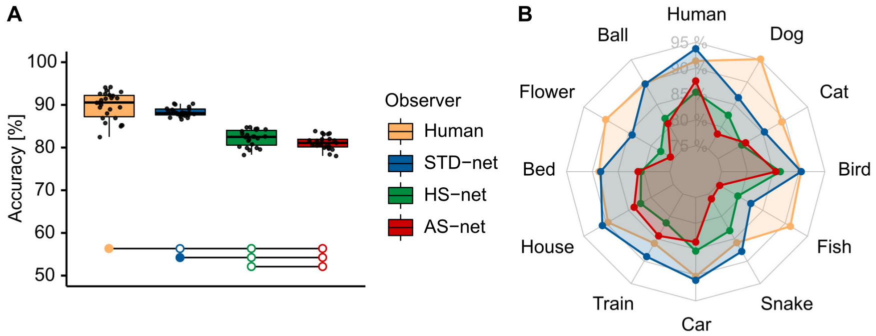


FIGURE 2
Mean object recognition accuracy of human participants and DCNN runs. **(A)** Mean accuracy of human participants, standardly fine-tuned STD-nets, purposefully manipulated HS-nets, and inversely manipulated AS-nets across all testing images. Groups marked by filled dots outperformed groups marked by empty dots significantly at the $p < 0.05$ level. **(B)** Mean accuracy profiles across categories.

and common misclassifications indicated unique behavioral patterns (see **Supplementary Figures**).

Visual attention

Average human eye tracking heatmaps acquired under long image presentation of 1,500 ms and DCNN GradCAMs were compared to test the human-likeness of individual models (for examples see **Supplementary Figures**). Therefore, mean Pearson correlation coefficients between eye tracking heatmaps and GradCAMs for individual testing images were transformed into Fisher's-Z-scores (Z_r) to aid linear comparability (see **Figure 3**). A vast distribution of scores was observed. Due to violated normal distribution of the data (indicated by significant Shapiro-Wilk tests), the following statistical analyses were performed non-parametrically. For two-factorial non-parametric tests, we use an advanced statistical procedure from the *nparLD* R software package and report its ANOVA-type statistic (ATS; Noguchi et al., 2012). Moreover, as pairwise correlations coefficients between eye tracking heatmaps and different GradCAMs could be considered as differences within the same group (i.e., human participants) and thereby may result in dependencies, we test the coefficients with overlap using the *cocor* R package (Diedenhofen and Musch, 2015) and also report the Hittner et al. (2003) modification of Dunn and Clark's (1969) approach of a back-transformed average Fisher's Z procedure.

As hypothesized, a Kruskal–Wallis test indicated significant differences between the three models [$H(2) = 26.52$, $p < 0.001$, $\eta^2 = 0.02$]. Generally, in terms of our guided focus manipulation, we expected HS-nets to demonstrate more human-like visual attention compared to their inversely conceptualized AS-net counterparts and commonly fine-tuned STD-nets. Most interestingly, Bonferroni-corrected pairwise Wilcoxon tests indeed revealed that AS-nets (Mdn = 0.10) focused on significantly less human-like image parts compared to STD-nets (Mdn = 0.24; $W = 77332$, $p < 0.001$, $r = 0.17$; $z = 4.07$, $p < 0.001$) and HS-nets (Mdn = 0.23, $W = 77149$, $p < 0.001$, $r = 0.16$; $z = 3.23$, $p < 0.001$). However, no overall significant differences between HS-nets and STD-nets were found ($W = 64527$, $p = 0.922$, $r = 0.003$; $z = -0.10$, $p = 0.459$). Remarkably, as hypothesized, a significant interaction effect between manipulation type and animacy was identified [$F_{ATS}(1.62) = 10.26$, $p < 0.001$] and as the overall patterns seemed to remain stable across animacy groups, an additional main effect of animacy may be assumed [$F_{ATS}(1) = 26.09$, $p < 0.001$]. Subsequent Bonferroni-corrected pairwise Wilcoxon tests revealed significant differences for both animate and inanimate objects between AS-nets compared to STD-nets (animate: $W = 20543$, $p < 0.001$, $r = 0.23$; $z = 3.60$, $p < 0.001$/inanimate: $W = 18353$, $p = 0.029$, $r = 0.11$; $z = 2.04$, $p = 0.021$) and HS-nets (animate: $W = 20507$, $p < 0.001$, $r = 0.23$; $z = 2.89$, $p = 0.002$/inanimate: $W = 18288$, $p = 0.035$, $r = 0.11$; $z = 1.61$, $p = 0.054$). Generally speaking, most human-like visual attention was obtained with STD-nets and HS-nets categorizing animate objects, while AS-nets categorizing inanimate objects elicited the least human-like attentional processes. Finally, to further pinpoint the central effects of our manipulation, the influence of face presence (i.e., human or animal faces) on the similarity of visual attention was analyzed. The analysis revealed a significant interaction effect between manipulation type and face presence [$F_{ATS}(1.56) = 7.48$, $p = 0.002$] with similar patterns to the reported animacy splits. Here again, in both face and non-face images, AS-nets focused on significantly more dissimilar image parts compared to STD-nets (face: $W = 11211$, $p < 0.001$, $r = 0.23$; $z = 3.14$, $p < 0.001$/non-face: $W = 30004$, $p = 0.002$, $r = 0.14$; $z = 2.67$, $p = 0.004$) and HS-nets (face: $W = 11456$, $p < 0.001$, $r = 0.26$; $z = 2.78$, $p = 0.003$/non-face: $W = 29364$, $p = 0.010$, $r = 0.12$; $z = 1.89$, $p = 0.029$). Taken together, these results imply that the distinct biases for living entities and especially their faces in human gaze patterns may have introduced the opposite effects in AS-nets. As an interim summary, it should be noted that the guided focus approach did not seem to increase the overall fit to human eye tracking data

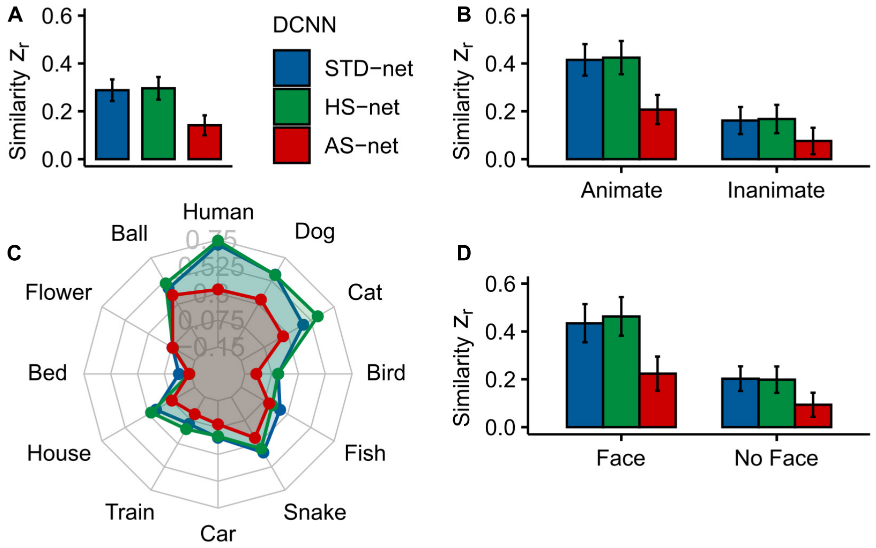


FIGURE 3

Mean Fisher's Z transformed Pearson correlations between DCNN GradCAMs and human eye tracking heatmaps for all individual testing images. **(A)** Overall similarity of standardly fine-tuned STD-nets, purposefully manipulated HS-nets, and inversely manipulated AS-nets compared to eye tracking heatmaps. **(B)** Similarity across animacy. **(C)** Similarity across categories. **(D)** Similarity across face presence. Error bars display bootstrap confidence intervals.

as intended in GradCAMs of HS-nets. However, the single most striking observation to emerge from the data was that HS-nets and AS-nets did show only a marginal accuracy difference and yet the AS manipulation lead to significantly less human-like visual attention. Besides that, it should be noted that individual categories were investigated through exploratory analyses only and no statistical procedures were applied due to the reported accuracy differences.

Face detection

Following up on the observed effects of face presence, especially in AS-nets, we compared the models' GradCAMs to face ROIs, which were visually segmented by two independent raters. Therefore, a face detection index (see section "Materials and methods") was computed. While data from human participants displayed a strong face bias (Mdn = 2.74), which varied considerably across time, DCNNs generally demonstrated clearly less focused attention towards face regions (STD-nets: Mdn = 0.59; HS-nets: Mdn = 0.57; AS-nets: Mdn = 0.42, see **Figure 4**). A Kruskal-Wallis test indicated a significant effect between the models [$H(2) = 7.54$, $p = 0.023$, $\eta^2 = 0.01$] and ensuing Bonferroni-corrected pairwise Wilcoxon tests further revealed significant differences between AS-nets compared to STD-nets ($W = 10351$, $p = 0.016$, $r = 0.15$) and HS-nets ($W = 10319$, $p = 0.019$, $r = 0.14$). These results highlight the importance of faces in human vision, as our AS manipulation seemed to robustly blur them from the fine-tuning dataset and AS-nets thereby shifted their visual attention away from them to classify images without losses in accuracy compared to HS-nets.

Visual attention across time

In a next step, we included the time course of human object recognition processes in our comparison. We hypothesized that during the first feedforward sweep, preattentive processing should prevail before attentive processing can set in and human observers saccade to relevant object features. Hence, in terms of the fit of the computational models, we expected the similarity of visual attention to increase rapidly and peak after this initial time window (i.e., due to similar bottom-up processes) and then gradually diminish due to the onset of increasingly feedback processes (i.e., missing top-down processes in the models). The data resembled our proposed time course only partly, as indeed the similarity increased reliably at ∼150–200 ms but seemed to reach a plateau without a noticeable decrease in later time windows (see **Figure 5**). Splits by animacy and face presence revealed similar patterns to the overall analysis while the previously implied but not significant difference between HS-nets and STD-nets in face images seemed to emerge at the beginning of attentive processing. Moreover, analyses by category revealed remarkably specific time courses, which seem to be the source of the vast distribution of similarity scores and

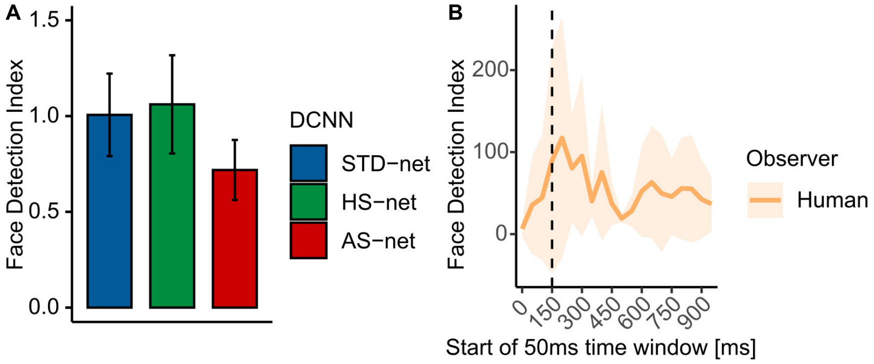


FIGURE 4
Mean face detection indices. The face detection index reflects the proportional attention within compared to outside of the face ROI. **(A)** Mean face detection indices for GradCAMs overall. **(B)** Mean face detection index for eye tracking heatmaps across time. Shaded areas display bootstrap confidence intervals. The dotted line marks the transition from feedforward to recurrent processing at ∼150 ms.

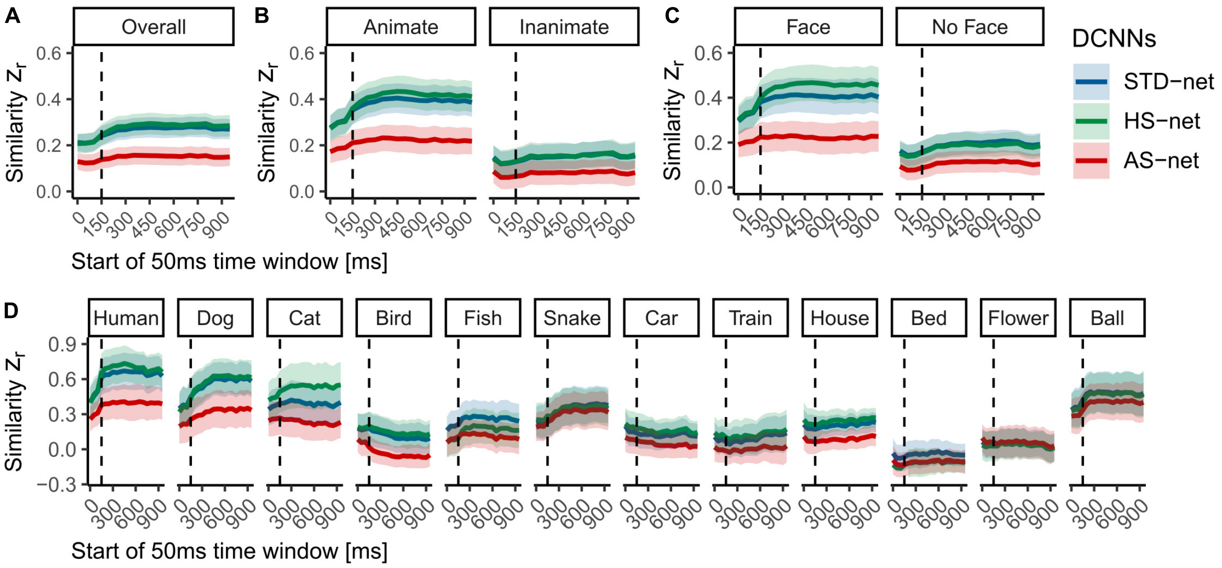


FIGURE 5
Mean Fisher's Z transformed Pearson correlations between DCNN GradCAMs and eye tracking heatmaps for all individual testing images across time. **(A)** Overall similarity of standardly fine-tuned STD-nets, purposefully manipulated HS-nets, and inversely manipulated AS-nets compared to eye tracking heatmaps across the first 1,000 ms within the image boundaries. **(B)** Similarity across time split by animacy. **(C)** Similarity across time split by face presence. **(D)** Similarity across time split by category. Shaded areas display bootstrap confidence intervals. The dotted line marks the transition from feedforward to recurrent processing at ∼150 ms.

should be treated with caution due to the reported accuracy differences.

Accuracy-similarity trade-off

Finally, to investigate possible methodological advantages and disadvantages of the applied approach, we performed accuracy and similarity analyses for ratios of HS and AS manipulations (see section "Materials and methods"). We hypothesized that accuracy and similarity change as a function of magnitude in respect to the directionality of the manipulation as smaller manipulation ratios were expected to return both measurements to the STD-net baseline. Hence, while in HS-nets, smaller manipulation ratios (i.e., more STD-images in the fine-tuning dataset) were expected to lead to increases in accuracy but decreases in human-likeness, in AS-nets both accuracy and similarity were expected to increase. The results endorsed this assumption partially (see **Figure 6**), as AS-nets but not HS-nets followed this pattern. Taken together, these findings endorse the idea that the guided focus approach influences the human-likeness of the models at least in the

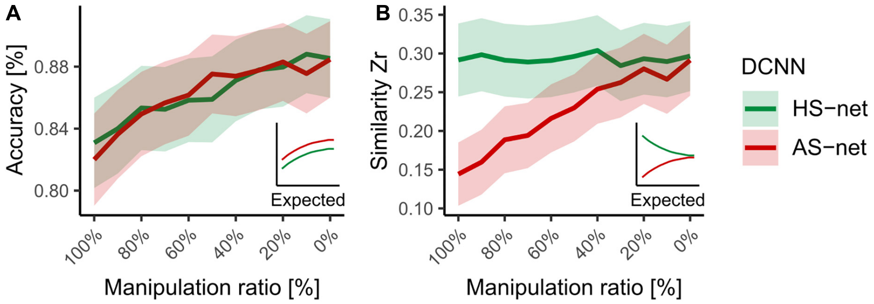


FIGURE 6
Accuracy versus similarity trade-off across HS and AS manipulation ratios ranging from 100 to 0% in steps of 10%. **(A)** Mean object recognition accuracy of DCNN runs for all individual testing images. **(B)** Mean Fisher's Z transformed Pearson correlations between DCNN GradCAMs and human eye tracking heatmaps for all individual testing images. Shaded areas display bootstrap confidence intervals.

negative direction considerably and with reasonable losses in accuracy.

Discussion

In this study, we address the question whether it is possible to influence the human-likeness of DCNNs through a purely data-driven approach. In two opposing guided focus manipulations (i.e., HS-nets and AS-nets), we used eye tracking data to selectively modify the informational content in the training examples and thereby attempted to shift the models' visual attention either toward or away from human-relevant features. Subsequently, we compared and validated the resulting models against new eye movement data.

Generally, analyses of object recognition accuracy implied that human benchmark performance, often reported for the applied architecture (Krizhevsky et al., 2012), was nearly reached by the standardly fine-tuned models (i.e., STD-nets). As anticipated, the guided focus manipulation led to significant losses in accuracy (i.e., in HS-nets and AS-nets) due to the overall reduction of informational content. While there were no significant accuracy differences between HS-nets and AS-nets, this finding turned out to be highly intriguing in the light of the proportional extent of the manipulations and the associated differences in visual attention. Even though in the HS manipulation more than three times as many features were blurred as in the AS manipulation, both models indicated similar accuracy. The HS manipulation was quantitatively more interfering but spared qualitatively more important features. This may illustrate the interplay between object-relevant features and contextual shortcut learning in DCNNs (for review see Geirhos et al., 2020), as both processes represent valid strategies and yet reflect the distribution of information across natural images (for review see Oliva and Torralba, 2007) through their unequal efficiency. Furthermore, the resulting similarities between GradCAMs and eye tracking heatmaps demonstrated significant differences in human-likeness among the models. Contrary to our expectations, these findings did not show a significantly increased overall human-likeness in the experimental condition (i.e., HS-nets). Nevertheless, the hypothesized effect was present in the control condition (i.e., AS-nets), as these models focused significantly more on dissimilar image parts compared to humans. Across following analyses, remarkable results were uncovered.

Firstly, the overall similarity for individual images indicated category-specific effects. This is in accordance with our previous findings on comparisons of human and DCNN visual attention (van Dyck et al., 2021) and may reflect the diverse processes captured by human eye movements (for review see Schütz et al., 2011) as well as their transferability to DCNNs through the applied approach. Here, it is important to note that, in the eye tracking experiment, we cannot determine the exact timepoint at which exogenous, recognition-relevant neural processes are completed and overt visual attention may be influenced by other endogenous, recognition-irrelevant neural processes. This way, it is entirely possible that the overlap between recognition-relevant and recognition-irrelevant but salient features may be highly variable across categories. To further investigate this, we analyzed the human-likeness of GradCAMs across different time windows. The results support the notion, that during feedforward, preattentive processing (i.e., ∼150–200 ms within the image boundaries) subsequent feedback, attentive processing (i.e., saccades and fixations) may be programmed. However, the fit to our computational models only partly resembled the expected time course, as the similarity rapidly increased at ∼150 ms but did not pursue the proposed decline, due to increasingly recurrent/feedback processes that lack in the models. It seems likely that the expected time course and especially the decrease in similarity later on were not observable for specific categories (e.g., *house*, *bed*, and *flower*), as their image examples may display predominantly non-critical, distributed features (e.g., textural surfaces) and not critical, focal features (e.g., shape configurations) as other categories (e.g., *human*, *dog*, and *cat*), which would attract fixations more reliably. Naturally, it is difficult to investigate the impact of feedforward and recurrent processing on eye movements as they may be a mere behavioral proxy of the underlying neural processes (Van der Stigchel et al., 2009). On the one hand, while the initial dip, interpreted as preattentive processing, could be an artifact of the central fixation bias (Tatler, 2007), both interpretations are not necessarily contradicting. On the other hand, as objects are usually, and in our case, located in the center of natural images (i.e., due to "photographer bias," see **Supplementary Figures**), this should have led to higher similarity right from the beginning.

Secondly, we were able to show that animate objects and face presence both significantly increase the effect of our manipulation. These findings are in line with previous results on visual processing of living beings, as they are known to elicit more stereotypical gaze patterns with more frequent and longer fixations compared to inanimate objects (New et al., 2007; Ković et al., 2009; Yang et al., 2012; Jackson and Calvillo, 2013). Following analyses of images including faces helped to further isolate the presumed driving force of our manipulation, as AS-nets detected significantly less face regions compared to other models. Our results corroborate a large body of literature on the unique attentional (Buswell, 1935; Gilchrist and Proske, 2006; Crouzet et al., 2010; Tatler et al., 2010) and computational (Kanwisher et al., 1997; Farah et al., 1998; Haxby et al., 2000; Kanwisher, 2000) status of faces in human vision as a basis for automatic and rapid attraction of fixations. We believe that through this "face bias," found in the eye tracking data, AS-nets learned to selectively ignore face configurations as informative features for object recognition. It remains debatable whether this presumably acquired "face blindness" may indeed support the face expertise hypothesis (Gauthier and Nelson, 2001; Gauthier and Bukach, 2007) and disagree with the innate face bias (Johnson et al., 1991; Johnson and Mareschal, 2001). Interestingly, while Xu et al. (2021) reported the emergence of a face module in fully face-deprived AlexNets, Blauch et al. (2021) discovered noticeable effects of expertise on face recognition in DCNNs that also support our findings. This way, the AS manipulation may not only be useful to further investigate this debate but could also offer plausible computational models for disrupted face perception in disorders such as autism spectrum disorder and prosopagnosia.

We are aware of the possibility that the guided focus manipulation induces merely cosmetic effects. This would explain the absent increased human-likeness in the HS manipulation, as DCNNs may not be able to process the features emphasized by human eye movements adequately. Therefore, future investigations on the interaction effect of biologically inspired architectures and human-relevant data will be needed. However, we believe that the human-likeness of DCNNs may not yet be directly linked to their performance, as manifest, accuracy-relevant, and latent, similarity-relevant information does not necessarily have to be identical. This further strengthens our confidence that the manipulation should principally work in both directions but is challenging to implement and verify. Thus, due to numerous limitations, the effect may have been solely found in the less challenging direction. While saliency maps were criticized to be highly noisy (Kim et al., 2019), our averaging approach should have increased the overall signal-to-noise ratio but also might have amplified the gravitational impact of the previously reported grid-phenomena in saliency maps toward certain points. Furthermore, our approach may be considerably underpowered, as fine-tuning influences especially the last fully connected layer, but saliency maps are extracted from the minimally reweighted last convolutional layer. While we still found the reported differences in human-likeness, future work will benefit from larger eye tracking datasets and fully trained models. From a neuroscience perspective, a visual search paradigm or controversial stimuli (Golan et al., 2020) with different foci of bottom-up and top-down visual attention may be better suited to disentangle the role of feedforward and recurrent processing in the human-likeness of DCNNs. However, they are challenging to implement in object recognition of mutually exclusive targets. From a computer science perspective, future work on recurrent architectures (Mnih et al., 2014) and soft spatial attention (Lindsay, 2020), reweighting solely the representations and not directly entire image parts, may be able to further push the boundaries of such a data-driven approach to design more human-like models without the reported disadvantages. Our findings could have several possible implications for computer vision in the form of explainable artificial intelligence. While architectural modifications may help to better understand which mechanisms a given model uses to come to a general decision (i.e., global interpretability), data-driven manipulations may be able to shed light on the mechanisms that underlie a single decision (i.e., local interpretability; Adadi and Berrada, 2018). We hope that our novel approach excites both fields and sparks innovative ideas for future computational models of vision.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the University of Salzburg Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LD, SD, and WG contributed to the conception and design of the study. LD and SD programmed the experiment and collected the eye tracking data. LD wrote the first draft of the manuscript and implemented the computational models. LD and WG analyzed and interpreted the data. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

An Open Access Publication Fee was granted by the Centre for Cognitive Neuroscience, University of Salzburg.

Acknowledgments

We thank our colleagues Charlotte Paulina Schöllkopf and Rade Kutil for providing insightful comments in our lively discussions. A preprint of this article was published on arXiv: https://arxiv.org/abs/2206.10587.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.975639/full#supplementary-material

References

Adadi, A., and Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6, 52138–52160. doi: 10.1109/ACCESS.2018.2870052

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Blauch, N. M., Behrmann, M., and Plaut, D. C. (2021). Computational insights into human perceptual expertise for familiar and unfamiliar face recognition. *Cognition* 208:104341. doi: 10.1016/j.cognition.2020.104341

Buswell, G. T. (1935). *How People Look at Pictures: A Study of the Psychology and Perception in Art*. Chicago, IL: University of Chicago Press.

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., et al. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Comput. Biol.* 10:e1003963. doi: 10.1371/journal.pcbi.1003963

Cichy, R. M., and Kaiser, D. (2019). Deep Neural Networks as Scientific Models. *Trends Cogn. Sci.* 23, 305–317. doi: 10.1016/j.tics.2019.01.009

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., and Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* 6:27755. doi: 10.1038/srep27755

Crouzet, S. M., Kirchner, H., and Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *J. Vision* 10:16. doi: 10.1167/10.4.16

Deng, J., Dong, W., Socher, R., Li, L., Kai, L., and Fei-Fei, L. (2009). "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (Miami, FL: IEEE), 248–255. doi: 10.1109/CVPR.2009.5206848

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010

Diedenhofen, B., and Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS One* 10:e0121945. doi: 10.1371/journal.pone.0121945

Dunn, O. J., and Clark, V. (1969). Correlation coefficients measured on the same individuals. *J.Am. Statist. Assoc.* 64, 366–377. doi: 10.1080/01621459.1969.10500981

Etter, A., and Biedermann, M. (2018). *Edf2Mat© matlab toolbox (1.20) [Computer software]*. Zürich: University of Zurich.

Farah, M. J., Wilson, K. D., Drain, M., and Tanaka, J. N. (1998). What is "special" about face perception? *Psychol. Rev.* 105, 482–498. doi: 10.1037/0033-295X.105.3.482

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1-a

Firestone, C. (2020). Performance vs. competence in human–machine comparisons. *Proc. Natl. Acad. Sci.* 117:26562. doi: 10.1073/pnas.1905334117

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. B* 360, 815–836. doi: 10.1098/rstb.2005.1622

Fukushima, K., and Miyake, S. (1982). "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition," in *Competition and Cooperation in Neural Nets*, eds S. Amari and M. A. Arbib (Berlin: Springer), 267–285. doi: 10.1007/978-3-642-46466-9_18

Funke, C. M., Borowski, J., Stosio, K., Brendel, W., Wallis, T. S., and Bethge, M. (2021). Five points to check when comparing visual perception in humans and machines. *J. Vision* 21:16. doi: 10.1167/jov.21.3.16

Gauthier, I., and Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition* 103, 322–330. doi: 10.1016/j.cognition.2006.05.003

Gauthier, I., and Nelson, C. A. (2001). The development of face expertise. *Curr. Opin. Neurobiol.* 11, 219–224. doi: 10.1016/S0959-4388(00)00200-2

Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., et al. (2020). Shortcut learning in deep neural networks. *Nat. Mach. Intell.* 2, 665–673. doi: 10.1038/s42256-020-00257-z

Geirhos, R., Temme, C. R., Rauber, J., Schütt, H. H., Bethge, M., and Wichmann, F. A. (2018). "Generalisation in humans and deep neural networks," in *Proceedings of the 32nd conference on neural information processing systems* (Montréal, QC: NeurIPS).

Gilchrist, I. D., and Proske, H. (2006). Anti-saccades away from faces: evidence for an influence of high-level visual processes on saccade programming. *Exp. Brain Res.* 173, 708–712. doi: 10.1007/s00221-006-0455-1

Golan, T., Raju, P. C., and Kriegeskorte, N. (2020). Controversial stimuli: Pitting neural networks against each other as models of human cognition. *Proc. Natl. Acad. Sci. U.S.A.* 117, 29330–29337. doi: 10.1073/pnas.1912334117

Greene, M. R., and Hansen, B. C. (2018). Shared spatiotemporal category representations in biological and artificial deep neural networks. *PLoS Comput. Biol.* 14:e1006327. doi: 10.1371/journal.pcbi.1006327

Güçlü, U., and van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *J. Neurosci.* 35:10005. doi: 10.1523/JNEUROSCI.5023-14.2015

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/S1364-6613(00)01482-0

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, (Heidelberg: IEEE), 770–778. doi: 10.1109/CVPR.2016.90

Hittner, J. B., May, K., and Silver, N. C. (2003). A Monte Carlo Evaluation of Tests for Comparing Dependent Correlations. *J. General Psychol.* 130, 149–168. doi: 10.1080/00221300309601282

Horikawa, T., and Kamitani, Y. (2017). Generic decoding of seen and imagined objects using hierarchical visual features. *Nat. Commun.* 8:15037. doi: 10.1038/ncomms15037

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* 148, 574–591. doi: 10.1113/jphysiol.1959.sp006308

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154. doi: 10.1113/jphysiol.1962.sp006837

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., and Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc. Natl. Acad. Sci. U.S.A.* 96:9379. doi: 10.1073/pnas.96.16.9379

Jackson, R. E., and Calvillo, D. P. (2013). Evolutionary relevance facilitates visual information processing. *Evol. Psychol.* 11, 1011–1026. doi: 10.1177/147470491301100506

Jang, H., McCormack, D., and Tong, F. (2021). Noise-trained deep neural networks effectively predict human vision and its neural responses to challenging images. *PLoS Biol.* 19:e3001418–e3001418. doi: 10.1371/journal.pbio.3001418

Johnson, M. H., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40, 1–19. doi: 10.1016/0010-0277(91)90045-6

Johnson, M. H., and Mareschal, D. (2001). Cognitive and perceptual development during infancy. *Curr. Opin. Neurobiol.* 11, 213–218. doi: 10.1016/S0959-4388(00)00199-9

Juefei-Xu, F., Verma, E., Goel, P., Cherodian, A., and Savvides, M. (2016). "Deepgender: Occlusion and low resolution robust facial gender classification via progressively trained convolutional neural networks with attention," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (Las Vegas, NV: IEEE), 68–77. doi: 10.1109/CVPRW.2016.24

Kanwisher, N. (2000). Domain specificity in face perception. *Nat. Neurosci.* 3, 759–763. doi: 10.1038/77664

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *J. Neurosci.* 17:4302. doi: 10.1523/JNEUROSCI.17-11-04302.1997

Kar, K., and DiCarlo, J. J. (2020). Fast recurrent processing via ventral prefrontal cortex is needed by the primate ventral stream for robust core visual object recognition. *bioRxiv* [Preprint]. doi: 10.1101/2020.05.10.086959

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., and DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* 22, 974–983. doi: 10.1038/s41593-019-0392-5

Khaligh-Razavi, S.-M., and Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* 10:e1003915. doi: 10.1371/journal.pcbi.1003915

Kheradpisheh, S. R., Ghodrati, M., Ganjtabesh, M., and Masquelier, T. (2016). Deep networks can resemble human feed-forward vision in invariant object recognition. *Sci. Rep.* 6:32672. doi: 10.1038/srep32672

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., and Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proc. Natl. Acad. Sci. U.S.A.* 116:21854. doi: 10.1073/pnas.1905544116

Kim, B., Seo, J., Jeon, S., Koo, J., Choe, J., and Jeon, T. (2019). "Why are Saliency Maps Noisy? Cause of and Solution to Noisy Saliency Maps," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, (Seoul: IEEE), 4149–4157. doi: 10.1109/ICCVW.2019.00510

Ković, V., Plunkett, K., and Westermann, G. (2009). Eye-tracking study of animate objects. *Psihologija* 42, 307–327. doi: 10.2298/PSI0903307K

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural. Inf.Proc. Syst.* 25, 1097–1105.

Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *TINS* 23, 571–579. doi: 10.1016/s0166-2236(00)01657-x

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Lindsay, G. W. (2020). Attention in Psychology, Neuroscience, and Machine Learning. *Front. Comput. Neurosci.* 14:19. doi: 10.3389/fncom.2020.00029

Loke, J., Seijdel, N., Snoek, L., van der Meer, M., van de Klundert, R., Quispel, E., et al. (2022). A critical test of deep convolutional neural networks' ability to capture recurrent processing in the brain using visual masking. *bioRxiv* [Preprint]. doi: 10.1101/2022.01.30.478404

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press.

Mehrer, J., Spoerer, C. J., Jones, E. C., Kriegeskorte, N., and Kietzmann, T. C. (2021). An ecologically motivated image dataset for deep learning yields better models of human vision. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2011417118. doi: 10.1073/pnas.2011417118

Mnih, V., Heess, N., and Graves, A. (2014). Recurrent models of visual attention. *arXiv* [Preprint].

Muckli, L. (2010). What are we missing here? Brain imaging evidence for higher cognitive functions in primary visual cortex V1. *Int. J. Imag. Syst. Technol.* 20, 131–139. doi: 10.1002/ima.20236

Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., et al. (2015). Contextual Feedback to Superficial Layers of V1. *Curr. Biol.* 25, 2690–2695. doi: 10.1016/j.cub.2015.08.057

New, J., Cosmides, L., and Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16598–16603. doi: 10.1073/pnas.0703913104

Niu, Z., Zhong, G., and Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing* 452, 48–62. doi: 10.1016/j.neucom.2021.03.091

Noguchi, K., Gel, Y. R., Brunner, E., and Konietschke, F. (2012). nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. *J. Statist. Softw.* 50, 1–23. doi: 10.18637/jss.v050.i12

Nonaka, S., Majima, K., Aoki, S. C., and Kamitani, Y. (2021). Brain hierarchy score: Which deep neural networks are hierarchically brain-like? *iScience* 24:103013. doi: 10.1016/j.isci.2021.103013

Oliva, A., and Torralba, A. (2007). The role of context in object recognition. *Trends Cogn. Sci.* 11, 520–527. doi: 10.1016/j.tics.2007.09.009

Rajaei, K., Mohsenzadeh, Y., Ebrahimpour, R., and Khaligh-Razavi, S.-M. (2019). Beyond core object recognition: Recurrent processes account for object recognition under occlusion. *PLoS Comput. Biol.* 15:e1007001. doi: 10.1371/journal.pcbi.1007001

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., et al. (2019). A deep learning framework for neuroscience. *Nat. Neurosci.* 22, 1761–1770. doi: 10.1038/s41593-019-0520-2

Rosch, E. (1978). *Principles of Categorization*. Cambridge, MA: MIT Press.

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., et al. (2020). Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like? *bioRxiv* [Preprint]. doi: 10.1101/407007

Schütz, A. C., Braun, D. I., and Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *J. Vision* 11:9. doi: 10.1167/11.5.9

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, (Venice: IEEE), 618–626. doi: 10.1109/ICCV.2017.74

Storrs, K. R., and Kriegeskorte, N. (2019). Deep learning for cognitive neuroscience. *arXiv*. [preprint]

Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Ortega Caro, J., et al. (2018). Recurrent computations for visual pattern completion. *Proc. Natl. Acad. Sci.U.S.A.* 115, 8835–8840. doi: 10.1073/pnas.1719397115

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *J. Vision* 7:4. doi: 10.1167/7.14.4

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., and Velichkovsky, B. M. (2010). Yarbus, Eye Movements, and Vision. *iPerception* 1, 7–27. doi: 10.1068/i0382

The MathWorks Inc (2022). *MATLAB and deep learning toolbox (R2022a) [Computer software].* Natick, MA: The MathWorks Inc.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522. doi: 10.1038/381520a0

Treisman, A. (1985). Preattentive processing in vision. *Comput. Vision Graphics Image Proc.* 31, 156–177. doi: 10.1016/S0734-189X(85)80004-9

Van der Stigchel, S., Belopolsky, A. V., Peters, J. C., Wijnen, J. G., Meeter, M., and Theeuwes, J. (2009). The limits of top-down control of visual attention. *Acta Psychol.* 132, 201–212. doi: 10.1016/j.actpsy.2009.07.001

van Dyck, L. E., Kwitt, R., Denzler, S. J., and Gruber, W. R. (2021). Comparing Object Recognition in Humans and Deep Convolutional Neural Networks—An Eye Tracking Study. *Front. Neurosci.* 15:750639. doi: 10.3389/fnins.2021.750639

VanRullen, R. (2011). Four Common Conceptual Fallacies in Mapping the Time Course of Recognition. *Front. Psychol.* 2:365. doi: 10.3389/fpsyg.2011.00365

Xu, S., Zhang, Y., Zhen, Z., and Liu, J. (2021). The Face Module Emerged in a Deep Convolutional Neural Network Selectively Deprived of Face Experience. *Front. Comput. Neurosci.* 15, 626259–626259. doi: 10.3389/fncom.2021.626259

Yamins, D. L., and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19, 356–365. doi: 10.1038/nn.4244

Yamins, D. L., Hong, H., Cadieu, C., and DiCarlo, J. J. (2013). "Hierarchical modular optimization of convolutional networks achieves representations similar to macaque IT and human ventral stream," in *Proceedings of the 26th International Conference on Neural Information Processing Systems (NeurIPS)*, (Lake Tahoe).

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111:8619. doi: 10.1073/pnas.1403112111

Yang, J., Wang, A., Yan, M., Zhu, Z., Chen, C., and Wang, Y. (2012). Distinct processing for pictures of animals and objects: Evidence from eye movements. *Emotion* 12, 540–551. doi: 10.1037/a0026848

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 01 February 2023
DOI 10.3389/fnins.2023.1052384

OPEN ACCESS

EDITED BY
Miguel Ángel Martín-Pascual,
Spanish Public Television, Spain

REVIEWED BY
Hui Zheng,
Zhejiang Normal University, China
Rebecca J. Lepping,
The University of Kansas Medical Center,
United States

*CORRESPONDENCE
Dabin Yeum
✉ dabin.yeum.gr@dartmouth.edu

†These authors have contributed equally to this work and share senior authorship

SPECIALTY SECTION
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

RECEIVED 23 September 2022
ACCEPTED 09 January 2023
PUBLISHED 01 February 2023

CITATION
Yeum D, Jimenez CA, Emond JA, Meyer ML,
Lansigan RK, Carlson DD, Ballarino GA,
Gilbert-Diamond D and Masterson TD (2023)
Differential neural reward reactivity in response
to food advertising medium in children.
Front. Neurosci. 17:1052384.
doi: 10.3389/fnins.2023.1052384

Differential neural reward reactivity in response to food advertising medium in children

Dabin Yeum[1,2]*, Courtney A. Jimenez[3], Jennifer A. Emond[1,4], Meghan L. Meyer[5], Reina K. Lansigan[2], Delaina D. Carlson[2], Grace A. Ballarino[2], Diane Gilbert-Diamond[2,4,6]† and Travis D. Masterson[7]†

[1]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States, [2]Department of Epidemiology, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States, [3]Department of Psychological and Brain Science at Dartmouth College, Hanover, NH, United States, [4]Department of Pediatrics, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States, [5]Department of Psychology, Columbia University, New York, NY, United States, [6]Department of Medicine, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States, [7]Department of Nutritional Sciences, College of Health and Human Development, The Pennsylvania State University, University Park, PA, United States

Introduction: Food cues including food advertisements (ads) activate brain regions related to motivation and reward. These responses are known to correlate with eating behaviors and future weight gain. The objective of this study was to compare brain responses to food ads by different types of ad mediums, dynamic (video) and static (images), to better understand how medium type impacts food cue response.

Methods: Children aged 9–12 years old were recruited to complete a functional magnetic resonance imaging (fMRI) paradigm that included both food and non-food dynamic and static ads. Anatomical and functional images were preprocessed using the fMRIPrep pipeline. A whole-brain analysis and a targeted region-of-interest (ROI) analysis for reward regions (nucleus accumbens, orbitofrontal cortex, amygdala, insula, hypothalamus, ventral tegmental area, substantia nigra) were conducted. Individual neural responses to dynamic and static conditions were compared using a paired *t*-test. Linear mixed-effects models were then constructed to test the differential response by ad condition after controlling for age, sex, BMI-z, physical activity, and % of kcal consumed of a participant's estimated energy expenditure in the pre-load prior to the MRI scan.

Results: A total of 115 children (mean=10.9 years) completed the fMRI paradigm. From the ROI analyses, the right and left hemispheres of the amygdala and insula, and the right hemisphere of the substantia nigra showed significantly higher responses for the dynamic food ad medium after controlling for covariates and a false discovery rate correction. From the whole-brain analysis, 21 clusters showed significant differential responses between food ad medium including the precuneus, middle temporal gyrus, superior temporal gyrus, and inferior frontal gyrus, and all regions remained significant after controlling for covariates.

Discussion: Advertising medium has unique effects on neural response to food cues. Further research is needed to understand how this differential activation by ad medium ultimately affects eating behaviors and weight outcomes.

KEYWORDS

food cues, fMRI, neural reactivity, visual stimuli, children, static ad, dynamic ad

Introduction

In the United States, the prevalence of obesity was estimated to be 19.7% among children and adolescents aged 2–19 years between 2017 and 2020 according to the National Health Statistics Report (Bryan et al., 2021). Obesity in childhood often continues into adulthood (Singh et al., 2008), and increased body weight is related to chronic disease including type 2 diabetes, cardiovascular disease, and hypertension (National Institutes of Health, 1998; Dietz, 2004; Daniels et al., 2005; Freedman et al., 2007). Many studies have indicated that food marketing plays a critical role in promoting the current obesity epidemic particularly in amplifying the obesogenic environment that younger children in the United States find themselves in Harris et al. (2009), Zimmerman (2011), Ustjanauskas et al. (2013), Harris and Kalnova (2018), Coates et al. (2019), Bragg et al. (2021), and Packer et al. (2022).

For children specifically, media is the primary source of food marketing (World Health Organization Regional Office for Europe, 2016; Anderson, 2018; Edwards et al., 2022). In 2021, children aged 8–12-years spent 4–6 h a day watching entertainment and using apps that include food marketing materials including smartphones, tablets, gaming consoles, TVs, and computers (The American Academy of Child and Adolescent Psychiatry [AACAP], 2020; Rideout et al., 2022). Specifically, between 2019 and 2021 the total amount of time spent on screen media in US children marked a much faster increase than the previous 4 years with the biggest increases seen in screen time activities including watching online videos, using social media, and browsing websites (Rideout et al., 2022). Through media, children are exposed to environmental food cues in the form of food advertisements, frequently of unhealthy foods and beverages (Elsey and Harris, 2016; Frazier and Harris, 2018). According to national Nielsen data, the average U.S. child or adolescent had viewed over 4,300 TV food ads on TV platforms in 2017 with an average of 10 food-related TV ads per day (Frazier and Harris, 2018). Children and adolescents are now constantly exposed to advertisements through digital media on smartphones, tablets, and laptops (World Health Organization Regional Office for Europe, 2016). Online marketing is presented in a decidedly unique way with a combination of both static and dynamic advertising being presented on popular entertainment,[1] social media,[2] and streaming[3] websites (Kelly et al., 2008; Guo et al., 2019). Static advertising is often found in the form of banner, sidebar, and click-through ads, which utilize eye-catching visuals and catch-phrases to increase brand exposure and familiarity that might impact viewers' food choice and intake (Pieters and Wedel, 2004; Edquist et al., 2011). The various layers integrated in this complex marketing environment make it especially challenging to investigate the impact of unhealthy food advertising exposure on younger children (Elsey and Harris, 2016; Tatlow-Golden et al., 2021).

Images in both dynamic and static ads serve as general food cues. Food cues, including those in food marketing, are known to generally activate the dopaminergic mesolimbic pathway of the brain which has implications to health (Stoeckel et al., 2008; Castellanos et al., 2009; Demos et al., 2012; Lawrence et al., 2012; Wagner et al., 2012; Gearhardt et al., 2014; Rapuano et al., 2016, 2017). For instance, regions of the mesolimbic pathway have been previously associated with reward based outcomes (Bassareo et al., 2015) and therefore may influence food intake or other food related behaviors which are tied to reward pathways (Douglass et al., 2017; Liu and Kanoski, 2018; Bond et al., 2020). According to previous functional magnetic resonance imaging (fMRI) research, brain response to food stimuli is also associated with memory, cognitive evaluation of salient stimuli, and overall decision making (Camara, 2008; Schultz, 2015). Previously implicated regions from these studies include nucleus accumbens, orbitofrontal cortex, amygdala, insula, hypothalamus, substantia nigra, and ventral tegmental area (Killgore et al., 2003; Kelley et al., 2005; Berridge, 2009; Kenny, 2011; van der Laan et al., 2011; García-García et al., 2013; Huerta et al., 2014; Meye and Adan, 2014; Sheng et al., 2014; Masterson et al., 2016; van Meer et al., 2016). These regions are also known to be related to food craving and appetitive motivation (Pelchat et al., 2004; Siep et al., 2012; Dietrich et al., 2016; Kahathuduwa et al., 2018; Contreras-Rodríguez et al., 2019). Moreover, activity in these dopaminergic reward regions in response to food cues has been previously associated with increased food consumption (Lawrence et al., 2012; Masterson et al., 2019b; Gearhardt et al., 2020) and weight gain (Stice et al., 2010; Demos et al., 2012; Gearhardt et al., 2014).

Thus far, previous studies focused on comparison of food cues such as food logos vs. non-food logos (Bruce et al., 2013, 2014; Fehse et al., 2017; Masterson et al., 2019b) and food commercials vs. non-food commercials (Gearhardt et al., 2014, 2020; Bruce et al., 2016; Rapuano et al., 2016, 2017; Masterson et al., 2019a); however, no studies have compared the differential reward activation between static and dynamic advertising mediums. While static advertisements present powerful graphics and eye-catching texts that elicit reward responses (Pieters and Wedel, 2004; Edquist et al., 2011), dynamic advertisements also include narrative and emotional aspects and soundtracks, that may heighten the reward response (Emond et al., 2015). Understanding if one type of medium is more effective at producing a neural response indicative of reward reactivity is critical to inform policies related to child-directed food marketing online.

Therefore, we sought to compare brain responses between food-related dynamic and static advertisements using a fMRI paradigm. We hypothesized that dynamic advertisements would elicit greater food-specific neural response in the reward regions than static advertisements. We conducted a targeted region of interest (ROI) analysis using *a priori* ROIs identified from previous literature as we hypothesized reward regions of the brain would be highly relevant to food cues. Additionally, we also conducted a non-hypothesis-driven whole-brain exploratory analysis to consider differences that may be related to medium type, independent of content.

Materials and methods

Study participants

This study utilized data from a larger study with the primary goal to investigate the relationship between children's response to dynamic food ads and snack food intake. A total of 146 children between the ages of 9 and 12 were enrolled. Children were recruited through community fliers, listservs, and events. Participants were excluded from the present study if they had allergies or dietary restrictions related to the foods served in the study meals, metal in or on the body, claustrophobia, psychiatric or neurological disorders, had appetite-or attention-altering disorders, or if they were not fluent in English. Additionally, participants were excluded if they

---

[1] http://youtube.com/

[2] http://instagram.com/

[3] https://www.twitch.tv/

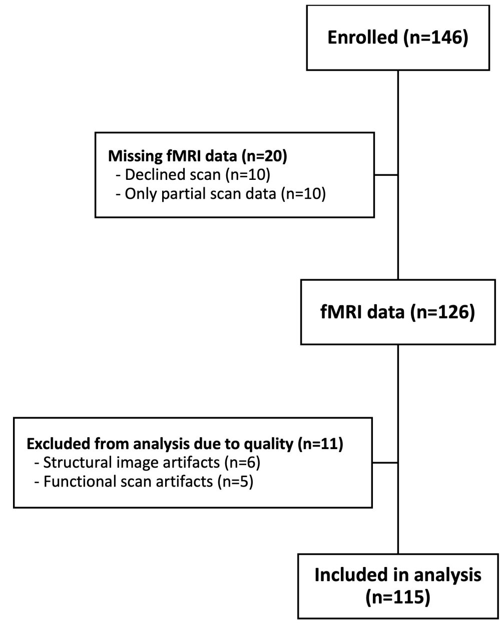


FIGURE 1
Participant flow chart.

had an immediate family member who had previously participated in the study. Caregivers and children provided informed consent and assent, and participants received monetary compensation for participating. Dartmouth's Committee for the Protection of Human Subjects approved all study protocols. A participant flow chart is shown in Figure 1. Data from 31 children were excluded due to: refusal to be scanned (n = 10); incomplete scans (n = 10); excessive movement in the scanner (n = 6); and functional scan artifacts (n = 5). Therefore, the final analysis sample consisted of 115 participants (66 male; mean age (SD) = 10.9 (1.17) years).

Study overview

Each child, accompanied by a parent, was scheduled for a lunchtime (11:00 a.m. to 1:00 p.m.) or dinnertime (4:00 p.m. to 6:00 p.m.) appointment when they completed a brief lab assessment and then the fMRI scan. During the lab assessment, children were provided with a standardized pre-load meal to ensure fullness in line with the protocol of the larger study. Hunger level was assessed prior to the scan using the Freddy Fullness scale (Keller et al., 2006), a validated scale for estimating satiety in children. A trained research staff member then measured children's height and weight and administered a battery of questionnaires to the parent. Following the parent questionnaires, the fMRI paradigm and protocol were explained to both children and their parents prior to conducting the scanning protocol.

Scanning paradigm

The scan protocol consisted of a series of videos and images that were presented to participants using E-Prime (Psychology Software Tools Inc., Sharpsburg, PA, United States), herein called "runs." The main portion of the protocol was designed to simulate a normal television viewing session and included four commercial runs and three 5-min TV show runs during which the children watched a popular age-appropriate science show (MythBusters). Following this naturalistic paradigm, participants viewed four runs of static ads. Participants viewed all videos and images on a screen through a mirror mounted to the head coil. Each MRI scan consisted of 12 functional runs total. For the purposes of this analysis, only eight runs (the four runs of dynamic ads and the four runs of static ads) were included in the analysis (Figure 2). Participants' structural scans were completed during one of the TV shows runs.

Each functional run was approximately 5-min in length. Each run began and ended with a 15 s presentation of a fixation cross. To promote participant engagement, a trained research staff verbally talked with the participants between each run and asked if they would like to continue. In the dynamic ad runs, five food and five non-food TV commercials were presented which alternated in an AB pattern (Smith et al., 2007; Maus et al., 2010). The block pattern for each run was randomized (AB or BA) along with which commercials were played within each block was also randomized. Each commercial was approximately 15 s in length. Static ad runs were similarly randomized but consisted of 10 food and 10 non-food static ads. Each ad image was displayed for 7.5 s followed immediately by another image of the same type (food or non-food) which was also displayed for 7.5 s for a total exposure time of 15 s. This back-to-back display of ads was arranged so that the ad exposure period matched that of the dynamic ad length (15 s). An additional 15 s fixation cross block was placed in the middle of each static ad run to ensure equal amounts of exposure to all stimulus types (i.e., static, dynamic, and fixation).

Image acquisition

Scanning was performed on a 3.0T Siemens MAGNETOM Prisma MRI scanner with a 32-channel head coil. The following parameters were used to obtain a T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) structural scan for each participant: Echo time (TE) = 2.32 ms; repetition time (TR) = 2,300 ms; flip angle = 8°; matrix size = 256 × 256 mm; field of view = 240 × 240 mm; 192 slices; slice thickness = 0.9 mm; voxel size = 0.9 × 0.9 × 0.9 mm. Functional images were acquired using a T2*-weighted echo planar imaging (EPI) sequence using the following parameters: TE = 33 ms; TR = 1,250 ms; flip angle = 64°; matrix size = 96 × 96; field of view = 240 × 240 mm; 56 slices; slice thickness = 2.5 mm; voxel size = 2.5 × 2.5 × 2.5 mm. Eight functional runs (4 dynamic ad runs with 144 TRs and 4 static ad runs with 157 TRs each) were included in the analysis for each participant.

Stimuli

For the ads presented in the dynamic condition, contemporary food, and toy commercials were selected based on relevance to the age group. In a previous study, children rated the commercials included here for interest and excitement and reported no overall difference in interest and excitement between the food and non-food commercials (Rapuano et al., 2016). The ads selected for the static condition were matched to the products and companies of the ads used in the dynamic condition and therefore expected to be similarly relevant and exciting for this age group.

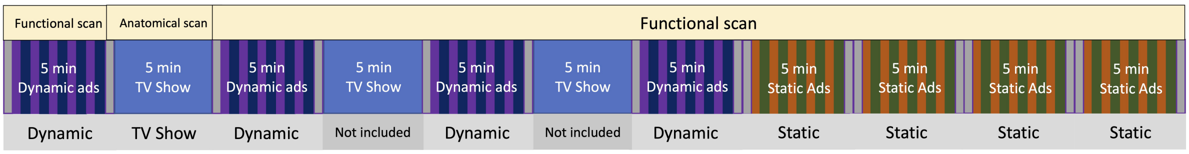


FIGURE 2
MRI paradigm. Food and non-food ads were presented which alternated in an AB/BA (randomized order) pattern during the dynamic and static runs (shown in two different colors). Each run contained five food and five non-food ads, and each ad was 15 s in length. Each fixation cross block between runs was 15 s in length. For the four dynamic runs, a total of 20 food ads (5 min) and 20 non-food ads (5 min) were presented; for the four static runs, a total of 40 food ads (5 min) and 40 non-food ads (5 min) were presented to the study participants.

Model covariates

Participant height and weight were measured using a Seca 703 Medical Scale and Seca 264 Stadiometer (Hamburg, Germany). Body mass index (BMI) was calculated using U.S. Center for Disease Control (CDC) 2000 age- and sex-specific distributions (Centers for Disease Control and Prevention [CDC], 2002). Healthy weight was defined as <85th percentile, overweight was defined as ≥85th – <95th percentile and obese was defined as ≥95th percentile.

The estimated daily energy requirement (EER) was calculated for each child according to Institute of Medicine guidelines given the child's sex, age, and measured height and weight (Institute of Medicine (U.S.), 2005). Per the larger study, children consumed a standardized pre-load meal that consisted of macaroni and cheese, apple sauce, corn, milk, and water; the calories of each item provided were defined to meet, in total, ∼25% of each child's calculated EER while maintaining a standardized ratio across items. The total kcal consumed at pre-load was then divided by the child's EER to derive the percent kcal (per EER) consumed at pre-load.

Caregivers reported child physical activity by answering, "During the past 7 days, on how many days was a child active for a total of at least 60 min per day?" for child's physical activity, which was categorized as "No days," "1 day," "2–3 days," "4–5 days," and "6–7 days." Child's screen time on a typical weekday and a weekend day was separately reported by caregivers. The total screen time per week was calculated by multiplying the screen time on a weekday by five and a weekend day by two. The caregiver also reported their child's date of birth, biological sex, race, and ethnicity.

MRI pre-processing

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 1.2.5 [(Esteban et al., 2017, 2019); RRID:SCR_016216], which is based on *Nipype* 1.1.6 [(Gorgolewski et al., 2011; Esteban et al., 2022); RRID:SCR_002502].

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) using N4BiasFieldCorrection, (Tustison et al., 2010) and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped using antsBrainExtraction.sh (ANTs 2.2.0), using Open Access Series of Imaging Studies (OASIS) as target template. Brain surfaces were reconstructed using recon-all [FreeSurfer 6.0.1, RRID:SCR_001847, (Dale et al., 1999)], and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle [RRID:SCR_002438, (Klein et al., 2017)]. Spatial normalization to the International Consortium for Brain Mapping (ICBM) 152 Non-linear Asymmetrical template version 2009c [(Fonov et al., 2009), RRID:SCR_008796] was performed through non-linear registration with antsRegistration [ANTs 2.2.0, RRID:SCR_004757, (Avants et al., 2008)], using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM), and gray-matter (GM) was performed on the brain-extracted T1w using fast [FSL 5.0.9, RRID:SCR_002823, (Zhang et al., 2001)].

Functional data preprocessing

For each of the eight blood oxygenation-level dependent (BOLD) runs found per participant (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl, 2009). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt [FSL 5.0.9, (Jenkinson et al., 2002)]. BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207 [(Cox and Hyde, 1997), RRID:SCR_005927]. The BOLD time-series, were resampled to surfaces on the following spaces: *fsaverage5*. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled to MNI152Nlin2009cAsym standard space, generating a *preprocessed BOLD run in MNI152Nlin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), delta variation signal (DVARS) and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* [following the definitions by Power et al. (2014)]. The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction [*CompCor*, (Behzadi et al., 2007)]. Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128 s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). Six tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, six components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and template spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Statistical analysis

Subject-level analysis

Following pre-processing, participants' individual fMRI data were analyzed using the NLTools Python package (Chang et al., 2018). For each participant's subject-level analysis, a general linear model (GLM) was performed by constructing a design matrix, running hemodynamic response function (HRF) convolution, adding nuisance variables related to intercepts, linear and quadratic trends, motion covariates (24 motion parameters; six demeaned realignment parameters, their squares, their derivatives, and their squared derivatives), and motion spikes [motion spikes between successive TRs and global spikes that exceed an overall average intensity change by 2.5 standard deviations (SD) as a threshold]. Data were spatially smoothed using a 6 mm full-width at half maximum (FWHM) Gaussian kernel. Structural images and any functional runs that did not pass standard visual inspection due to excessive motion artifacts were excluded from the analysis (n = 11). From standard visual inspection, 12 participants (∼10%) had data from one functional run excluded from the dynamic condition; 20 participants (∼17%) had data from one functional run excluded from the static condition; three participants (∼2%) were missing one run each in both dynamic and static runs. However, they did not meet the criteria for overall exclusion from the analysis, therefore, three runs were included for the individual-level regression. Any functional run in which the number of spikes lead to greater than 25% of the total run length to be censored were excluded from further analyses (n = 0). Furthermore, any participants for whom the number of runs was missing or excluded in more than 15% in either ad condition were excluded from further analyses (n = 10). Individual-level regression coefficient (beta) maps were averaged across functional runs and created contrasting the activation between food ads and non-food ads to generate beta maps that reflected the unique effect of food ads for both the dynamic and static conditions. All subject-level maps were then moved forward to both whole-brain and targeted region of interest analyses.

Region of interest analyses

For the *a priori* region of interest (ROI) analysis, seven ROIs were selected as candidate reward regions based on previous literature (Kenny, 2011): the nucleus accumbens (NAcc), orbitofrontal cortex (OFC), amygdala, insula, hypothalamus, ventral tegmental area, and substantia nigra. Masks of these bilateral regions were generated using the Talairach Daemon and Montreal Neurological Institute (MNI) atlas using AFNI [Analysis of Functional NeuroImages version: 21.0.06 (Cox and Hyde, 1997)]. The mask of the ventral tegmental area was defined by the sphere with a radius of 5mm centered at MNI coordinate [4, −16, −10] (Carter, 2009). The ROI masks are shown in **Figure 3**. The beta coefficients within each ROI mask were extracted from the individual's food > non-food beta maps from the dynamic and static conditions separately, then were averaged using 3dMaskave in AFNI. Beta values were then imported and analyzed using the R Language and Environment for Statistical Computing, version 4.0.2 (R Core Team, 2020). Using the ROI-specific beta maps, linear mixed effects (LME) models were used to test the neural response by ad condition (dynamic or static condition); a random effect was included at the participant level to account for the repeated measures within participant. Analyses were repeated, controlling for age, sex, BMI-z, physical activity, and% of kcal consumed of a participant's estimated energy expenditure in the pre-load. LME models were fitted using the lmerTest package in R. A false discovery rate (FDR) multiple comparison correction was performed on the p-values in seven bilateral regions-of-interest (14 regions total) at $q < 0.05$. An additional exploratory analysis was conducted to examine the interaction between the use of binary weight status (healthy weight vs. with overweight/obesity) and the ad condition (dynamic vs. static) to explore of child weight status modified the effect of ad condition on neural response. Furthermore, a sensitivity analysis was conducted to include a total screen time per week as another covariate to account for participants' exposure to food and non-food cues from the screen time exposure.

Exploratory whole-brain analysis

A whole-brain analysis was conducted using the individual beta maps as input to compare dynamic and static conditions. Initial voxel-wise paired t-test was conducted using individual's food > non-food beta maps for the dynamic and static conditions with 3dttest++ in AFNI. A gray matter mask derived from the Talairach Daemon in AFNI was applied. To determine significance, an initial voxel-wise significance threshold of p-value < 0.001 was applied and was then cluster-corrected using a threshold of a cluster size of k = 180 for an overall p-value < 0.05. These parameters were based on 10,000 Monte-Carlo simulation determined using 3dClustSim from AFNI. The voxel with the peak response in each cluster was identified using the automated anatomical atlas (AAL) in the label4MRI package in R. Additionally, average beta coefficients were extracted from all clusters and extracted for additional analysis using LME models controlling for age, sex, BMI-z, physical activity, and% of kcal consumed of a participant's estimated energy expenditure in the pre-load. An FDR multiple comparison correction was also applied to this set of analyses at $q < 0.05$.

Results

A total of 115 participants were included in the analysis (**Table 1**). Most participants were white (93.0%) and non-Hispanic (93.9%). The average (SD) BMI Z-score was 0.60 (0.96), and 36% of participants were categorized as either having overweight or obesity.

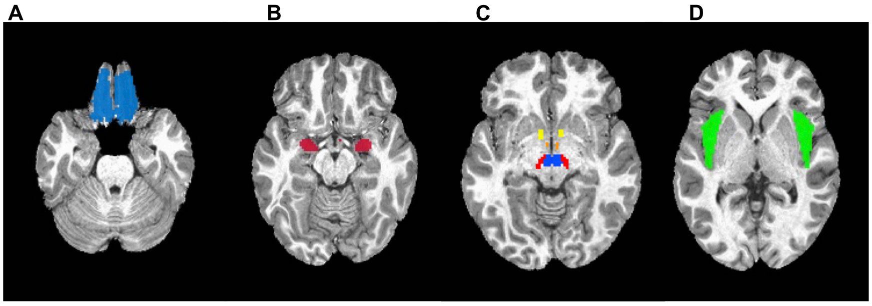


FIGURE 3
Masks used in the region-of-interest (ROI) analysis. **(A)** Orbitofrontal cortex. **(B)** Amygdala. **(C)** Yellow: Nucleus accumbens; Orange: Hypothalamus; Red: Substantia nigra; Blue: Ventral tegmental area. **(D)** Insula.

ROI analyses

The food > non-food contrast maps in the dynamic and static conditions were separately shown in Supplementary Figure 1. Unadjusted and adjusted LME models demonstrated differential reward activation to food ads in the dynamic condition versus static condition (Table 2). Of the 14 bilateral regions tested, six regions showed a statistically significantly higher food-related neural response in the dynamic as compared to the static condition. Specifically, in both unadjusted and adjusted models and after the FDR correction, the right and left amygdala, the right and left insula, and right substantia nigra showed statistically significant higher reward-related response to dynamic ads as compared to static ads. The right ventral tegmental area and left substantia nigra showed significantly higher reward-related response to dynamic ads as compared to static ads before the FDR correction but not after. There was no significant interaction between food ad medium and child weight status. Adjusted models that includes a total screen time exposure as a covariate did not change the results in the 14 bilateral regions (Supplementary Table 1).

TABLE 1 Baseline characteristics of study participants (*N* = 115).

	Mean (SD) or *N* (%)
Age (years)	10.9 (1.17)
Sex	
Female	48 (41.7%)
Male	66 (57.4%)
Prefer not to answer	1 (0.9%)
Ethnicity	
Hispanic	4 (3.5%)
Not Hispanic	108 (93.9%)
Prefer not to answer	2 (1.7%)
Unknown	1 (0.9%)
Race	
Non-white	8 (7.0%)
White	107 (93.0%)
BMI *Z*-score	0.60 (0.96)
BMI category	
Normal weight	74 (64.3%)
Overweight	19 (16.5%)
Obese	22 (19.1%)
Physical activity	
(Active for at least 60 min per day in the past 7 days)	
No days	2 (1.7%)
1 day	38 (33.0%)
2–3 days	42 (36.5%)
4–5 days	29 (25.2%)
6–7 days	1 (0.9%)
Missing	3 (2.6%)

Exploratory whole-brain analysis

Finding from the whole brain analysis are summarized in Table 3 and significant clusters are shown in Figure 4. Fourteen clusters showed a statistically significant higher response to dynamic ads compared to static ads. These included the left gyrus rectus, left middle temporal gyrus, left superior temporal gyrus, left superior frontal gyrus, right precuneus, right inferior frontal gyrus, right supplementary motor area, right cerebellum, and right and left calcarine fissure. Seven clusters showed a higher response to the static ads compared to dynamic ads including the left superior temporal gyrus, left inferior frontal gyrus, right cerebellum, right middle frontal gyrus, and left and right middle occipital gyrus. The results of the adjusted LME models were consistent with unadjusted models and all regions remained significant after FDR correction. Child weight status (healthy weight vs. having overweight and obesity) did not modify the effect of ad condition on neural response.

Discussion

This study examined whether advertising medium (dynamic or static) elicited a differential food-related neural responses in children. Children's media use *via* TV, electronic mobile apps, and social media sites has increased dramatically between 2019 and 2021 (Rideout et al., 2022) and is likely increasing children's exposure to unhealthy food advertising, which could motivate craving and lead to increased food consumption. Previous fMRI studies have compared the neural activation in response to food and non-food cues in either the

TABLE 2 Region-of-interest (ROI) analysis ($N = 115$).

	L/R	Unadjusted LME models[1,2,4]			Adjusted LME models[1,2,3,4]		
		t-value	*p*-value	FDR *q*-value	*t*-value	*p*-value	FDR *q*-value
Nucleus accumbens	R	−1.49	0.140	0.218	−1.49	0.138	0.215
	L	−1.20	0.232	0.325	−1.24	0.218	0.305
Orbitofrontal cortex	R	−0.94	0.351	0.406	−0.97	0.331	0.386
	L	−1.02	0.309	0.393	−1.07	0.287	0.365
Amygdala	R	5.34	**<0.001**	**<0.001**	5.34	**<0.001**	**<0.001**
	L	2.43	**0.016**	0.056	2.43	**0.016**	**0.048**
Insula	R	3.07	**0.003**	**0.019**	3.15	**0.002**	**0.014**
	L	2.31	**0.023**	0.064	2.42	**0.017**	**0.048**
Hypothalamus	R	−0.89	0.377	0.406	−0.89	0.373	0.402
	L	0.09	0.929	0.929	0.10	0.919	0.919
Ventral tegmental area	R	2.07	**0.039**	0.089	2.09	**0.037**	0.086
	L	1.95	0.052	0.091	1.94	0.054	0.094
Substantia nigra	R	2.94	**0.004**	**0.019**	2.94	**0.004**	**0.019**
	L	2.04	**0.044**	0.089	2.04	**0.043**	0.086

[1] Linear mixed effects models.
[2] FDR-corrected threshold at $q < 0.05$ was used.
[3] Covariates include BMI-z, age, sex, % caloric intake at preload, and physical activity.
[4] Bold values represent the statistical significance.

dynamic medium (Gearhardt et al., 2014, 2020; Bruce et al., 2016; Rapuano et al., 2016, 2017; Masterson et al., 2019a) or static medium (Bruce et al., 2013, 2014; Fehse et al., 2017; Masterson et al., 2019b) of advertisement. As it is critical to examine children's brain reactivity to food cues in a naturalistic and ecologically valid presentation, this is the first study that aimed to understand this gap in literature and compared the differential neural reward response to environmental food cues on two types of advertising mediums.

Several dopaminergic reward regions showed higher neural response in the dynamic medium of food ads compared to the static medium of food ads. These regions have been implicated in food cue responsivity previously using static food versus non-food images in the amygdala (Sayer et al., 2016; Melhorn et al., 2018; Luo et al., 2019; Roth et al., 2019) and insula (Schur et al., 2009; Scharmüller et al., 2012; Sayer et al., 2016; Roth et al., 2019), which are brain regions involved in salience, memory, and emotional regulation (Malik et al., 2011; Nummenmaa et al., 2012; Blechert et al., 2016; Roth et al., 2019). The ventral tegmental area and substantia nigra, areas characterized by their dopaminergic neurons and involved in dopamine transmission (Ikemoto, 2007; Kenny, 2011; Ilango et al., 2014), have also shown food cue responsivity to static images (Melhorn et al., 2018). Additionally, previous literature has found that the activation in the amygdala, insula, ventral tegmental area, and substantia nigra are related to food craving and appetitive motivation and regulation, (Pelchat et al., 2004; Siep et al., 2012; Dietrich et al., 2016; Douglass et al., 2017; Bond et al., 2020) which may further drive subsequent food consumption. Our findings suggest that children may be more responsive to food cues when presented in a dynamic medium, and further implies that dynamic ads may be more motivating and engaging to young children than static ads.

From the whole brain analysis, we found increased activation in the precuneus and the occipital lobe in response to the dynamic versus static ad condition, regions that are involved in identifying and detecting the salience of appetitive cue (Tang et al., 2012; Gearhardt et al., 2014; Rapuano et al., 2016). We also observed greater activation in the supplementary motor area to dynamic ads, and this may be indicative of activation of the action observation network where neurons in the motor cortex are active when observing another subject performing the action as if one is performing the action itself (Caspers et al., 2010).

Additionally, the whole brain analysis showed that the bilateral middle occipital gyrus and right middle frontal gyrus had greater activation to the static ad medium. The middle occipital gyrus is involved in visual processing and attention (Murdaugh et al., 2012), and middle frontal gyrus has been implicated in self-control (Harris A. et al., 2013) and response inhibition (Garavan et al., 1999). Studies that compared the neural response to the food versus non-food logos in younger children also found a food-cue-related response in the right occipital cortex (Bruce et al., 2014) and the right middle frontal gyrus (Bruce et al., 2013). Together, findings suggest that there may be activation in self-regulatory behavior networks when static food ads are present. Therefore static ads may play a distinct role in altering behavior. This may be an important consideration for policy makers as both static and dynamic ads are often displayed alongside one another, particularly on online platforms.

Though we did not examine the mechanisms by which dynamic ads elicit a greater reward response than static ads, we hypothesize that the narrative and emotional aspects of the dynamic food ads may increase their saliency compared to their static counterparts. Static ads such as banners on websites utilize powerful visuals and catch-phrases to increase brand exposure and awareness (Pieters and Wedel, 2004; Taylor et al., 2006; Edquist et al., 2011) while dynamic ads present messages using both visual images and audio recordings. A previous study that presented a content analysis of ads for children's packaged foods and beverages showed that child-directed ads highlighted fun, taste,

TABLE 3 Whole-brain analysis ($N = 115$)[1,2].

Paired *t*-test (peak voxel)[3]						Adjusted LME models[4,5,6]	
Brain regions	Side	*X*	*Y*	*Z*	No. voxels[4]	*t*-value	FDR *q*-value
Dynamic > Static condition							
Middle temporal gyrus	L	60	−50	18	8,414	9.32	<**0.001**
	L	−52	−60	24	936	6.15	<**0.001**
Precuneus	R	2	−62	42	3,609	7.26	<**0.001**
	R	50	6	50	388	5.77	<**0.001**
Superior temporal gyrus: temporal pole	L	−36	10	−20	884	7.28	<**0.001**
Superior temporal gyrus	L	−54	−32	14	290	5.78	<**0.001**
Superior frontal gyrus, medial	L	2	62	18	725	5.99	<**0.001**
Inferior frontal gyrus, orbital part	R	52	24	−8	630	6.41	<**0.001**
Inferior frontal gyrus, triangular part	R	46	22	24	244	5.33	<**0.001**
Supplementary motor area	R	6	10	72	471	6.86	<**0.001**
Calcarine fissure and surrounding cortex	R	12	−82	6	391	6.04	<**0.001**
(Occipital lobe)	L	−8	−86	2	218	5.28	<**0.001**
Cerebellum crus^{2}	R	−20	−74	−36	292	5.99	<**0.001**
Gyrus rectus	L	−2	48	−16	222	5.36	<**0.001**
Static > Dynamic condition							
Middle occipital gyrus	L	−32	−90	4	7,102	−7.37	<**0.001**
	R	30	−66	32	4,345	−6.79	<**0.001**
Superior temporal gyrus	L	−62	−12	4	1,025	−8.54	<**0.001**
Cerebellum crus^{1}	R	32	−78	−20	550	−5.67	<**0.001**
Middle frontal gyrus, orbital part	R	34	48	−14	282	−5.26	<**0.001**
Inferior frontal gyrus, triangular part	L	−46	48	8	246	−4.60	<**0.001**
Inferior frontal gyrus, opercular part	L	−44	6	26	207	−4.55	<**0.001**

[1] Cluster size of $k = 180$.
[2] Each voxel has a voxel size of 2.5 × 2.5 × 2.5 mm.
[3] Reported MNI x, y, z coordinate is the peak voxel location within each cluster.
[4] Linear mixed effects models; Covariates include BMI-z, age, sex, % caloric intake at preload, and physical activity.
[5] Models based on the averaged activation in each cluster.
[6] Bold values represent the statistical significance.

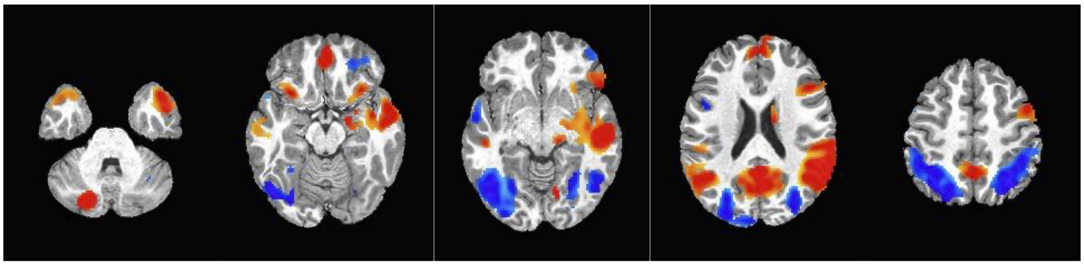


FIGURE 4
Clusters obtained from the whole brain analysis. Red = Activation in the positive direction (Dynamic > Static condition); Blue = Activation in the negative direction (Static > Dynamic condition).

and humor (Emond et al., 2015). The narrative and emotional aspects of the dynamic food ads (versus static food ads) may draw more attention and prompt higher neural responses in younger children.

Collectively, the present findings have implications for the design and interpretation of future food-cue reactivity studies. Researchers should recognize that the medium of food cue presentation may impact the magnitude and location of observed brain response, and comparisons between studies should be cognizant of the food cue medium used. Future studies that assess food marketing exposure in children should distinguish between marketing *via* different mediums as they may have varied effects on eating behavior and other health-related outcomes.

Results from this study lend further support to the need for increased regulations regarding child-directed to food marketing. In the U.S., marketing regulations are voluntary but not statutory,[4] which are consistently scrutinized as insufficient in reducing children's exposure to marketing for unhealthy foods and drinks (Frazier and Harris, 2018; Jensen et al., 2022). The political will to support a government regulatory framework to reduce child-directed food marketing remains a challenge in the US, and more policy options to protect children from unhealthy TV food advertising are still needed (Harris J. L. et al., 2013; Boyland and Harris, 2017; Taillie et al., 2019; Fleming-Milici and Harris, 2020). Nevertheless, recent advances made by the Children's Food and Beverage Advertising Initiative recognize the reach and power of online food marketing, and our findings suggest that the efforts to limit the ad medium may reduce the effectiveness of child-directed food marketing without enforcing more stringent restrictions on ad placement.

Our study has limitations that should be noted. First, because this study was embedded in a larger study, there was lack of randomization in the order of dynamic and static ad runs with the four dynamic runs always preceding the four static runs. We therefore cannot entirely rule out order effects. However, the runs and stimuli within those runs were sufficiently randomized. Additionally, the length of the fMRI scanning paradigm may have affected the scan quality given the younger age group. However, we had a fairly conservative motion screening threshold for excess motion, and no functional run was removed due to excess motion. Additionally, other studies have also successfully conducted fMRI paradigms that are similar or longer in length in similarly aged children (English et al., 2017; Adise et al., 2018; Masterson et al., 2019b). Second, our sample was largely white, non-Hispanic, and of a higher socioeconomic status, and therefore our results may not be generalizable to other populations. Future studies are needed to investigate whether these associations vary by ethnicity, race, and socioeconomic status. Third, the fMRI paradigm was conducted after the consumption of a full meal, which may have differentially reduced the neural response to food cues in the dynamic and static conditions (Goldstone et al., 2009). Future research should also explore concurrent presentation of static and dynamic food ads and or other outlets that combine presentation medium such as advergaming, a dynamic medium often accompanied by static advertising that is frequently used to market unhealthy foods and beverages.

Conclusion

In conclusion, we show that the food advertising medium affects the neural reward response to food cues in children. Because greater neural activity in the regions involved in the dopaminergic pathways in response to food cues has been related to greater consumption (Lawrence et al., 2012; Masterson et al., 2019b; Gearhardt et al., 2020) and weight gain (Stice et al., 2010; Demos et al., 2012; Gearhardt et al., 2014), future research is needed to understand differential impacts of food advertising in different mediums on subsequent eating behaviors and unhealthy weight gain. Given the high current level of media exposure and constant exposure to food marketing across platforms, children are also constantly exposed to a variety of unhealthy food marketing, and it is crucial to understand the health effects of this new level of exposure.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board at Dartmouth College. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

TM and DG-D designed and supervised the data collection and formulated the hypothesis. DY analyzed the data and wrote the first draft of the manuscript. All authors contributed to the various stages of the study, read and edited several draft versions, and approved the final manuscript.

Funding

This study was supported by the National Institutes of Health (5R01HD092604). The study sponsors had no role in the study design, the collection, analysis, and interpretation of data, the writing of the report, and the decision to submit the manuscript for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1052384/full#supplementary-material

---

[4] https://bbbprograms.org/programs/all-programs/cfbai

References

Adise, S., Geier, C. F., Roberts, N. J., White, C. N., and Keller, K. L. (2018). Is brain response to food rewards related to overeating? A test of the reward surfeit model of overeating in children. *Appetite* 128, 167–179. doi: 10.1016/j.appet.2018.06.014

Anderson, M. (2018). *Teens, social media and technology 2018. Pew Research Center: Internet, Science & Tech*. Available online at: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ (accessed December 17, 2022).

Avants, B., Epstein, C., Grossman, M., and Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. doi: 10.1016/j.media.2007.06.004

Bassareo, V., Cucca, F., Musio, P., Lecca, D., Frau, R., and Chiara, G. D. (2015). Nucleus accumbens shell and core dopamine responsiveness to sucrose in rats: Role of response contingency and discriminative/conditioned cues. *Eur. J. Neurosci.* 41, 802–809. doi: 10.1111/ejn.12839

Behzadi, Y., Restom, K., Liau, J., and Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37, 90–101. doi: 10.1016/j.neuroimage.2007.04.042

Berridge, K. C. (2009). 'Liking' and 'wanting' food rewards: Brain substrates and roles in eating disorders. *Physiol. Behav.* 97, 537–550. doi: 10.1016/j.physbeh.2009.02.044

Blechert, J., Klackl, J., Miedl, S. F., and Wilhelm, F. H. (2016). To eat or not to eat: Effects of food availability on reward system activity during food picture viewing. *Appetite* 99, 254–261. doi: 10.1016/j.appet.2016.01.006

Bond, C. W., Trinko, R., Foscue, E., Furman, K., Groman, S. M., Taylor, J. R., et al. (2020). Medial nucleus accumbens projections to the ventral tegmental area control food consumption. *J. Neurosci.* 40, 4727–4738. doi: 10.1523/JNEUROSCI.3054-18.2020

Boyland, E. J., and Harris, J. L. (2017). Regulation of food marketing to children: Are statutory or industry self-governed systems effective? *Public Health Nutr.* 20, 761–764. doi: 10.1017/S1368980017000465

Bragg, M., Lutfeali, S., Greene, T., Osterman, J., and Dalton, M. (2021). How food marketing on Instagram shapes adolescents' food preferences: Online randomized trial. *J. Med. Internet Res.* 23:e28689. doi: 10.2196/28689

Bruce, A. S., Bruce, J. M., Black, W. R., Lepping, R. J., Henry, J. M., Cherry, J. B. C., et al. (2014). Branding and a child's brain: An fMRI study of neural responses to logos. *Soc. Cogn. Affect. Neurosci.* 9, 118–122. doi: 10.1093/scan/nss109

Bruce, A. S., Lepping, R. J., Bruce, J. M., Cherry, J. B. C., Martin, L. E., Davis, A. M., et al. (2013). Brain responses to food logos in obese and healthy weight children. *J. Pediatr.* 162, 759–764.e2. doi: 10.1016/j.jpeds.2012.10.003

Bruce, A. S., Pruitt, S. W., Ha, O.-R., Cherry, J. B. C., Smith, T. R., Bruce, J. M., et al. (2016). The influence of televised food commercials on children's food choices: Evidence from ventromedial prefrontal cortex activations. *J. Pediatr.* 177, 27–32.e1. doi: 10.1016/j.jpeds.2016.06.067

Bryan, S., Afful, J., Carroll, M., Te-Ching, C., Orlando, D., Fink, S., et al. (2021). *NHSR 158. National health and nutrition examination survey 2017–March 2020 pre-pandemic data files*. Hyattsville, MD: National Center for Health Statistics. doi: 10.15620/cdc:106273

Camara, E. (2008). Functional connectivity of reward processing in the brain. *Front. Hum. Neurosci.* 2:19. doi: 10.3389/neuro.09.019.2008

Carter, R. M. (2009). Activation in the VTA and nucleus accumbens increases in anticipation of both gains and losses. *Front. Behav. Neurosci.* 3, 21. doi: 10.3389/neuro.08.021.2009

Caspers, S., Zilles, K., Laird, A. R., and Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage* 50, 1148–1167. doi: 10.1016/j.neuroimage.2009.12.112

Castellanos, E. H., Charboneau, E., Dietrich, M. S., Park, S., Bradley, B. P., Mogg, K., et al. (2009). Obese adults have visual attention bias for food cue images: Evidence for altered reward system function. *Int. J. Obes.* 33, 1063–1073. doi: 10.1038/ijo.2009.138

Centers for Disease Control and Prevention [CDC] (2002). *2000 CDC growth charts for the United States*. Hyattsville, MD: Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

Chang, L., Jolly, E., Cheong, J., Burnashev, A., and Chen, A. (2018). *cosanlab/nltools: 0.3. 11*. Genève: Zenodo.

Coates, A. E., Hardman, C. A., Halford, J. C. G., Christiansen, P., and Boyland, E. J. (2019). Social media influencer marketing and children's food intake: A randomized trial. *Pediatrics* 143:e20182554. doi: 10.1542/peds.2018-2554

Contreras-Rodríguez, O., Cano, M., Vilar-López, R., Rio-Valle, J. S., Verdejo-Román, J., Navas, J. F., et al. (2019). Visceral adiposity and insular networks: Associations with food craving. *Int. J. Obes.* 43, 503–511. doi: 10.1038/s41366-018-0173-3

Cox, R. W., and Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR Biomed.* 10, 171–178. doi: 10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. *Neuroimage* 9, 179–194. doi: 10.1006/nimg.1998.0395

Daniels, S. R., Arnett, D. K., Eckel, R. H., Gidding, S. S., Hayman, L. L., Kumanyika, S., et al. (2005). Overweight in children and adolescents. *Circulation* 111, 1999–2012. doi: 10.1161/01.cir.0000161369.71722.10

Demos, K. E., Heatherton, T. F., and Kelley, W. M. (2012). Individual differences in nucleus accumbens activity to food and sexual images predict weight gain and sexual behavior. *J. Neurosci.* 32, 5549–5552. doi: 10.1523/jneurosci.5958-11.2012

Dietrich, A., Hollmann, M., Mathar, D., Villringer, A., and Horstmann, A. (2016). Brain regulation of food craving: Relationships with weight status and eating behavior. *Int. J. Obes.* 40, 982–989. doi: 10.1038/ijo.2016.28

Dietz, W. H. (2004). Overweight in childhood and adolescence. *N. Engl. J. Med.* 350, 855–857. doi: 10.1056/nejmp048008

Douglass, A. M., Kucukdereli, H., Ponserre, M., Markovic, M., Gründemann, J., Strobel, C., et al. (2017). Central amygdala circuits modulate food consumption through a positive-valence mechanism. *Nat. Neurosci.* 20, 1384–1394. doi: 10.1038/nn.4623

Edquist, J., Horberry, T., Hosking, S., and Johnston, I. (2011). Effects of advertising billboards during simulated driving. *Appl. Ergon.* 42, 619–626. doi: 10.1016/j.apergo.2010.08.013

Edwards, C. G., Pollack, C. C., Pritschet, S. J., Haushalter, K., Long, J. W., and Masterson, T. D. (2022). Prevalence and comparisons of alcohol, candy, energy drink, snack, soda, and restaurant brand and product marketing on Twitch, Facebook Gaming and YouTube gaming. *Public Health Nutr.* 25, 1–12. doi: 10.1017/S1368980021004420

Elsey, J. W., and Harris, J. L. (2016). Trends in food and beverage television brand appearances viewed by children and adolescents from 2009 to 2014 in the USA. *Public Health Nutr.* 19, 1928–1933. doi: 10.1017/S1368980015003274

Emond, J. A., Smith, M. E., Mathur, S. J., Sargent, J. D., and Gilbert-Diamond, D. (2015). Children's food and beverage promotion on television to parents. *Pediatrics* 136, 1095–1102. doi: 10.1542/peds.2015-2853

English, L. K., Fearnbach, S. N., Wilson, S. J., Fisher, J. O., Savage, J. S., Rolls, B. J., et al. (2017). Food portion size and energy density evoke different patterns of brain activation in children. *Am. J. Clin. Nutr.* 105, 295–305. doi: 10.3945/ajcn.116.136903

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., et al. (2017). *poldracklab/fmriprep: 1.0.0-rc5*. Genève: Zenodo. doi: 10.5281/zenodo.996169

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., et al. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nat. Methods* 16, 111–116. doi: 10.1038/s41592-018-0235-4

Esteban, O., Markiewicz, C. J., Burns, C., Goncalves, M., Jarecka, D., Ziegler, E., et al. (2022). *nipy/nipype: 1.7.1*. Genève: Zenodo. doi: 10.5281/ZENODO.596855

Fehse, K., Simmank, F., Gutyrchik, E., and Sztrókay-Gaul, A. (2017). Organic or popular brands—food perception engages distinct functional pathways. An fMRI study. *Cogent. Psychol.* 4:1284392. doi: 10.1080/23311908.2017.1284392

Fleming-Milici, F., and Harris, J. L. (2020). Food marketing to children in the United States: Can industry voluntarily do the right thing for children's health? *Physiol. Behav.* 227:113139. doi: 10.1016/j.physbeh.2020.113139

Fonov, V., Evans, A., McKinstry, R., Almli, C., and Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* 47, S102. doi: 10.1016/S1053-8119(09)70884-5

Frazier, W., and Harris, J. (2018). *Trends in television food advertising to young people: 2017 update*. Available online at: http://uconnruddcenter.org/files/Pdfs/TVAdTrends2018_Final.pdf (accessed January 20, 2021).

Freedman, D. S., Mei, Z., Srinivasan, S. R., Berenson, G. S., and Dietz, W. H. (2007). Cardiovascular risk factors and excess adiposity among overweight children and adolescents: The bogalusa heart study. *J. Pediatr.* 150, 12–17.e2. doi: 10.1016/j.jpeds.2006.08.042

Garavan, H., Ross, T. J., and Stein, E. A. (1999). Right hemispheric dominance of inhibitory control: An event-related functional MRI study. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8301–8306. doi: 10.1073/pnas.96.14.8301

García-García, I., Narberhaus, A., Marqués-Iturria, I., Garolera, M., Rădoi, A., Segura, B., et al. (2013). Neural responses to visual food cues: Insights from functional magnetic resonance imaging: Brain and sight of food. *Eur. Eat. Disord. Rev.* 21, 89–98. doi: 10.1002/erv.2216

Gearhardt, A. N., Yokum, S., Harris, J. L., Epstein, L. H., and Lumeng, J. C. (2020). Neural response to fast food commercials in adolescents predicts intake. *Am. J. Clin. Nutr.* 111, 493–502. doi: 10.1093/ajcn/nqz305

Gearhardt, A. N., Yokum, S., Stice, E., Harris, J. L., and Brownell, K. D. (2014). Relation of obesity to neural activation in response to food commercials. *Soc. Cogn. Affect. Neurosci.* 9, 932–938. doi: 10.1093/scan/nst059

Goldstone, A. P., Prechtl de Hernandez, C. G., Beaver, J. D., Muhammed, K., Croese, C., Bell, G., et al. (2009). Fasting biases brain reward systems towards high-calorie foods. *Eur. J. Neurosci.* 30, 1625–1635. doi: 10.1111/j.1460-9568.2009.06949.x

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front. Neuroinform.* 5:13. doi: 10.3389/fninf.2011.00013

Greve, D. N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48, 63–72. doi: 10.1016/j.neuroimage.2009.06.060

Guo, F., Ye, G., Hudders, L., Lv, W., Li, M., and Duffy, V. G. (2019). Product placement in mass media: A review and bibliometric analysis. *J. Advertising* 48, 215–231. doi: 10.1080/00913367.2019.1567409

Harris, A., Hare, T., and Rangel, A. (2013). Temporally dissociable mechanisms of self-control: Early attentional filtering versus late value modulation. *J. Neurosci.* 33, 18917–18931. doi: 10.1523/JNEUROSCI.5816-12.2013

Harris, J. L., and Kalnova, S. S. (2018). Food and beverage TV advertising to young children: Measuring exposure and potential impact. *Appetite* 123, 49–55. doi: 10.1016/j.appet.2017.11.110

Harris, J. L., Pomeranz, J. L., Lobstein, T., and Brownell, K. D. (2009). A crisis in the marketplace: How food marketing contributes to childhood obesity and what can be done. *Annu. Rev. Public Health* 30, 211–225. doi: 10.1146/annurev.publhealth.031308.100304

Harris, J. L., Sarda, V., Schwartz, M. B., and Brownell, K. D. (2013). Redefining "child-directed advertising" to reduce unhealthy television food advertising. *Am. J. Prev. Med.* 44, 358–364. doi: 10.1016/j.amepre.2012.11.039

Huerta, C. I., Sarkar, P. R., Duong, T. Q., Laird, A. R., and Fox, P. T. (2014). Neural bases of food perception: Coordinate-based meta-analyses of neuroimaging studies in multiple modalities: Coordinate-based meta-analyses of neuroimaging studies in multiple modalities. *Obesity* 22, 1439–1446. doi: 10.1002/oby.20659

Ikemoto, S. (2007). Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Res. Rev.* 56, 27–78. doi: 10.1016/j.brainresrev.2007.05.004

Ilango, A., Kesner, A. J., Keller, K. L., Stuber, G. D., Bonci, A., and Ikemoto, S. (2014). Similar roles of substantia nigra and ventral tegmental dopamine neurons in reward and aversion. *J. Neurosci.* 34, 817–822. doi: 10.1523/JNEUROSCI.1703-13.2014

Institute of Medicine (U.S.) (2005). *Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein, and amino acids*. Washington, D.C: National Academies Press. doi: 10.17226/10490

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132

Jensen, M. L., McCann, M., Fleming-Milici, F., Mancini, S., and Harris, J. L. (2022). *Food industry self-regulation: Changes in nutrition of foods and drinks that may be advertised to children.* Hartford, CT: UConn Rudd Center for Food Policy & Health, 63.

Kahathuduwa, C. N., Davis, T., O'Boyle, M., and Binks, M. (2018). Do scores on the food craving inventory and three-factor eating questionnaire correlate with expected brain regions of interest in people with obesity? *Physiol. Behav.* 188, 1–10. doi: 10.1016/j.physbeh.2018.01.018

Keller, K. L., Assur, S. A., Torres, M., Lofink, H. E., Thornton, J. C., Faith, M. S., et al. (2006). Potential of an analog scaling device for measuring fullness in children: Development and preliminary testing. *Appetite* 47, 233–243. doi: 10.1016/j.appet.2006.04.004

Kelley, A., Baldo, B., Pratt, W., and Will, M. (2005). Corticostriatal-hypothalamic circuitry and food motivation: Integration of energy, action and reward. *Physiol. Behav.* 86, 773–795. doi: 10.1016/j.physbeh.2005.08.066

Kelly, B., Bochynska, K., Kornman, K., and Chapman, K. (2008). Internet food marketing on popular children's websites and food product websites in Australia. *Public Health Nutr.* 11, 1180–1187. doi: 10.1017/S1368980008001778

Kenny, P. J. (2011). Reward mechanisms in obesity: New insights and future directions. *Neuron* 69, 664–679. doi: 10.1016/j.neuron.2011.02.016

Killgore, W. D. S., Young, A. D., Femia, L. A., Bogorodzki, P., Rogowska, J., and Yurgelun-Todd, D. A. (2003). Cortical and limbic activation during viewing of high- versus low-calorie foods. *Neuroimage* 19, 1381–1394. doi: 10.1016/S1053-8119(03)00191-5

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., et al. (2017). Mindboggling morphometry of human brains. *PLoS Comput. Biol.* 13:e1005350. doi: 10.1371/journal.pcbi.1005350

Lanczos, C. (1964). Evaluation of noisy data. *J. Soc. Industr. Appl. Math. Ser. B Numeric. Anal.* 1, 76–85. doi: 10.1137/0701007

Lawrence, N. S., Hinton, E. C., Parkinson, J. A., and Lawrence, A. D. (2012). Nucleus accumbens response to food cues predicts subsequent snack consumption in women and increased body mass index in those with reduced self-control. *Neuroimage* 63, 415–422. doi: 10.1016/j.neuroimage.2012.06.070

Liu, C. M., and Kanoski, S. E. (2018). Homeostatic and non-homeostatic controls of feeding behavior: Distinct vs. common neural systems. *Physiol. Behav.* 193, 223–231. doi: 10.1016/j.physbeh.2018.02.011

Luo, S., Alves, J., Hardy, K., Wang, X., Monterosso, J., Xiang, A. H., et al. (2019). Neural processing of food cues in pre-pubertal children. *Pediatr. Obes.* 14:e12435. doi: 10.1111/ijpo.12435

Malik, S., McGlone, F., and Dagher, A. (2011). State of expectancy modulates the neural response to visual food stimuli in humans. *Appetite* 56, 302–309. doi: 10.1016/j.appet.2011.01.005

Masterson, T. D., Bermudez, M. A., Austen, M., Lundquist, E., Pearce, A. L., Bruce, A. S., et al. (2019a). Food commercials do not affect energy intake in a laboratory meal but do alter brain responses to visual food cues in children. *Appetite* 132, 154–165. doi: 10.1016/j.appet.2018.10.010

Masterson, T. D., Kirwan, C. B., Davidson, L. E., and LeCheminant, J. D. (2016). Neural reactivity to visual food stimuli is reduced in some areas of the brain during evening hours compared to morning hours: An fMRI study in women. *Brain Imaging Behav.* 10, 68–78. doi: 10.1007/s11682-015-9366-8

Masterson, T. D., Stein, W. M., Beidler, E., Bermudez, M., English, L. K., and Keller, K. L. (2019b). Brain response to food brands correlates with increased intake from branded meals in children: An fMRI study. *Brain Imaging Behav.* 13, 1035–1048. doi: 10.1007/s11682-018-9919-8

Maus, B., van Breukelen, G. J. P., Goebel, R., and Berger, M. P. F. (2010). Optimization of blocked designs in fMRI studies. *Psychometrika* 75, 373–390. doi: 10.1007/s11336-010-9159-3

Melhorn, S. J., Askren, M. K., Chung, W. K., Kratz, M., Bosch, T. A., Tyagi, V., et al. (2018). FTO genotype impacts food intake and corticolimbic activation. *Am. J. Clin. Nutr.* 107, 145–154. doi: 10.1093/ajcn/nqx029

Meye, F. J., and Adan, R. A. H. (2014). Feelings about food: The ventral tegmental area in food reward and emotional eating. *Trends Pharmacol. Sci.* 35, 31–40. doi: 10.1016/j.tips.2013.11.003

Murdaugh, D. L., Cox, J. E., Cook, E. W., and Weller, R. E. (2012). fMRI reactivity to high-calorie food pictures predicts short- and long-term outcome in a weight-loss program. *Neuroimage* 59, 2709–2721. doi: 10.1016/j.neuroimage.2011.10.071

National Institutes of Health, (1998). Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults–the evidence report. National Institutes of Health. *Obes. Res.* 6 Suppl. 2, 51S–209S.

Nummenmaa, L., Hirvonen, J., Hannukainen, J. C., Immonen, H., Lindroos, M. M., Salminen, P., et al. (2012). Dorsal striatum and its limbic connectivity mediate abnormal anticipatory reward processing in obesity. *PLoS One* 7:e31089. doi: 10.1371/journal.pone.0031089

Packer, J., Russell, S. J., Siovolgyi, G., McLaren, K., Stansfield, C., Viner, R. M., et al. (2022). The impact on dietary outcomes of celebrities and influencers in marketing unhealthy foods to children: A systematic review and meta-analysis. *Nutrients* 14:434. doi: 10.3390/nu14030434

Pelchat, M. L., Johnson, A., Chan, R., Valdez, J., and Ragland, J. D. (2004). Images of desire: Food-craving activation during fMRI. *Neuroimage* 23, 1486–1493. doi: 10.1016/j.neuroimage.2004.08.023

Pieters, R., and Wedel, M. (2004). Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *J. Market.* 68, 36–50. doi: 10.1509/jmkg.68.2.36.27794

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* 84, 320–341. doi: 10.1016/j.neuroimage.2013.08.048

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rapuano, K. M., Huckins, J. F., Sargent, J. D., Heatherton, T. F., and Kelley, W. M. (2016). Individual differences in reward and somatosensory-motor brain regions correlate with adiposity in adolescents. *Cereb. Cortex* 26, 2602–2611. doi: 10.1093/cercor/bhv097

Rapuano, K. M., Zieselman, A. L., Kelley, W. M., Sargent, J. D., Heatherton, T. F., and Gilbert-Diamond, D. (2017). Genetic risk for obesity predicts nucleus accumbens size and responsivity to real-world food cues. *Proc. Natl. Acad. Sci. U.S.A.* 114, 160–165. doi: 10.1073/pnas.1605548113

Rideout, V., Peebles, A., Supreet, M., and Michael, R. (2022). *Common sense census: Media use by tweens and teens, 2021*. San Francisco, CA: Common Sense Media.

Roth, C. L., Melhorn, S. J., Elfers, C. T., Scholz, K., De Leon, M. R. B., Rowland, M., et al. (2019). Central nervous system and peripheral hormone responses to a meal in children. *J. Clin. Endocrinol. Metab.* 104, 1471–1483. doi: 10.1210/jc.2018-01525

Sayer, R. D., Amankwaah, A. F., Tamer, G. G., Chen, N., Wright, A. J., Tregellas, J. R., et al. (2016). Effects of dietary protein and fiber at breakfast on appetite, ad libitum energy intake at lunch, and neural responses to visual food stimuli in overweight adults. *Nutrients* 8:E21. doi: 10.3390/nu8010021

Scharmüller, W., Übel, S., Ebner, F., and Schienle, A. (2012). Appetite regulation during food cue exposure: A comparison of normal-weight and obese women. *Neurosci. Lett.* 518, 106–110. doi: 10.1016/j.neulet.2012.04.063

Schultz, W. (2015). Neuronal reward and decision signals: From theories to data. *Physiol. Rev.* 95, 853–951. doi: 10.1152/physrev.00023.2014

Schur, E. A., Kleinhans, N. M., Goldberg, J., Buchwald, D., Schwartz, M. W., and Maravilla, K. (2009). Activation in brain energy regulation and reward centers by food cues varies with choice of visual stimulus. *Int. J. Obes.* 33, 653–661. doi: 10.1038/ijo.2009.56

Sheng, Z., Santiago, A. M., Thomas, M. P., and Routh, V. H. (2014). Metabolic regulation of lateral hypothalamic glucose-inhibited orexin neurons may influence midbrain reward neurocircuitry. *Mol. Cell Neurosci.* 62, 30–41. doi: 10.1016/j.mcn.2014.08.001

Siep, N., Roefs, A., Roebroeck, A., Havermans, R., Bonte, M., and Jansen, A. (2012). Fighting food temptations: The modulating effects of short-term cognitive reappraisal, suppression and up-regulation on mesocorticolimbic activity related to appetitive motivation. *Neuroimage* 60, 213–220. doi: 10.1016/j.neuroimage.2011.12.067

Singh, A. S., Mulder, C., Twisk, J. W. R., van Mechelen, W., and Chinapaw, M. J. M. (2008). Tracking of childhood overweight into adulthood: A systematic review of the literature. *Obes. Rev.* 9, 474–488. doi: 10.1111/j.1467-789X.2008.00475.x

Smith, S., Jenkinson, M., Beckmann, C., Miller, K., and Woolrich, M. (2007). Meaningful design and contrast estimability in FMRI. *Neuroimage* 34, 127–136. doi: 10.1016/j.neuroimage.2006.09.019

Stice, E., Yokum, S., Bohon, C., Marti, N., and Smolen, A. (2010). Reward circuitry responsivity to food predicts future increases in body mass: Moderating effects of DRD2 and DRD4. *Neuroimage* 50, 1618–1625. doi: 10.1016/j.neuroimage.2010.01.081

Stoeckel, L. E., Weller, R. E., Cook, E. W., Twieg, D. B., Knowlton, R. C., and Cox, J. E. (2008). Widespread reward-system activation in obese women in response to pictures of high-calorie foods. *Neuroimage* 41, 636–647. doi: 10.1016/j.neuroimage.2008.02.031

Taillie, L. S., Busey, E., Stoltze, F. M., and Dillman Carpentier, F. R. (2019). Governmental policies to reduce unhealthy food marketing to children. *Nutr. Rev.* 77, 787–816. doi: 10.1093/nutrit/nuz021

Tang, D. W., Fellows, L. K., Small, D. M., and Dagher, A. (2012). Food and drug cues activate similar brain regions: A meta-analysis of functional MRI studies. *Physiol. Behav.* 106, 317–324. doi: 10.1016/j.physbeh.2012.03.009

Tatlow-Golden, M., Jewell, J., Zhiteneva, O., Wickramasinghe, K., Breda, J., and Boyland, E. (2021). Rising to the challenge: Introducing protocols to monitor food marketing to children from the World Health Organization Regional Office for Europe. *Obes. Rev.* 22:e13212. doi: 10.1111/obr.13212

Taylor, C. R., Franke, G. R., and Bang, H.-K. (2006). Use and effectiveness of billboards: Perspectives from selective-perception theory and retail-gravity models. *J. Advertising* 35, 21–34. doi: 10.2753/JOA0091-3367350402

The American Academy of Child and Adolescent Psychiatry [AACAP] (2020). *Screen time and children*. Available online at: https://www.aacap.org/AACAP/Families_and_Youth/Facts_for_Families/FFF-Guide/Children-And-Watching-TV-054.aspx (accessed May 24, 2022).

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010). N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging* 29, 1310–1320. doi: 10.1109/TMI.2010.2046908

Ustjanauskas, A. E., Harris, J. L., and Schwartz, M. B. (2013). Food and beverage advertising on children's web sites. *Pediatr. Obes.* 9, 362–372. doi: 10.1111/j.2047-6310.2013.00185.x

van der Laan, L. N., de Ridder, D. T. D., Viergever, M. A., and Smeets, P. A. M. (2011). The first taste is always with the eyes: A meta-analysis on the neural correlates of processing visual food cues. *Neuroimage* 55, 296–303. doi: 10.1016/j.neuroimage.2010.11.055

van Meer, F., Charbonnier, L., and Smeets, P. A. M. (2016). Food decision-making: Effects of weight status and age. *Curr. Diab. Rep.* 16:84. doi: 10.1007/s11892-016-0773-z

Wagner, D. D., Boswell, R. G., Kelley, W. M., and Heatherton, T. F. (2012). Inducing negative affect increases the reward value of appetizing foods in dieters. *J. Cognit. Neurosci.* 24, 1625–1633. doi: 10.1162/jocn_a_00238

World Health Organization Regional Office for Europe (2016). *Tackling food marketing to children in a digital workd: Trans-disciplinary perspectives*. Copenhagen: World Health Organization.

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20, 45–57. doi: 10.1109/42.906424

Zimmerman, F. J. (2011). Using marketing muscle to sell fat: The rise of obesity in the modern economy. *Annu. Rev. Public Health* 32, 285–306. doi: 10.1146/annurev-publhealth-090810-182502

Frontiers | Frontiers in Neuroscience

TYPE Correction
PUBLISHED 23 March 2023
DOI 10.3389/fnins.2023.1170370

OPEN ACCESS

EDITED AND REVIEWED BY
Celia Andreu-Sánchez,
Autonomous University of Barcelona, Spain

*CORRESPONDENCE
Dabin Yeum
✉ dabin.yeum.gr@dartmouth.edu

†These authors have contributed equally to this work and share senior authorship

SPECIALTY SECTION
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

RECEIVED 20 February 2023
ACCEPTED 08 March 2023
PUBLISHED 23 March 2023

CITATION
Yeum D, Jimenez CA, Emond JA, Meyer ML, Lansigan RK, Carlson DD, Ballarino GA, Gilbert-Diamond D and Masterson TD (2023) Corrigendum: Differential neural reward reactivity in response to food advertising medium in children.
Front. Neurosci. 17:1170370.
doi: 10.3389/fnins.2023.1170370

Corrigendum: Differential neural reward reactivity in response to food advertising medium in children

Dabin Yeum[1,2]*, Courtney A. Jimenez[3], Jennifer A. Emond[1,4], Meghan L. Meyer[5], Reina K. Lansigan[2], Delaina D. Carlson[2], Grace A. Ballarino[2], Diane Gilbert-Diamond[2,4,6]† and Travis D. Masterson[7]†

[1]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States, [2]Department of Epidemiology, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States, [3]Department of Psychological and Brain Science at Dartmouth College, Hanover, NH, United States, [4]Department of Pediatrics, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States, [5]Department of Psychology, Columbia University, New York, NY, United States, [6]Department of Medicine, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States, [7]Department of Nutritional Sciences, College of Health and Human Development, The Pennsylvania State University, University Park, PA, United States

KEYWORDS

food cues, fMRI, neural reactivity, visual stimuli, children, static ad, dynamic ad

A corrigendum on
Differential neural reward reactivity in response to food advertising medium in children

by Yeum, D., Jimenez, C. A., Emond, J. A., Meyer, M. L., Lansigan, R. K., Carlson, D. D., Ballarino, G. A., and Masterson, T. D. (2023). *Front. Neurosci.* 17, 1052384. doi: 10.3389/fnins.2023.1052384

In the original article, there was an error in Figure 3 as published. An error was caught with the ventral tegmental area (VTA) masks. We have identified the MNI coordinates for VTA, created a mask, and updated the analysis.

The corrected Figure 3 and its caption appear below.

Following the incorrect mask used for VTA, there was an error in Table 2/Supplementary Table 1 as published. Using the corrected mask, *t*-, *p*-, *q*-values for VTA changed. *q*-values (FDR-corrected statistical significance) for some other regions slightly changed because they are derived using the *p*-values of the multiple tests, but did not affect the interpretation of statistical significance. The VTA was statistically significantly associated with the dynamic advertising condition, but the statistical significance did not survive FDR correction.

The corrected Table 2/Supplementary Table 1 appears below.

Three corrections have been made to the main text due to the error in the VTA mask.

1. A correction has been made to **the abstract**, *Result*, line 46.

This sentence previously stated:

“From the ROI analyses, the right and left hemispheres of the amygdala and insula, and the right hemisphere of the ventral tegmental area and substantia nigra showed significantly higher responses for the dynamic food ad medium after controlling for covariates and a false discovery rate correction."

The corrected sentence appears below:

"From the ROI analyses, the right and left hemispheres of the amygdala and insula, and the right hemisphere of the substantia nigra showed significantly higher responses for the dynamic food ad medium after controlling for covariates and a false discovery rate correction."

2. A correction has been made to **the method**, *Region of Interest Analyses*, paragraph 1, line 312.

This sentence previously stated:

"Masks of these bilateral regions were generated using the Talairach Daemon and Montreal Neurological Institute (MNI) atlas using AFNI (Analysis of Functional NeuroImages version: 21.0.06, (Cox and Hyde, 1997) and are shown in Figure 3."

The corrected sentence appears below:

"Masks of these bilateral regions were generated using the Talairach Daemon and Montreal Neurological Institute (MNI) atlas using AFNI [Analysis of Functional NeuroImages version: 21.0.06 (Cox and Hyde, 1997)]. The mask of the ventral tegmental area was defined by the sphere with a radius of 5 mm centered at MNI coordinate [4, −16, −10] (Carter, 2009). The ROI masks are shown in Figure 3."

3. A correction has been made to **the results**, *ROI Analyses*, paragraph 1, line 357.

This sentence previously stated:

"Specifically, in both unadjusted and adjusted models and after the FDR correction, the right and left amygdala, the right and left insula, right ventral tegmental area, and right substantia nigra showed statistically significant higher reward-related response to dynamic ads as compared to static ads."

The corrected sentence appears below:

"Specifically, in both unadjusted and adjusted models and after the FDR correction, the right and left amygdala, the right and left insula, and right substantia nigra showed statistically significant higher reward-related response to dynamic ads as compared to static ads. The right ventral tegmental area and left substantia nigra showed significantly higher reward-related response to dynamic ads as compared to static ads before the FDR correction but not after."

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

References

Carter, R. M. (2009). Activation in the VTA and nucleus accumbens increases in anticipation of both gains and losses. *Front. Behav. Neurosci.* 3, 21. doi: 10.3389/neuro.08.021.2009

Cox, R. W., and Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR Biomed.* 10, 17196178. doi: 10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453<3.0.CO;2-L

The corrected Figure 3 with caption:

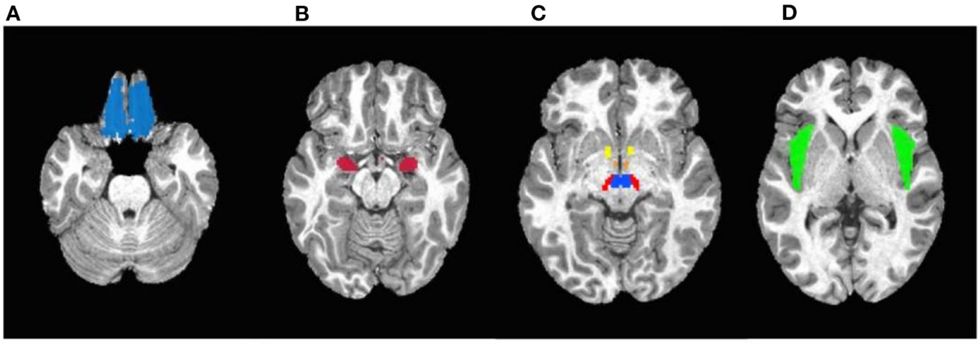


FIGURE 3
Masks used in the region-of-interest (ROI) analysis. **(A)** Orbitofrontal cortex. **(B)** Amygdala. **(C)** Yellow: Nucleus accumbens; Orange: Hypothalamus; Red: Substantia nigra; Blue: Ventral tegmental area. **(D)** Insula.

The corrected Table 2:

TABLE 2 Region-of-interest (ROI) analysis ($N = 115$).

		Unadjusted LME models[1,2,4]			Adjusted LME models[1,2,3,4]		
	L/R	*t*-value	*p*-value	FDR *q*-value	*t*-value	*p*-value	FDR *q*-value
Nucleus accumbens	R	−1.49	0.140	0.218	−1.49	0.138	0.215
	L	−1.20	0.232	0.325	−1.24	0.218	0.305
Orbitofrontal cortex	R	−0.94	0.351	0.406	−0.97	0.331	0.386
	L	−1.02	0.309	0.393	−1.07	0.287	0.365
Amygdala	R	5.34	**<0.001**	**<0.001**	5.34	**<0.001**	**<0.001**
	L	2.43	**0.016**	0.056	2.43	**0.016**	**0.048**
Insula	R	3.07	**0.003**	**0.019**	3.15	**0.002**	**0.014**
	L	2.31	**0.023**	0.064	2.42	**0.017**	**0.048**
Hypothalamus	R	−0.89	0.377	0.406	−0.89	0.373	0.402
	L	0.09	0.929	0.929	0.10	0.919	0.919
Ventral tegmental area	R	2.07	**0.039**	0.089	2.09	**0.037**	0.086
	L	1.95	0.052	0.091	1.94	0.054	0.094
Substantia nigra	R	2.94	**0.004**	**0.019**	2.94	**0.004**	**0.019**
	L	2.04	**0.044**	0.089	2.04	**0.043**	0.086

[1] Linear mixed effects models.
[2] FDR-corrected threshold at $q < 0.05$ was used.
[3] Covariates include BMI-z, age, sex, % caloric intake at preload, and physical activity.
[4] Bold values represent the statistical significance.

The corrected Supplementary Table 1:

SUPPLEMENTARY TABLE 1 Sensitivity analysis with total screen exposure time as a covariate.

		Adjusted LME models[1,2,3]		
	L/R	*t*-value	*p*-value	FDR *q*-value
Nucleus accumbens	R	−1.37	0.172	0.268
	L	−1.23	0.219	0.307
Orbitofrontal cortex	R	−0.87	0.384	0.419
	L	−1.04	0.300	0.382
Amygdala	R	5.26	**<0.001**	**<0.001**
	L	2.43	**0.016**	**0.045**
Insula	R	3.17	**0.002**	**0.014**
	L	2.43	**0.016**	**0.045**
Hypothalamus	R	−0.86	0.389	0.419
	L	0.13	0.895	0.895
Ventral tegmental area	R	2.09	**0.037**	0.086
	L	1.93	0.055	0.096
Substantia nigra	R	2.95	**0.004**	**0.019**
	L	2.00	**0.046**	0.092

frontiers | Frontiers in Neuroscience

TYPE Review
PUBLISHED 27 April 2023
DOI 10.3389/fnins.2023.1155750

OPEN ACCESS

EDITED BY
Miguel Ángel Martín-Pascual,
Spanish Public Television,
Spain

REVIEWED BY
Christin Scholz,
University of Amsterdam,
Netherlands
Pia Tikka,
Tallinn University,
Estonia

*CORRESPONDENCE
Ralf Schmälzle
✉ schmaelz@msu.edu

SPECIALTY SECTION
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

RECEIVED 31 January 2023
ACCEPTED 28 March 2023
PUBLISHED 27 April 2023

CITATION
Schmälzle R and Huskey R (2023) Integrating media content analysis, reception analysis, and media effects studies.
Front. Neurosci. 17:1155750.
doi: 10.3389/fnins.2023.1155750

Integrating media content analysis, reception analysis, and media effects studies

Ralf Schmälzle[1]* and Richard Huskey[2,3,4]

[1]Department of Communication, Michigan State University, East Lansing, MI, United States, [2]Department of Communication, University of California, Davis, Davis, CA, United States, [3]Cognitive Science Program, University of California, Davis, Davis, CA, United States, [4]Center for Mind and Brain, University of California, Davis, Davis, CA, United States

Every day, the world of media is at our fingertips, whether it is watching movies, listening to the radio, or browsing online media. On average, people spend over 8h per day consuming messages from the mass media, amounting to a total lifetime dose of more than 20years in which conceptual content stimulates our brains. Effects from this flood of information range from short-term attention bursts (e.g., by breaking news features or viral 'memes') to life-long memories (e.g., of one's favorite childhood movie), and from micro-level impacts on an individual's memory, attitudes, and behaviors to macro-level effects on nations or generations. The modern study of media's influence on society dates back to the 1940s. This body of mass communication scholarship has largely asked, *"what is media's effect on the individual?"* Around the time of the cognitive revolution, media psychologists began to ask, *"what cognitive processes are involved in media processing?"* More recently, neuroimaging researchers started using real-life media as stimuli to examine perception and cognition under more natural conditions. Such research asks: *"what can media tell us about brain function?"* With some exceptions, these bodies of scholarship often talk past each other. An integration offers new insights into the neurocognitive mechanisms through which media affect single individuals and entire audiences. However, this endeavor faces the same challenges as all interdisciplinary approaches: Researchers with different backgrounds have different levels of expertise, goals, and foci. For instance, neuroimaging researchers label media stimuli as "naturalistic" although they are in many ways rather artificial. Similarly, media experts are typically unfamiliar with the brain. Neither media creators nor neuroscientifically oriented researchers approach media effects from a social scientific perspective, which is the domain of yet another species. In this article, we provide an overview of approaches and traditions to studying media, and we review the emerging literature that aims to connect these streams. We introduce an organizing scheme that connects the causal paths from media content→brain responses→media effects and discuss network control theory as a promising framework to integrate media content, reception, and effects analyses.

KEYWORDS

neuroscience, movies, media, naturalistic, media effects, content analysis, reception analysis

1. Introduction

Media messages permeate our lives; they stimulate rich neurocognitive responses and serve important, much-debated functions within modern information societies. On average, we spend about 8 h per day consuming media (Twenge et al., 2019). Effects of exposure to media range from micro-level impacts on an individual's memory, attitudes, and behaviors to macro-level effects on nations or generations (Bryant and Oliver, 2008; Larzabal et al., 2017). In short, we live in a world where media content flows through our brains much like blood through our veins.

In recent years, researchers have begun to use theories and methods from neuroscience to examine the neural mechanisms of media effects (Weber, 2013; Schmälzle and Grall, 2020a,b; Schmälzle, 2022). This approach is motivated by the fact that the brain is the biological organ underlying all media effects, regardless of whether the study is about movies, narratives (books and audiobooks), or other media types. After all, if a message did not arrive in a recipient's brain, it could not have any effect. This notion of the brain as the central processor of media content is undisputed. It is what motivates the use of neuroimaging to study brain responses to media in the hope of revealing the actual mechanisms that underlie media's effects on perception, attention, comprehension, affect - or whatever the focal topic of a concrete neuroscientific investigation that uses media may be.

However, while the promise of neuroimaging in this area is generally recognized, the complexity of the enterprise cannot be underestimated. Media are a highly complex kind of 'stimulus', actually, they are a sequence of a multitude of individual stimuli. Moreover, media evoke multiplex brain responses. And finally, media result in a mosaic of consequences - from short-term to long-term effects and from individual to collective outcomes.

Given this complexity, it is no surprise that multiple disciplines exist at the nexus of media and the brain. Researchers in the fields of communication and media studies have largely focused on issues related to media content and the effects of exposure to such content (Figure 1, left; Riff et al., 2014; Neuendorf, 2017). By comparison, psychology and media psychology investigate the cognitive processes that subserve media processing and effects (Figure 1, middle; Weber et al., 2008; Lang and Ewoldsen, 2010). By comparison, the cognitive sciences and cognitive neurosciences primarily use media as a tool for studying cognition and the brain (Figure 1, right; Spiers and Maguire, 2007; Hasson and Honey, 2012; Sonkusare et al., 2019; Vanderwal et al., 2019).

Of course, these broad generalizations mask substantial disciplinary and topical heterogeneity such that inquiry surrounding media and the brain is a bit reminiscent of people feeling an elephant in a dark room (Figure 2): In this parable, each person brings their own experience and perspective to the endeavor of identifying the elephant, but each person is only able to feel just one small part of the large animal. In the same way, many different perspectives about

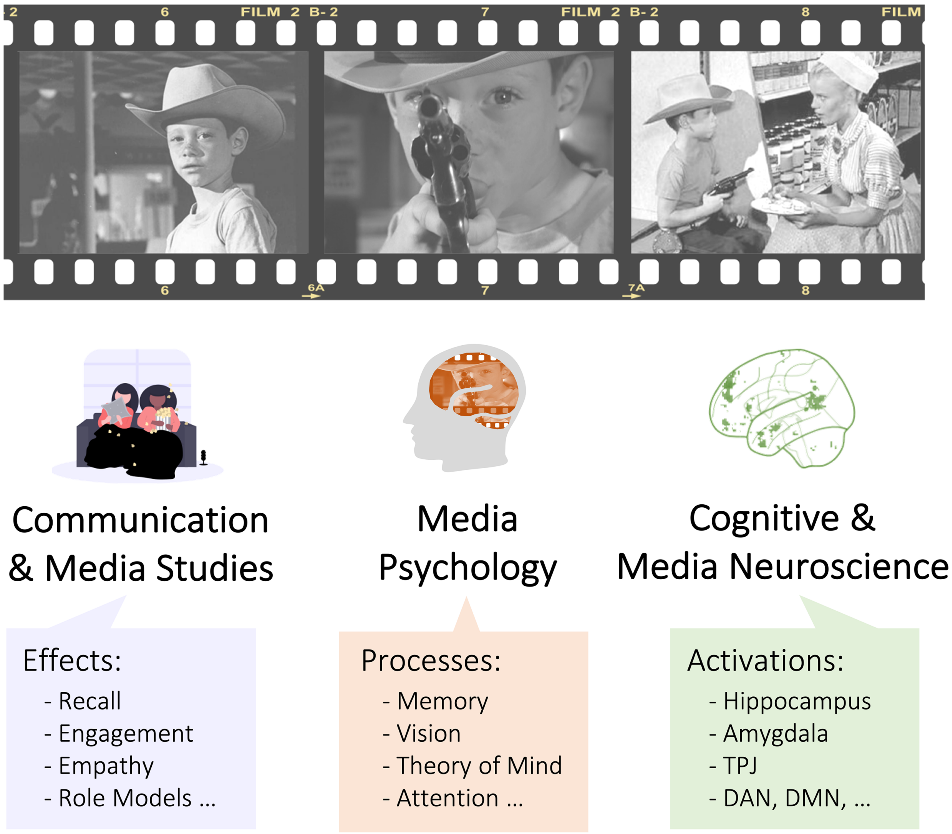


FIGURE 1
Connecting siloed fields of media effects, media processes, and media neuroscience. Traditionally, these topics have been studied by different academic disciplines.

FIGURE 2
An elephant in a dark room. In this classic parable, people investigate an elephant in a dark room. Each can only feel part of the elephant and cannot identify the whole. Misunderstanding ensues.

media and the brain coexist - all valuable in and of themselves - but there is a lack of integration and a lot of confusion. In fact, early career researchers who consider working at the intersection of media and the brain will find themselves in a complex theoretical and methodological landscape that spans disciplines and even paradigms from the humanities, traditional STEM disciplines, and the social sciences. This state of affairs can make it difficult to see the proverbial elephant in the room, and one can almost ask oneself: If "naturalistic neuroimaging" or "movie fMRI" is the answer, what is the question (see Kosslyn, 1999)?

With this in mind, this article suggests a conceptual framework to integrate these disparate research streams of media effects, media processing, and media neuroscience. We begin by introducing and discussing each area and provide the logical division into *content analysis*, *reception analysis*, and *effects analysis* as an organizing scheme. Then, we suggest *network control theory (NCT)* as a framework with the potential for integrating these siloed traditions. We believe this framework can shed light on the elephant in the dark room and reveal causal mechanisms by which the content of media messages affects brain responses and how the resulting message effects in single individuals aggregate into media effects in large populations.

2. The arrow of causality: from media content to reception responses to media effects

So far, we have discussed how different areas of disciplinary inquiry are largely organized around levels of analysis (media effects on individuals and society, media processing within individuals, neural responses within individuals). As this section will show, a framework organized around levels of analysis does not cleanly map onto a causal path that begins with exposure to media content and ends with media effects. In this section, we give an overview of our conceptual model that starts with media as a stimulus (a brief text message, an audiovisual movie, a social-media video clip, an audiobook) containing conceptual content that is analyzed by the brain and results in what has traditionally been called media or message effects (Figure 3).

2.1. Assaying the ingredients: media content analysis

Media are inherently content-rich and, therefore, complex. To demonstrate by selecting one possible example, consider movies. Among the most popular types of media, movies comprise multimodal content (images and soundtracks) that include a wealth of semantic and social-pragmatic dimensions that vary over time. The term movie emerged as a shortcut for moving images - essentially by stitching together photographs in rapid succession. For example, a typical Netflix HD movie streams about 3-7GB of data, containing over 100,000 individual frames, each containing many pixels. It becomes clear that if we consider the pixel-level information of any given movie, the information contained in a movie quickly reaches billions. These flickering pixels form the manifest content of the movie as it emerges from your TV screen.

Clearly, though, looking at movies as a multitude of pixels misses the point - just as it makes little sense to use a microscope to examine ink-saturated paper when reading a fiction book. Typically, when discussing movies, we mean their higher-order information, such as narrative and social-cognitive content. Clearly, we also do not remember the surface-level information (the pixels), but we recall and retell what happens to characters and the overall trajectory of a plot (like heroes and villains, or a rags-to-riches story, etc.; see Kintsch, 1998 for a similar argument about language comprehension).

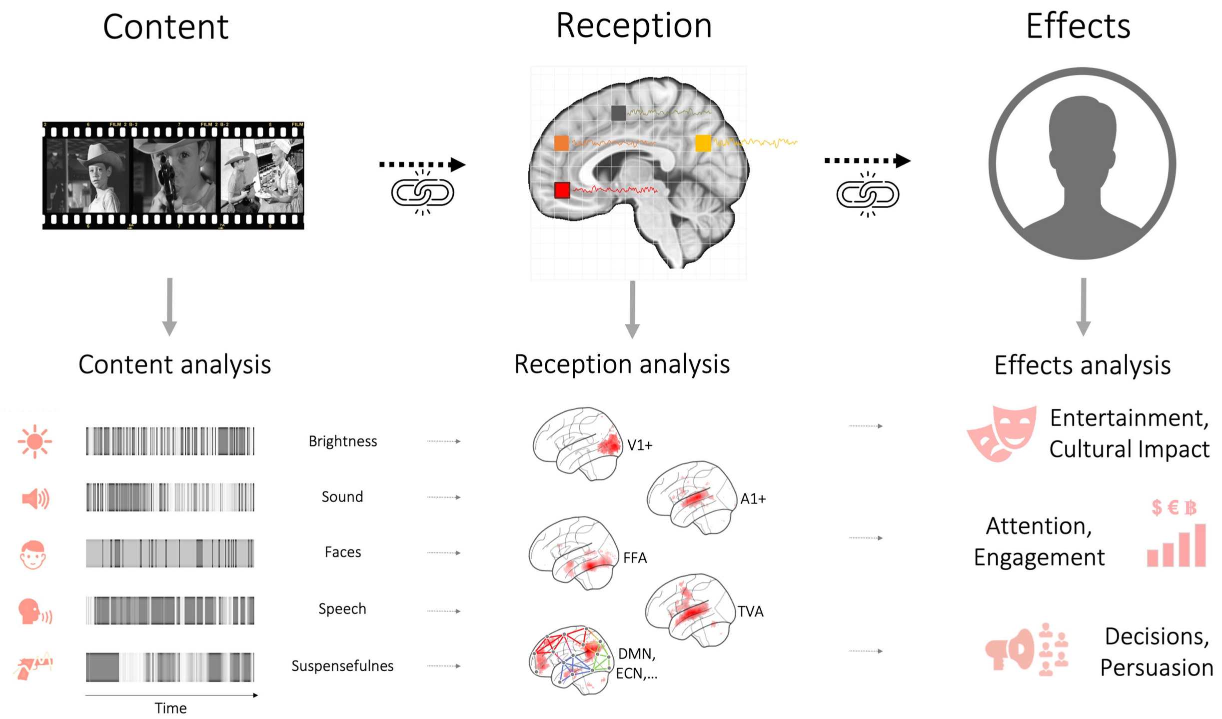


FIGURE 3
The arrow of causality from media content to reception responses to media effects. The bottom left panel illustrates how content analysis quantifies content (e.g., for use as a design matrix), allowing to map out brain systems responding to specific content elements. These can then be linked to effects of media consumption. Of note, reception responses to an incoming (causal) media stimulus can also be modulated by receiver state, background knowledge, beliefs, and so forth.

Between the pixels as the lowest-level content features and the macro- or plot-level content features lie numerous intermediary-level features. For example, consider now the soundtrack of a movie (instead of the video track containing the pixels). At a lower level, a movie's soundtrack is characterized by physical properties like its constituent amplitude, frequency content, etc. However, this all is embedded in a nested, hierarchical structure: Stretches of sound encode particular phonemes, which in turn represent words, words are nested in sentences, and a couple of sentences by one speaker are typically followed by a response from another speaker, reflecting a dialogue in a scene. The same case can be made for visual content (e.g., Hasson et al., 2008a). Thus, it becomes clear that *the content* of a movie - a deceptively simple singular word - actually encompasses multiple content elements that can be organized along a hierarchy of abstraction (see Figures 3, 4). Which specific content element is of interest to researchers often depends on their home discipline - just like in the elephant in the dark room parable. Arguably, since movies are largely created and consumed to entertain, the most relevant level is the plot level. Still, it is clear that all lower levels (sounds, words, sentences, paragraphs or pixels, images, scenes) are necessary to convey the plot-level content of a movie (or a book or whatever the media format).[1]

One way of quantifying all this higher-order and often latent content (or subtext) is through a procedure known as content analysis (Krippendorff, 2004; Riff et al., 2014; Neuendorf, 2017). Historically, content-analyzing movies and media more broadly was an arduous task. For example, the famous National Television Violence Study (Federman, 1995) relied on manual labor from human coders to annotate over 10,000 h of content over a three-year period. Most content analyses are smaller in scale, but the effort required is still a key bottleneck (Greenberg and Atkin, 1980; Masters et al., 1991; Hahn et al., 2017). Consequently, classical content analyses usually feature sparse sampling frames (e.g., the first 10 min of content from a sample of movies) that often incompletely describe the entire media corpus. Another, not unrelated problem, is that there is often substantial error in human annotations, which can be quite difficult or even impossible to resolve in some circumstances (Weber et al., 2018).

Advances in computational analysis make this task much faster, scalable, feasible, and accurate. The catch, however, is that computational analyses are currently only able to quantify manifest lower- and mid-level features of the content-abstraction hierarchy. We assume that readers will be familiar with the explosion of research on natural language processing and computer vision. As of 2023, computers can automatically quantify many sound characteristics

---

[1] Of note, while we believe that upper levels (i.e., beyond sensory-perceptual content) are important and largely understudied, at least when it comes to neuroimaging, this does not mean that it was not worthwhile to study lower levels. For instance, researchers in visual neuroscience and neurolinguistics have both discovered the benefit of using movies as stimuli, and many other lower- and mid-level neurocognitive processes (e.g., event segmentation, situation model building, etc.) can be examined fruitfully using media.

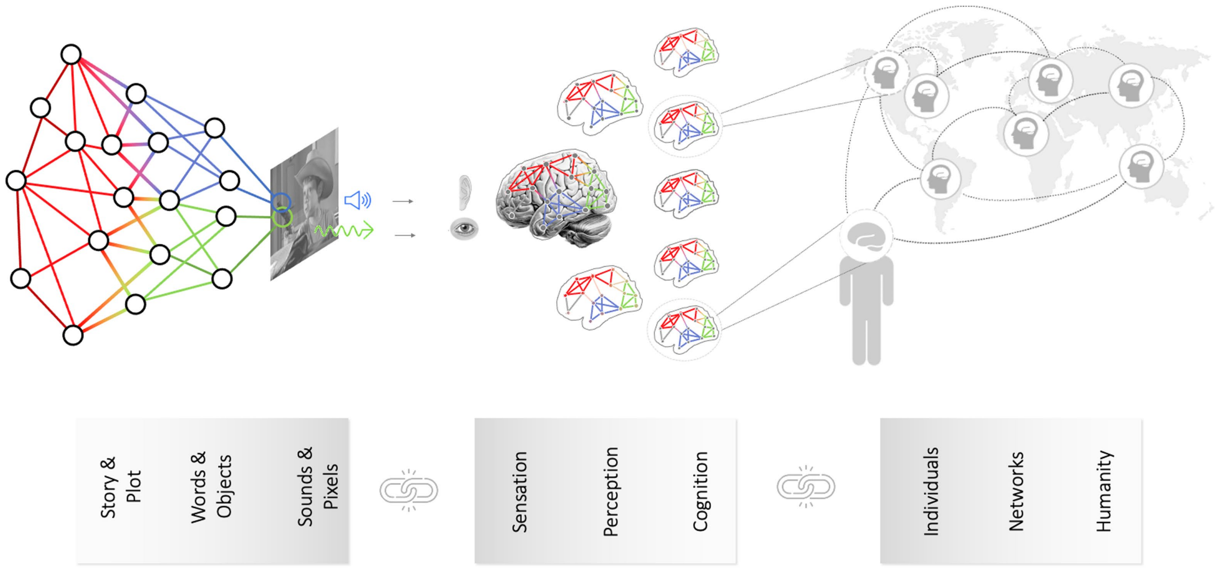


FIGURE 4
Linking hierarchies of content with matching reception mechanisms and integration with media effects. Content is a deceptively singular word, but it encompasses multiple layers - from manifest (e.g., pixels) to latent content (e.g., subtext, story morals). Understanding content as a network of information layers (left panel) allows for its comprehensive quantification, facilitating the identification of corresponding reception mechanisms. In a similar fashion, we can also integrate individual-specific neurocognitive processes during media reception to media effects and social levels.

(such as energy and pitch) and even transcribe spoken content into writing (Radford et al., 2022), and in the visual domain, they excel at quantifying image properties, recognizing objects, or even detecting actions in image sequences (Rohrbach et al., 2017).

These advances can be leveraged to analyze media content in a rigorously quantitative fashion and at scale. For instance, researchers have used face-detection systems to detect characters in movies or natural language processing (NLP) techniques to identify characters from scripts, both of which can be used to create character networks (e.g., communities of characters that co-occur in the same scenes; Hopp et al., 2020; Baldwin and Schmälzle, 2022; Malik et al., 2022) or to create time-locked explanatory variables for neuroimaging analyses. Or researchers have used NLP techniques to study moral language in media (Weber et al., 2018). Perhaps the most systematic yet still young approach in this area is the NeuroScout platform and the related pliers python package (McNamara et al., 2017; De la Vega et al., 2022). NeuroScout provides easy access to machine-learning methods capable of automatically extracting hundreds of features that range from the very concrete (like root mean squared amplitude of the sound signal) to more abstract ones (like concept-level image tags from Google's or Clarif.ai's computer vision systems).

Overall, computational tools for extracting content features are immensely promising for studying content in a rigorously quantitative and automatic manner. However, we must acknowledge that even the most advanced machine-learning systems fail to achieve human-level understanding (Marcus and Davis, 2019; McClelland et al., 2020). In fact, even though impressive progress is made in modeling so-called common sense knowledge, current systems still fall short in many regards when it comes to coding abstract categories of content, such as sarcasm or humor, or detecting sequential narrative information related to story schemata or scripts, or visual action depictions (Vicol et al., 2018; Choi et al., 2021; Zellers et al., 2021). Taken together, automatic approaches can excel at quantifying lower- and mid-level properties of content, but they still face a barrier (Karpathy, 2012; Mitchell, 2020) when it comes to analyzing higher-order media content.

Said differently, the content of a movie (or other media) can be analyzed very concretely and efficiently in terms of physical properties, such as brightness and contrast, and also for intermediate levels, like the presence of objects, such as guns and faces. At a yet more abstract level, however, the movie has an event structure (separated by cuts) and a plot that conveys the overall narrative. This type of abstract content is currently much harder to quantify, even with advanced machine learning and NLP techniques. Indeed, we often find ourselves resorting to psychological terms to describe content-level properties whose "ingredients" in content remain somewhat unclear, such as the 'suspensefulnes' of a movie to describe its potential to elicit suspense (see Cummins, 2000). However, it is clear that these content elements matter for a movie's impact on viewers' brain responses.[2]

---

[2] In fact, they may matter even more than lower-level content features because we can easily turn any story into a movie and vice versa, which completely exchanges the lower-level content types, but keeps the higher-level information intact (Honey et al., 2012; Regev et al., 2018). Likewise, one can also transform a given movie in many ways, like into a comic, or a reissue with newer actors, thus changing all lower-level features, yet it will still stay the same movie. However, we want to avoid creating the impression that we give priority to higher-level content elements or that we consider lower-level elements as less important. This is not the case. For example, content creators (directors and camera operators) often make strategic use of lower-level content elements (e.g., angle, shot sequence) to create specific

The upshot of all this is that if our goal is to understand the effects of content of the brain, then a diverse and rapidly improving toolkit for quantifying media content already exists for more concrete features, and we can rely on traditional human content annotations to quantify higher-level aspects of content that are still beyond the capabilities of computational tools. In the next section, we discuss how this quantified content is the key to deciphering the brain responses.[3]

2.2. Reception analysis: how brains respond to media

It is clear that media content's arrival in the brain sets forth a cascade of reactions (Kepplinger, 1989; Bryant and Zillmann, 1990; Potter and Bolls, 2012; Schmälzle and Grall, 2020a). Just like we started our analysis of movie content at the pixel level, we can begin our quest into the brain at what can be considered the neural counterpart of the pixel: an individual cell (rod/cone) in the receiver's retina that gets stimulated by light and converts the televised movie's signal into a neural impulse. Due to space limits, we cannot trace this signal's neural itinerary in fine detail, but a rough sketch goes like this: From the retina, information travels along the optic nerve into the thalamus, gets relayed in the lateral geniculate nucleus, and arrives *via* the optic radiation in the primary visual cortex (Mesulam, 1998; Chalupa and Werner, 2003; Fuster, 2003; Poeppel et al., 2020), and so forth. The seminal work by Hubel & Wiesel on receptive fields provides perhaps the most concrete examination of content-extractors (or feature detectors) in the brain; that is, neurocognitive mechanisms that match certain content elements, like oriented lines, edges, or motion (Hubel and Wiesel, 1962).

However, just like with the analogy of trying to read a book with a microscope, studying movies as purely visual stimuli that activate the retina and V1-edge-detectors runs the risk of missing the point: We clearly do not watch movies simply to obtain visual stimulation, and we do not read or listen to books solely because we like letters and sounds, or processing any of the intermediary representations like objects, action sequences, or speech. Instead, we typically use media to engage their higher-level, albeit more difficult to quantify, content.

Few researchers would question the statement that "content is key" for understanding how media impact the brain. However, looking into the emerging literature on media and neuroscience, it is apparent that content is often simply ignored. In some ways, this is understandable. Modern neuroscience already requires extensive training in neuroanatomy, physiology, physics, statistics, engineering and signal processing, psychology, philosophy, programming, high-performance computing, and so on, such that there is little time left to also train in scholarship on complicated and sometimes even poorly specified content features that come with media stimuli (e.g., narratives, characters). Similarly, when using media as stimuli, it is not always so clear exactly what needs to be accounted for in either experimental design or statistical analysis. Should we account for luminance? Sound amplitude? The presence of faces? If so, how? The difficult answers and unappealing tradeoffs associated with these questions have spurred clever solutions optimized for designing around all of this complexity. Such approaches include calculating intersubject correlations (ISC; Hasson et al., 2004, 2008b), or borrowing other methods from resting-state fMRI, dynamic causal analyses (e.g., Granger causality or DCM methods), or introducing other advanced tools to decipher entangled brain responses (Di and Biswal, 2020; Van Der Meer et al., 2020; Busch et al., 2022).

It is not our goal to criticize this research as it has already led to important new discoveries about the brain. Nevertheless, these approaches are largely content-blind. We argue that without an equal appreciation of the content, this endeavor will yield only limited insights (see Okdie et al., 2014, for a parallel argument about media psychology). After all, it is clearly the content where the causal arrow originates that evokes the brain responses. Thus one should devote equal sophistication to content analysis as to reception analysis (i.e., analysis of neural or other types of data).

Not all neuroimaging analyses are content-blind, though. In fact, some go to great lengths to quantify or manipulate content. However, we claim that even these approaches are still limited when it comes to identifying the kinds of higher-level content elements that prompt conceptual and affective reactions to media and drive media selection and consumption behavior. For example, in studies of natural vision, movies are increasingly adopted as stimuli because they depict relatively natural scenes (except for things like cuts and blends; Hasson et al., 2008c; Çukur et al., 2013). Such studies also tend to do a great job quantifying aspects of content that are relevant to their area of study, like meticulously annotating visual content properties such as contrast, individual objects, and so forth, or manipulating content *via* scrambling (Hasson et al., 2008c; Çukur et al., 2013; Huth et al., 2016b; Wen et al., 2017). Studies like these make great use of movies as an experimental stimulus, and they can serve as role models for how content analysis can inform reception analysis. These studies represent the kind of work that examines carefully one specific part of the proverbial elephant (e.g., visual processing). As such, they are extremely valuable for understanding vision. However, although vision clearly is central to movie viewing and the entertainment experiences it produces, vision alone is only one piece of a larger mosaic of movie-evoked brain responses. Moreover, to the extent that higher-level content properties (such as suspense fluctuations in a movie) impact attention, it is probably the case that the measurements might be biased (e.g., Van Berkum et al., 2009; Gantman and Van Bavel, 2014; Schmälzle and Grall, 2020a,b).

Much like the visual neurosciences have begun to adopt media as a more naturalistic alternative to traditional stimuli, neurolinguistics has also begun to embrace media (like stories, audiobooks, and movies with dialogue). In the early days of neuroimaging, language studies were notoriously artificial single-word studies (e.g., using sparse sampling event-related designs). The trend towards more naturalistic neuroimaging prompted an upsurge of studies using natural, running speech as stimuli - often taken from audiobooks and similar story-based media formats. Like their counterparts in the visual domain, neurolinguistics studies do a great job at annotating word-level linguistic properties, such as word length, frequency, syntactic role, or even basic semantic aspects (e.g., GloVe or Word2Vec embeddings) and relating these to the stimulus-evoked brain activity in a forward-inference manner (Lerner et al., 2011; Huth et al., 2016a; Broderick et al., 2018). As this trend advanced, the stimulus characteristics that

---

impressions. Studying these techniques, their impact on brain activity, and their effects on viewers is as valuable as analyzing, e.g., the plot narrative.

[3] *Via* standard forward inference (Henson, 2006).

were coded became more nuanced; for instance, it has been demonstrated that decoding results become better if one uses sentence-based embeddings as opposed to word-level-only embeddings. However, the key point is that these neurolinguistic studies also struggle to consider content elements that go beyond the linguistic level (McClelland et al., 2020; Arana et al., 2023). However, just like reading a book with a microscope, we claim that we do not consume stories because they provide linguistic stimulation. Rather, it is the supralinguistic content and the responses this evokes that are critical: stories entertain, satisfy social needs, pique our curiosity, and so forth.[4]

A still small but growing number of studies attempt to link higher-level media content, which influences post-perceptual processes like attention, semantic comprehension, and particularly affective and social-cognitive responses, to brain responses (Hasson et al., 2008b; Yeshurun et al., 2017; Richardson et al., 2018; Tikka et al., 2018; Nguyen et al., 2019; Schmälzle and Grall, 2020a,b; Baek and Parkinson, 2022; Grady et al., 2022). For example, it is well known that movies are highly social in content and that their social and affective content is key to why we engage with them in the first place. In fact, movies are bursting with depictions of social interactions, including love, aggression, betrayal, etc. - and viewers take an intense interest in the fate of characters (Bryant and Zillmann, 1990; Oatley, 2002; Tannenbaum, 2014). Because of this, movies and other fiction-based media are almost ideal tools for studying social cognitive processes like empathy, perspective-taking, trait inferences, and so on (Vorderer, 1996; Klimmt et al., 2006). These characteristics of movies are increasingly recognized by neuroimagers interested in the neural basis of such processes (Salmi et al., 2013; Byrge et al., 2015; Richardson et al., 2018; Nguyen et al., 2019; Broom et al., 2021; Chang et al., 2021), even beyond human neuroimaging (Mantini et al., 2012; Sliwa and Freiwald, 2017).

Similarly, these social-cognitive responses to movies are intimately interwoven with affective reactions. For instance, viewer affect reliably tracks character victories and failures, good fortune and suffering, trials and tribulations such that audiences experience strong participatory responses (e.g., goosebumps during the hero's victory at the end, crying during 'all is lost' moments when it seems that the hero is doomed to failure). In fact, it has been said that Hollywood is - at its core - a giant experimental psychology lab specializing in creating emotional stimuli that can effectively affect mass audiences. Likewise, Alfred Hitchcock, the famous master of suspense, described his profession as "based on an exact science of audience reactions" (Hasson et al., 2008a). Because of this capacity, entire genres of movies are devoted to catering to certain segments of the affect spectrum, and a few neuroimaging studies have explored such phenomena. For instance, suspense movies take audiences on an emotional rollercoaster that blends future-oriented cognitions like hope and anxiety (Bezdek et al., 2017; Schmälzle and Grall, 2020b). Action movies can stimulate intense bursts of arousal (Hermans et al., 2011; Kautonen et al., 2018). Comedy tickles our funny bone (Sawahata et al., 2013; Amir et al., 2015; Jääskeläinen et al., 2016; Schmälzle et al., 2022), drama/tragedy deals with human responses to suffering (Raz et al., 2014, 2016). And, while often hushed up, pornography is certainly quite powerful in stimulating experiences (Prause et al., 2015; Schmälzle et al., 2017; Chen et al., 2020; Grubbs and Kraus, 2021).

In sum, it is clear that media feature a host of content that can elicit and precisely steer social-cognitive and affective processes. In fact, due to this capacity, media are very promising to study the neural basis of these phenomena in a way that is more appropriate to their nature than, say, event-related studies of single words, affective images, and so forth (Hasson and Honey, 2012; Saarimäki, 2021).

The challenge, then, is to quantify the social and affective content characteristics to be able to unlock its mechanism of action *via* neuroimaging. The studies presented above are in an advantageous position because the content properties that we care about are relatively well understood and can be coded straightforwardly (as done in the NeuroScout system or *via* the Matlab vision toolbox or some natural language processing toolbox). By contrast, when the research focus is on social-cognitive and affective phenomena, the task of coding the conceptual content is considerably more difficult,[5] although some clever ways exist to attempt to parametrize these more challenging factors (Heider and Simmel, 1944; Meyer et al., 2019; Nguyen et al., 2019). But it is clear that if we ignore higher-level content altogether, then we cannot expect to meaningfully relate brain responses to their elicitors - at least not beyond relatively simple sensory-perceptual brain responses, and if top-down attention comes into play, even these will get affected. This is the problem with 'content-blind' neuroimaging.

2.3. Media effects: how media influence individuals and large-scale populations

The last link in the causal chain from content to reception is the question of how exposure to media changes memories, attitudes, or behaviors. The term media effects refers to these psychological or behavioral outcomes of stimulation with media. Of note, the term media effects is used to refer to individual-level as well as population effects (Bryant and Oliver, 2008). The latter clearly depend on the

---

[4] Again, we are not shy to admit that the content properties that cater to these processes are difficult to quantify: For example, computing a sound envelope/RMSE feature is easy. Nowadays, computing BERT-embeddings for every word of a story is also quite doable. However, even though these properties are relevant to understanding a story, they alone are insufficient. Parallel arguments about this exact issue are also made in the NLP community, where debate rages about the capabilities and limitations of large-language models (Bender and Koller, 2020). Yet, again, we want to emphasize that our goal is not to declare only the plot level as the only level worth quantifying. Rather, examinations of specific linguistic and sound features, their creation, their effects on the brain, and their impact on audiences are inherently relevant and worth studying.

[5] Researchers often rely on their intuition. In fact, most movies used in fMRI studies seem to be chosen for their social-affective elicitation potential. Movies that have been used include Bang Bang! And you are dead; The Present; Partly Cloudy; Curb your Enthusiasm; The Office; Sherlock; Memento. These are all great, and it seems clear that "researchers felt something" when they opted to use these movies. However, none of the papers devoted more than one or two sentences to the content and theoretical reasons why it was chosen.

former, but in practice, they tend to be studied by different research communities who focus either on micro- (intraindividual) or macro (social) levels of analysis.

The origin of the field can be traced back to social scientific research in the 1920s and 30s, which is the era when the first distant mass media (radio, TV) emerged. Historically, the field has swung back and forth between periods in which researchers postulated relatively strong media effects and those of weaker effects. For example, in the period between 1920 and 1950, much research attention centered on the putatively strong influence of propaganda (Hovland and Lumsdaine, 2017). Modern efforts showcase that media effects tend to be smaller in nature and more contextually dependent (Lang, 2013; Rains et al., 2018). Nevertheless, and despite substantial evidence to the contrary, today's pressing topics like radicalization, fake news, deep fakes, and the influence of social media are often cast in overly simplistic terms and assume overly powerful effects. Neuroimagers looking to use media as stimuli should recognize that, contrary to common perceptions, media effects tend to be quite small in practice.

The list of media effects and media effects theories is too long to discuss here. Still, a partial list of interesting phenomena and theories includes, e.g., the third-person effect - the belief that media influence others more than oneself (Perloff, 2002). Readers are likely familiar with the famous Bobo Doll Study that helped give rise to Social-Cognitive or Social Learning Theory (Bandura, 1977). Central to this theory is the notion of observational learning and role models - both of which can occur during media consumption - and therefore Social Cognitive Theory is widely used to explain social media effects (Bandura, 1994). Similarly, Affective Disposition Theory (Zillmann and Cantor, 1972; Raney, 2004) links characters and plot elements to affective audience responses. There are, of course, many other interesting effects and theories of media influence to highlight, but for the sake of space, we refer readers to key reference works (Zillmann and Vorderer, 2000; Bryant and Oliver, 2008; Littlejohn and Foss, 2009; Nabi and Oliver, 2009; Dill, 2013; DeFleur, 2016).

In essence, any result of media stimulation could be considered as a media effect, whether it is short-term memory (e.g., recalling last night's news), long-term memory (e.g., remembering a childhood TV show), a change in attitude, a belief (e.g., being more open to immigration after watching a refugee drama), or behavior (e.g., donating money to charity after viewing an ad). These effects are often linked to their elicitors in content, but how the brain mediates between content and effects has traditionally been ignored. Instead, because neuroimaging measures were unavailable until recently, researchers had to rely on self-report methods that were usually taken after the media consumption ended (Lang, 2014).

Critically, media effects are not only studied in single individuals but often with an eye toward aggregate audiences. The field most closely associated with this perspective is mass communication. In brief, mass communication describes a *one-to-many* mode of communication in which the same message is sent out to multiple recipients. For instance, early mass media were newspapers where the same article would be read by all readers. Radio marked another milestone, then most notably followed by Television. And, although social media has now upended the traditional "one-to-many" model of mass communication, providing a many-to-many mode of communication instead, it is still true that a single social media message can be sent out to a large audience, and the brains of audience members would then still respond to the same message (Schmälzle and Grall, 2020a,b; Gong et al., 2022).

Given the important effects media can have on the masses and public opinion (Lippmann, 1922; Noelle-Neumann, 1991), it is clearly of interest to examine how reception responses relate to such large-scale media effects. In other words, might media-evoked brain responses allow researchers to predict subsequent effects? Indeed, several emerging neuroimaging studies (and a large body of non-neuroimaging studies from the social sciences more broadly) have begun to examine this question. For instance, Hasson et al. showed that brain imaging data captured during viewing could predict memory, a very concrete and clear-cut media effect (Hasson et al., 2008a). Falk et al. showed that brain responses to health messages could predict message-consistent behavior change at later points (Falk et al., 2010), and several other articles examine effects related to persuasion, broadly defined, or engagement with and sharing of messages in social networks (Weber et al., 2015a; Baek et al., 2017; Huskey et al., 2017; Coronel et al., 2021). These studies point to the potential of using brain imaging data to predict individual-level outcomes, that is, how to link reception responses captured in individuals to the ensuing media effects.

Another intriguing twist for using brain imaging data is to predict collective outcomes. By that, we mean that it is possible to record the brain's responses during reception from a smaller test audience and link them to aggregate outcomes in larger groups (Berkman and Falk, 2013). For example, in the neuroeconomics literature, researchers have predicted the cultural popularity of music from brain responses (Berns and Moore, 2012). Similarly, Dmochowski et al. (2012), used brain responses to SuperBowl commercials to predict online engagement (tweet volume; Dmochowski et al., 2014), and Falk et al. used brain responses to health messages to predict campaign success (call volume to an anti-smoking quitline; Falk et al., 2012).

The broader reasoning behind these efforts, which connect the brain responses of single individuals or small groups to large-scale population-level media effects, is based on the one-to-many mass communication logic: A message is sent out and processed by multiple individuals comprising an audience. If a given test audience is representative of a larger population, their brain responses can serve as a potential predictor of aggregate outcomes. That this works is just as logical as it is logical to use self-reports from samples to forecast larger outcomes (Knutson and Genevsky, 2018). At present, this approach has been used only in a few studies. Still, given the desirability of movies and media as stimuli, we can expect that many others will follow: After all, movies often even galvanize culturally shared, long-lasting collective memories (e.g., the famous shower scene in Hitchcock's Psycho), suggesting that these effects have a shared basis in the brains of people who saw the specific footage (see, e.g., Kauttonen et al., 2018 for a neuroimaging study of key-frames). The same logic can also be applied to study how movie content produces any kind of convergent audience response, from collective suspense and fear during a horror movie to collective laughter during comedy (Schmälzle, 2022; Schmälzle et al., 2022).

Taken together, media effects are clearly consequential, of enormous interest to social scientists, and one of the most attractive areas that neuroscience researchers would like to seize. Especially the widespread ability of digital data (e.g., time-locked comments during movies and shows, social network metrics; Dmochowski et al., 2014; O'Donnell and Falk, 2015; Ni and Coupé, 2023) increases, there are unprecedented opportunities to link neural data to media effects. However, doing so in a meaningful way will - again - require keeping an eye on the content that starts the logical sequence from media content to brain responses to media effects. Said differently, we can only hope to explain media effects if we trace them back to the preceding brain responses and these, in turn to their elicitors in content.

To summarize, the previous section presented content analysis (2.1), reception analysis (2.2), and effects studies (2.3), arguing that these domains stand in a logical relationship with each other. And in each of these sections, we have pointed to the ways researchers have typically engaged in linking media, neural responses, and effects. These projects, while groundbreaking in their own right, often only investigate a subset of the causal chain from media content to reception responses to media effects. In what follows, we introduce Network Control Theory (NCT, Liu et al., 2011) as an integrative analytical framework that is well-suited to help further integrate these domains.

3. Network control theory: examining how media bring brains into specific states

In this article (and the special issue in which it appears), the brain takes center stage as the organ of media reception; that is, the site of action where complex content sets forth the activities that ultimately produce media effects. However, it is clear then that quantifying content is only half the battle - the other half deciphering the brain's reactions to it. This, in turn, requires a general theory of brain function to motivate an analytical framework for studying content-brain relationships. Our model of brain function is based on current cognitive neuroscience research that views the brain as a complex, hierarchical network (Mesulam, 1998; Fuster, 2003).[6] Entry-points into the network and its lower-level nodes (the eye, retina, optic nerve, LGN, and V1+; or the ear, cochlea, auditory nerve, olivary colliculi, and A1+) are relatively localized, and they correspond rather directly to specific lower-level content features (e.g., Hubel & Wiesel-type feature detectors). Subsequent layers of neural processing, however, tend to be more distributed, which calls for more multivariate analysis methods.

Over the past decade, network-based multivariate methods have been applied to neuroimaging data, and several large-scale brain networks have been identified (e.g., Medaglia et al., 2015). However, much of this work has been based on data captured in the so-called resting state, i.e., with participants only lying in the scanner. While this work has led to substantial and important insights, it is clear that the unconstrained nature of the resting state task is a limiting factor. By contrast, movies and media more broadly are ideal candidates to advance this research: They provide a rich and relevant stimulus for participants and one that is controlled insofar as it provides exactly the same input for everyone. Moreover, media can steer neurocognitive responses related to perception, attention, memory, and emotion, and it is this property that makes them ideally suited for studying cognitive neuroscience but also relevant for social science research trying to understand their mechanisms of influence. With this in mind, we will next introduce a mathematical framework - Network Control Theory - that uses external control forces (here: a movie and its content) to steer networked systems (here: the brains of audiences exposed to the movie).

3.1. What is network control theory?

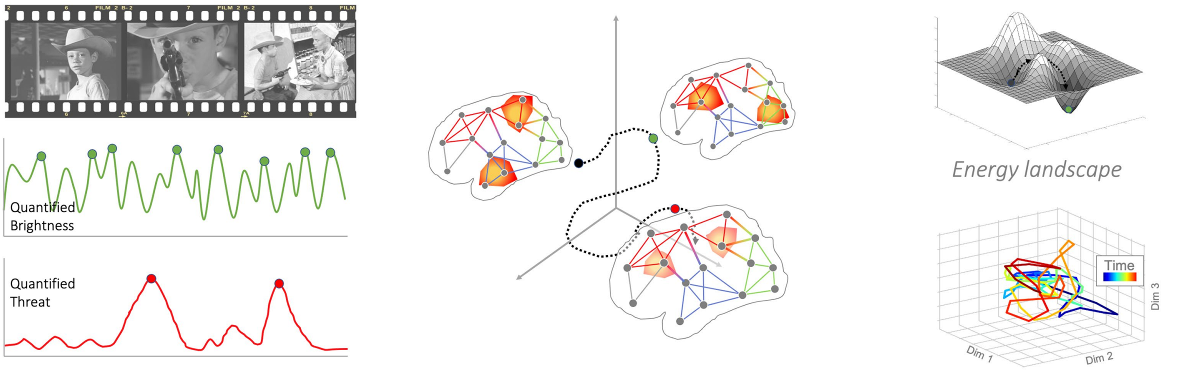
Network control theory is a branch of control theory in engineering and a subfield of the larger network sciences (Gu et al., 2015). It deals specifically with the question of how networked systems can be controlled. What does it mean to *control* a network? Simply put, network control theory is a computational model that specifies if and how interventions, and their corresponding energetic costs, drive complex systems between different topological organizations with different energetic requirements (Muldoon et al., 2016; Tang and Bassett, 2018; Kim and Bassett, 2020; Lydon-Staley et al., 2020). More specifically, a given network topology requires energy costs to maintain.[7] Networks can shift between different topological organizations, each with a different energetic requirement and these topological shifts can have their own energetic requirements, as well (see Figure 5).

To make this idea concrete via example, imagine a system of highways connecting several cities. The topological organization of this series of highways requires energy to construct and requires energy to maintain. Imagine now that the topology is updated; a new highway is built and added to the existing network. Constructing the new highway will also require energy, as will maintaining that new

---

[6] Network science is an application of graph theory where systems of information can be grouped into nodes (specific elements) and edges (the relationships between those elements). The beauty of network science is its domain generality. Network systems can be constructed to represent social organization (e.g., each node is an individual, each edge represents if individuals are friends or not), information on the internet (e.g., nodes represent a webpage, edges represent hyperlinks between websites), civil infrastructure (e.g., nodes represent cities, edges represent highways connecting cities), biological systems (e.g., nodes represent gray matter corresponding to specific brain structures, edges represent white matter fiber tracts connecting gray matter), and more (for a review, see Newman, 2010). The constellation of edges connected by nodes describes a network's organization. This organization is commonly referred to as a network's topology.

[7] There is the energy necessary to maintain a given topological organization. In the case of brains and brain networks (both structural and functional), this is associated with energetic costs, most notably metabolic costs (Bullmore and Sporns, 2012). There is also the energy that is necessary to transition a system into, and maintain, a given topological organization. In neurscientific contexts, this includes things external energy sources such as an experimental task, a pharmacological intervention, a specific stimulus, and so on, that drive the brain from one functional topological organization to another. Network control theory can be used to account for both cases (see, e.g., Gu et al., 2015; Lydon-Staley et al., 2020, respectively). In our application, we are particularly focused on the latter case without denying the former.



Quantifying movie content over time (examples)

Snapshots of network states during movie viewing

Trajectory of multivariate brain activity during movie viewing

FIGURE 5
Network Control Theory. Left panel: Movie events are the causal forces that push the brain (or brains of entire audiences) into different states. For instance, the sudden reveal of a betrayal will likely engage theory-of-mind processes associated with social-cognitive brain networks. The depiction of a potential shooter approaching an innocent victim will engage affective systems. In this particular example, brightness and threat could be used as a time varying energetic source to use in a control theoretic analysis. The idea being, that each is analogous to an energetic source that should impact specific nodes (visual cortex, PFC, respectively) differently, and have different cascading impacts on time-varying network topology. Middle panel: A snapshot of network states in a single individual. As the individual views the movie, their brain responds to the time-varying content. Visual changes will prompt visual networks to come online and reconfigure (the example focuses only on brightness, but other visual features could be modeled, such as the presence or number of faces, contrast, objects, etc.). In parallel, higher-level content elements (though conveyed *via* the concrete sensory-perceptual features) prompt changes in networks related to executive control, salience processing, etc. Right panel, top: Example of an energy landscape in which valleys represent equilibrium states. By application of control energy, the brain can be pushed to leave one state and settle down in another. Note that the actual landscape has a higher dimensionality. Right panel, bottom: Example of a multivariate brain activity trajectory from actual movie-viewers. The input movie was Bang-Bang! You're dead by Alfred Hitchcock. Brain activity from different regions (268-dimensional) is embedded in a lower-dimensional (3-dimensions) space (Heusser et al., 2018). Color represents time. As can be seen, the time-varying movie content steers brain activity into different positions.

highway. Introducing the new highway might also alter the energetic maintenance costs of the highways that already existed in the network (e.g., the new highway may increase or decrease traffic flows on preexisting highways). Similar ideas can be applied to the brain as a network, although empirical evidence about energetic costs and network structure is less robust. For instance, the creation and maintenance of specific connections (white matter pathways) in the brain's network are associated with energetic costs, and the topological structure is associated with the kinds of states and functions that the network can settle into and perform (e.g., Margulies et al., 2016).

Network control theory can be used to calculate a number of metrics that describe these energetic costs. Importantly, some network topologies are easier to reach - requiring less energy to obtain - than others. How can these energetic requirements be quantified? One of the most common approaches is known as *controllability.* Controllability is the ability to transition a network from one topological organization to another *via* external energetic input (Kalman, 1962). This controllability metric can be further subdivided into *average controllability*, *modal controllability*, and *boundary controllability*.

Average controllability (Shaker and Tahavori, 2013) describes how much energy needs to be applied to the system to transition the system into a different topological organization. Higher average controllability means that less energy input is necessary to drive the system to different topological organizations. One constraint on average controllability is that it only captures how much energy it takes to move the system into easily reached topological organizations. *Modal controllability* (Hamdan and Nayfeh, 1989) accounts for circumstances where it takes substantial energy to transition the system into a hard-to-reach topological organization. Finally, *boundary controllability* (Bassett et al., 2013) identifies nodes within a network that, when targeted with energy, can elicit connection or disconnection among other nodes in the network. Together, these metrics provide insight into the energetic costs and target nodes necessary to drive a network from one topological organization to another.

3.2. How has network control theory already been applied?

Network control theory has been increasingly applied to study the controllability of structural and functional brain networks (Medaglia et al., 2017), but it is not confined to brains alone. Instead, it is also perfectly feasible to apply network control theory to social or psychological networks (Abelson, 1964; Cremonini and Casamassima, 2017; Borsboom et al., 2021). For example, in neurology and neuropsychology, one can use network control theory to examine how strokes at specific anatomical (structural) sites affect cognitive (functional) processes (Popova et al., 2022). Similarly, in the case of social networks, it becomes possible to ask how structural changes affect function (Proskurnikov and Tempo, 2017, 2018). For instance, how do changes in leadership structure impact a group, its communication, and ultimately performance? Finally, turning to psychological networks such as attitude and belief networks, network control theory enables simulating how targeted influence (e.g., message-based persuasion attempts geared towards a specific belief) would impact the targeted belief, its associates, and the belief network as a whole (Schlicht-Schmälzle et al., 2018; Chambon et al., 2022).

Turning specifically to brain organization, network control theory has revealed some crucial findings about brain structure and function. Possibly most important is that the brain's intrinsic architecture, that is, the white matter fiber tracts connecting gray matter structures, facilitate controllability in different ways. In a pathbreaking study, Gu et al. (2015) demonstrated that different neural subnetworks had different levels of controllability. For instance, the default mode network has a topological organization that facilitates transitions into other easily reached topological organizations. By comparison, other subnetworks (e.g., fronto-parietal control networks) are better suited to facilitate transitions into difficult-to-reach topological organizations. These controllability characteristics appear to guide high-level cognitive and behavioral responses within organisms (Rouse et al., 2021).

3.3. How can network control theory integrate media content with reception responses and media effects?

How can network control theory be applied to the media content → brain reception mechanism → media effects framework presented above, and what can we gain from it? In a nutshell, our core argument is that under a normal mass communication regime (i.e., one-to-many: same message, many recipients), the arrow of causality starts with the message content. Therefore, understanding the content is the key to understanding downstream effects.[8]

To give an example, consider the case of a movie that contains a morally evocative event, such as an innocent person being shot and killed.[9] Such key moments of the story (Wilensky, 1983) evoke predictable audience reactions that are highly consistent across viewers (Hasson et al., 2008a; Dmochowski et al., 2012; Naci et al., 2014; Schmälzle and Grall, 2020a,b). It is clear that flickering pixels, moving images, and so forth are required to transmit the movie into peoples' brains. However, the main "effective ingredient" of this content sits at a higher level of plot abstraction. We also know that filmmakers, screenwriters, and fiction authors are very skilled at "pushing" people into certain psychological states (see Figure 5). In fact, even the designation 'director' clearly alludes to the potential to exert control, that is, by influencing the content creation process in such a way that certain audience reactions follow predictably.

With neuroimaging, we can now capture how brain networks reconfigure dynamically during movie watching, such as how movie events trigger attentional reorienting responses, how close-up shots of protagonists are important events that evoke theory-of-mind processing, or how morality violations engage brain networks involved in emotion and socio-moral cognition. If we can successfully integrate these higher-level layers of the media's content with the more easily quantifiable characteristics of content that engage sensory and perceptual brain systems, then we can hope to close the explanatory gaps between movie content, reception response, and media effects under one cohesive framework.

To make this all more concrete, consider the following example: We know that simple narratives are easier to follow than complex ones. From a cognitive perspective, we further know that following a complex narrative taxes working memory. Neurally, we know that working memory is associated with (although not in a 1:1 fashion) activity in the executive control and default mode networks. Thus, at a very simple level, we might examine network controllability metrics for different narratives that vary in complexity, and we could expect that simple vs. complex narratives are associated with different controllability values.[10] Further, we might also ask if these controllability values can be used to predict box office revenues of a given narrative, much in the same way as Dmochowski and colleagues (Dmochowski et al., 2014) used neural reliabilities to predict audience preferences. In this case, we would link a high-level media content characteristic (plot complexity), with an equally high-level reception response (controllability), and media effect (box office revenue, a measure of popularity).

Of course, it should also be possible - and maybe more interesting - to apply the approach to a single movie to examine finer-grained elements along the media content, reception response, and media effects pathway. In this case, the time-varying properties of the movie would comprise the input to the system, i.e., the energy that is applied to the network. Mathematically, this can be modeled *via* impulse response models (Blaauw et al., 2017) when targeting a single node or more generalized control models (Tang and Bassett, 2018) that target multiple nodes in a network (for a review, see Lydon-Staley et al., 2020).

The question, then, is what type of media content we should model, to what node or nodes (targets) in the network the resulting energy would get applied, and what sort of outcomes we might expect? Although answers to these questions remain speculative because - to our knowledge - NCT has yet to be applied to content-rich media (as opposed to simpler stimuli and tasks), the cumulative body of knowledge from sensory and cognitive neuroscience, combined with nearly six decades of entertainment research and mass communication research can offer direction.

---

[8] We thank a reviewer for pointing out that based on the picture presented here (content → brain → effects), readers may infer that media reception is a strictly passive process, which we ultimately do not believe to be correct. Rather, there are additional receiver-sided factors that can affect the reception process. For example, the degree of interest among receivers can modulate how people respond to the same incoming message content; the same is true for the belief-consistency of a message, the background knowledge audiences have about a topic, or simply their degree of vigilance. Thus, in reality, the way in which audiences select and engage with media content is going to be more dynamic, creating message-receiver interactions beyond simple message main effects, potential dynamic feedback loops, and other audience effects (e.g., during co-viewing vs. individual viewing). In sum, real audiences are more active (e.g., Biocca, 1988; Huskey et al., 2020) and these factors must be taken into account. Nevertheless, even if these (or other) additional external or internal factors come into play, it is clear that the proximal causal role of media content is critical and must be quantified.

[9] From the perspective of Moral Foundations Theory and the Model of Intuitive Morality and Media Exemplars, this could be considered a violation of the harm/care foundation (Tamborini, 2011; Graham et al., 2013).

[10] Readers who are familiar with traditional approaches to fMRI data analysis, such as the GLM framework, will realize that this approach is conceptually similar, the main difference being that it is applied here to network metrics as the dependent variable rather than to the activity of individual voxels.

Starting with basic sensory and perceptual features, we can extract these in much the same way as is currently done for topical studies of vision, audition, or language (e.g., Kauttonen et al., 2015; McNamara et al., 2017), and we can relate quantified content properties (e.g., over-time variations in brightness, sound energy, etc.) to brain imaging measures. To the extent that the reception mechanisms that correspond to specific content properties are localized, one may not even need to resort to network-based analyses but could even rely on standard brain mapping-style analyses.

Then, as we move from simple features like brightness or sound energy to more complex media content, we need to not only adjust the kinds of content features that are quantified and used to model brain responses but also the kinds of brain response features that are modeled (i.e., moving from localized univariate response models to model networked responses and state-reorganizations, which is what network control theory excels in). With regard to the quantification of content, we argued above that it will no longer be sufficient to model pixels, brightness, or the occurrence of faces. Rather, media psychological research points to the importance of characters, the actions they perform and the outcomes that befall them, and so forth. Using this understanding (for a review, see Grizzard and Eden, 2022), the kinds of content we should attend to, and their putative brain targets become clearer. With regard to response features, we can rely on methods from network neuroscience, including parcellations of canonical brain networks, network estimation methods, and knowledge about structure–function relationships (e.g., between the TPJ, a core node of the DMN and social-affective processes, e.g., Yeshurun et al., 2017).

Imagine a researcher interested in empathy. Two narratives could be constructed, one where a liked character suffers a dramatic setback (which should elicit an empathetic response), and one where the setback is edited out (which should not elicit an empathetic response). The timing of this empathy-inducing outcome could be used in an impulse response model that targets a specific node in the network, like the temporal–parietal junction, which has long been implicated in empathy processing (Saxe and Kanwisher, 2003; Decety and Lamm, 2007; Alcalá-López et al., 2018).[11] Then, one would analyze how this intervention (i.e., film event) changes the brain network topology and how this differs between the experimental and control version of the narrative. Moving onwards, if a negative event befalling a liked character changes the brain network into a state of empathy, then that change should be associated with a corresponding change in audience responses (e.g., self-reported empathy), thus completing the sequence from media content, reception response, to media effects.

Another example could be suspense: We know that suspense in media strongly affects the audience, and screenwriters and directors possess a lot of knowledge about how to elicit this phenomenon (e.g., Brewer and Lichtenstein, 1982; Douchet, 1985; Vorderer et al., 2006). Moreover, some prior work has focused on the brain mechanisms of suspense precisely because of its potential to take control of audiences (Bezdek et al., 2017; Schmälzle and Grall, 2020a,b). Much like in the example about empathy above, it would be possible to create different branches of the same story that incorporate directing techniques, music, narrative devices, or other methods to increase suspense and examine their impact on brain systems.[12] Again, one could then analyze how variations in suspense (either between experimental conditions or variations of suspense over time) impact the brain network topology. One broad prediction, for example, is that ebbs and flows in suspense should impact the saliency and executive control networks, which are associated with attention. Although more difficult to resolve with present-day functional neuroimaging methods (because of limitations in spatial and temporal resolutions), suspense should also impact ascending arousal networks and cortico-subcortical loops associated with emotional arousal. Indeed, previous neuroimaging work points to such responses (e.g., Hermans et al., 2011; Naci et al., 2014; Young et al., 2017; Schmälzle and Grall, 2020a,b), but whereas much of this work is data-driven and more exploratory in nature, network control theory holds potential to integrate this research and provide a common platform for bringing together content (directors, creators), brain response (cognitive neuroscientists) and effects studies (media psychology and entertainment research).

These represent just a few possible examples that use network control theory as a framework that connects the domains of content analysis, reception analysis, and media effects. The appeal of network control theory is that it enables us to start from media-informed hypotheses about what will be driving brain network dynamics and how while honoring the complexity and hierarchical nature of the content (from simple objective features to more abstract semantic and pragmatic contents), brain responses (from evoked sensory responses to reorganization of higher-level brain systems), and media effects (from effects on individuals to populations, and from obligatory effects in all individuals to effects that could vary based on individual difference, cultural background, or an individual's position in a larger social network topology).

Although many unknowns and challenges remain,[13] this approach holds the potential to integrate domains that have henceforth been studied separately. Viewed from afar, this endeavor is almost reminiscent of the seminal work of Penfield (1950), who used intracranial stimulation techniques to map out functional brain systems, but with the difference that movies now offer a way to influence brain systems and associated affective, social, and conceptual reactions, and not only in individuals but multiple brains comprising an audience.[14]

---

[11] Of course, this approach requires a strong *a-priori* hypothesis. Data-driven approaches are also available. For instance, a researcher could, one by one, apply the energy source to each node in the network in a round-robin style, and observe the outcome.

[12] Of note, here we discuss only standard experimental paradigms, but it would not seem infeasible to even create closed-loop, neurofeedback-type systems that feedback audience activity into the creation process, thereby further enhancing collaboration and integration between filmmakers and neuroscience (e.g., Tikka et al., 2012; Raz and Hendler, 2014).

[13] Especially regarding the quantification of content that lies at or behind the "barrier of meaning" and the accurate measurement of the networked structures.

[14] We would like to thank a reviewer for suggesting that this all sounds a bit like a "content-powered TMS machine," an idea that we find thought-provoking and appropriate. However, the reviewer is also right to warn against overstretching this analogy because current TMS methods allow causal targeting of single (or few) and localized brain functions. By contrast, when

4. Future directions

4.1. From traditional mass media to new media

We are not the first to make arguments about the necessity of quantifying naturalistic and multi-modal media stimuli for understanding the brain, or media effects (see, e.g., Weber et al., 2006, 2015b; Spiers and Maguire, 2007; Dudai, 2012; Sonkusare et al., 2019; Aliko et al., 2020; Finn et al., 2022). Important work headed in this direction already exists, and we have worked to note these developments at relevant points in our manuscript. The point is, however, this approach has not yet reached widespread adoption. We think this is for two key reasons: (1) uncertainty about how to quantify media content, and (2) ambiguity about how to link content's complex, hierarchically organized, and time-varying effects across complex, hierarchically organized, and time-varying brain systems. The approach outlined above, which advocates jointly studying media content, reception responses, and media effects and suggests NCT as a framework for doing so, addresses these two challenges and is directly applicable to a wide variety of traditional mass media, including TV, cinema, and written or spoken narratives.

However, the notion of mass media today is no longer quite what it was when relevant definitions and theories of mass media were first formulated. Rather, these days the media ecosystem is constantly in flux, and new ways to stimulate brains and entertain audiences are constantly invented. Traditional mass media, most notably radio and television, followed the classical one-to-many model in which a sender emitted the same message that was carried *via* a medium to a large audience, like when people listened to Orson Wells' "War of the Worlds" broadcast that prompted them to fear an alien invasion. Similarly, TV and cinema movie viewing also fall under this kind of paradigm (same message, millions of simultaneous receivers), which is very compatible with neuroimaging and leads to a constant increase in papers and publicly available datasets featuring audiobooks and movies (Aliko et al., 2020; Willems et al., 2020).

The advent of streaming platforms (e.g., Netflix for movies and shows, YouTube for all kinds of content, Spotify for music) prompted a shift in the landscape because previously more homogenous mass audiences became increasingly fragmented and can now consume content at their own pace and *via* increasingly niche content. Despite the self-timed nature of such video streaming, however, the basic notion of same-message - many receivers still remains. Thus, these kinds of media models lend themselves exceptionally well to neuroscientific studies like the ones outlined above.

Social media add another layer of complexity, but we argue that key principles of mass communication still remain relevant. Modern social media, like Twitter and TikTok, can be characterized as instant mass media; that is, they deliver the same messages to many recipients in a very swift manner. Moreover, they add novel affordances to engage with content *via* liking, sharing, and commenting. The resulting mode of communication has been called "masspersonal communication" because it blends elements of interpersonal communication into the mass communication model (O'Sullivan and Carr, 2018). Thus, the content of social media messages can still be studied and linked to brain reception responses, and the additional affordances of social media (like sharing, liking, commenting) can also be studied from a neuroscientific perspective (Meshi et al., 2015; Scholz et al., 2019).

4.2. Games and virtual reality as emerging trends

Reflecting on what the future may hold, we see two areas on the rise: Gaming and Virtual Reality (VR). Gaming and VR are both among the fastest-growing media types. Both offer interaction potential,[15] distinguishing them from movies and stories (TV, radio, podcasts, etc.) that are consumed more passively, although even for the latter, audiences can vary in their level and degree of internal activity (e.g., interest, involvement, vigilance). At first glance, the interactive and thus constantly changing nature of gaming and VR media may seem incompatible with the "same stimulus sequence" notion that is so characteristic of movies, audiobooks, and other fixed-type mass media. However, we note that even in games and VR, there are clearly shared aspects as well and that the experiences users have are far from idiosyncratic. In games, for instance, many sub-scenes are prerecorded and thus the same for all audience members, and the same holds true for VR. Moreover, for both games and VR experiences, it is exceptionally well possible to quantify and precisely time-lock contents (Bente et al., 2007; Huskey et al., 2018a; Lammers et al., 2019; Calcagnotto et al., 2021).

Thus, although studying brain responses during games and VR will require special consideration, we argue again that the basic model outlined above still applies: As long as fixed content is consumed, one can code it just like one would do for movies or narratives (see above), and to the extent that content varies by person, one can still content-analyze each individual screen-recording using the same principles (Dmochowski et al., 2018; Huskey et al., 2018b, 2022; Ki et al., 2020).

5. Summary and conclusion

In sum, we have argued that the time is ripe for creating a new substantive science at the intersection of media and neuroscience. The

---

content "targets" brain systems (such as the TPJ in the empathy example above), the TPJ would not receive direct input from content, but its input would consist of the preprocessed visual and auditory information conveying the empathy evoking narrative. Despite the caveats, we believe that the notion that media can precisely steer neurocognitive processes and evoke strong effects like emotional arousal, empathy, and so forth is convincing and that the NCT framework provides a way to examine how this is mediated by brain networks and their dynamic reconfigurations.

[15] Clearly, interactive media add immense complexity to the simple linear-causal content-reception-effects perspective offered for movies and comparable consumption media. Note that even for these types, however, there is interaction potential insofar as exposure to movies can shape preferences, thereby affecting future selection decisions and so forth. Thus, the overall picture is clearly more dynamic than described here. Yet, even a complex, convoluted episode with interactive media can be disentangled and causally arranged along the 'arrow of time', for which the content, reception, response framework should still hold.

neuroscientifically informed study of media reception processes provides the missing link between media content and media effects, enabling fascinating insights into the hidden mechanisms by which media affect us. However, to avoid reinventing the wheel or creating a mayfly-like field, neuroscientists should engage with research that has studied media content and media effects. The current article offers a springboard for doing so. We have introduced an organizing framework that connects the domains of media content, media reception, and media effects in a logical, sequential manner. In that framework, content is the key to understanding brain responses and subsequent media effects. We then suggested network control theory as a way to link the domains of media content, media reception mechanisms, and media effects (in individuals and social networks) in one multi-layered (or multi-staged) network. This framework offers a clear agenda for future research that uses media in combination with neural or other reception response measures and applies to studies focusing on specific neurocognitive processes (e.g., vision, language, or memory) as well as more integrative investigations of audience responses to movies and narratives. The ideas articulated here are most directly applicable to one-to-many mass communication models (which include neurocinematics, neuroscience of stories, etc.) but can also be adapted to modern social and interactive media.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abelson, R. P. (1964). Mathematical models of the distribution of attitudes under controversy. Contributions to Mathematical Psychology. eds. N. Fredericksen and H. Gullicksen, Holt, Rinehart & Winston. Available at: https://ci.nii.ac.jp/naid/10020196760/

Alcalá-López, D., Vogeley, K., Binkofski, F., and Bzdok, D. (2018). Building blocks of social cognition: Mirror, mentalize, share? *Cortex* 118, 4–18. doi: 10.1016/j.cortex.2018.05.006

Aliko, S., Huang, J., Gheorghiu, F., Meliss, S., and Skipper, J. I. (2020). A naturalistic neuroimaging database for understanding the brain using ecological stimuli. *Sci. Data* 7:347. doi: 10.1038/s41597-020-00680-2

Amir, O., Biederman, I., Wang, Z., and Xu, X. (2015). Ha ha! versus aha! a direct comparison of humor to nonhumorous insight for determining the neural correlates of mirth. *Cereb. Cortex* 25, 1405–1413. doi: 10.1093/cercor/bht343

Arana, S., Lerousseau, J. P., and Hagoort, P. (2023). Deep learning models to study sentence comprehension in the human brain. *arXiv*

Baek, E. C., and Parkinson, C. (2022). Shared understanding and social connection: integrating approaches from social psychology, social network analysis, and neuroscience. *Soc. Personal. Psychol. Compass* 16:e12710. doi: 10.1111/spc3.12710

Baek, E. C., Scholz, C., O'Donnell, M. B., and Falk, E. B. (2017). The value of sharing information: a neural account of information transmission. *Psychol. Sci.* 28, 851–861. doi: 10.1177/0956797617695073

Baldwin, J., and Schmälzle, R. (2022). A character recognition tool for automatic detection of social characters in visual media content. Computational. *Commun. Res.* 4, 350–371. doi: 10.5117/ccr2022.1.010.bald

Bandura, A. (1977). *Social learning theory*, Englewood Cliffs, N.J: Prentice Hall.

Bandura, A. (1994). "Social cognitive theory of mass communication" in *Media effects: Advances in theory and research*. eds. J. Bryant and D. Zillmann (Hillsdale, NJ: Erlbaum), 61–90.

Bassett, D. S., Porter, M. A., Wymbs, N. F., Grafton, S. T., Carlson, J. M., and Mucha, P. J. (2013). Robust detection of dynamic community structure in networks. *Chaos* 23:013142. doi: 10.1063/1.4790830

Bender, E. M., and Koller, A. (2020). "Climbing towards NLU: On meaning, form, and understanding in the age of data." in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. pp. 5185–5198.

Bente, G., Eschenburg, F., and Krämer, N. C. (2007). Virtual gaze. A pilot study on the effects of computer simulated gaze in avatar-based conversations. *In Virtual Reality: Second International Conference, ICVR 2007, Held as part of HCI International 2007* (Beijing, China: Springer Berlin Heidelberg), 185–194.

Berkman, E. T., and Falk, E. B. (2013). Beyond brain mapping: using neural measures to predict real-world outcomes. *Curr. Dir. Psychol. Sci.* 22, 45–50. doi: 10.1177/0963721412469394

Berns, G. S., and Moore, S. E. (2012). A neural predictor of cultural popularity. *J. Consum. Psychol.* 22, 154–160. doi: 10.1016/j.jcps.2011.05.001

Bezdek, M. A., Wenzel, W. G., and Schumacher, E. H. (2017). The effect of visual and musical suspense on brain activation and memory during naturalistic viewing. *Biol. Psychol.* 129, 73–81. doi: 10.1016/j.biopsycho.2017.07.020

Biocca, F. A. (1988). Opposing conceptions of the audience: the active and passive hemispheres of mass communication theory. *Ann. Int. Commun. Assoc.* 11, 51–80. doi: 10.1080/23808985.1988.11678679

Blaauw, F. J., van der Krieke, L., Emerencia, A. C., Aiello, M., and de Jonge, P. (2017). Personalized advice for enhancing well-being using automated impulse response analysis. *arXiv*

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., et al. (2021). Network analysis of multivariate data in psychological science. *Nat. Rev. Dis. Primers* 1, 1–18. doi: 10.1038/s43586-021-00055-w

Brewer, W. F., and Lichtenstein, E. H. (1982). Stories are to entertain: a structural-affect theory of stories. *J. Pragmat.* 6, 473–486. doi: 10.1016/0378-2166(82)90021-2

Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., and Lalor, E. C. (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Curr. Biol.* 28, 803–809.e3. doi: 10.1016/j.cub.2018.01.080

Broom, T. W., Chavez, R. S., and Wagner, D. D. (2021). Becoming the king in the north: identification with fictional characters is associated with greater self-other neural overlap. *Soc. Cogn. Affect. Neurosci.* 16, 541–551. doi: 10.1093/scan/nsab021

Bryant, J., and Oliver, M. B. (2008). *Media effects: Advances in theory and research. 3rd* Edn, New York: Routledge.

Bryant, J., and Zillmann, D. (1990). *Responding to the screen: Reception and reaction processes, Routledge Communication Series*, Mahwah, N.J: Routledge.

Bullmore, E., and Sporns, O. (2012). The economy of brain network organization. *Nat. Rev. Neurosci.* 13, 336–349. doi: 10.1038/nrn3214

Busch, E. L., Huang, J., Benz, A., Wallenstein, T., Lajoie, G., Wolf, G., et al. (2022). Multi-view manifold learning of human brain state trajectories. *bioRxiv*:2022.05.03.490534. doi: 10.1101/2022.05.03.490534

Byrge, L., Dubois, J., Tyszka, J. M., Adolphs, R., and Kennedy, D. P. (2015). Idiosyncratic brain activation patterns are associated with poor social comprehension in autism. *J. Neurosci.* 35, 5837–5850. doi: 10.1523/JNEUROSCI.5182-14.2015

Calcagnotto, L., Huskey, R., and Kosicki, G. M. (2021). The accuracy and precision of measurement. *Comp. Commun. Res.* 3, 1–20. doi: 10.5117/CCR2021.2.001.CALC

Chalupa, L. M., and Werner, J. S. (2003). *The visual neurosciences, Bradford Books*. Cambridge, MA: The MIT Press.

Chambon, M., Dalege, J., Waldorp, L. J., Van der Maas, H. L. J., Borsboom, D., and van Harreveld, F. (2022). Tailored interventions into broad attitude networks towards the COVID-19 pandemic. *PLoS One* 17:e0276439. doi: 10.1371/journal.pone.0276439

Chang, L. J., Jolly, E., Cheong, J. H., Rapuano, K. M., Greenstein, N., Chen, P.-H. A., et al. (2021). Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. Science. *Advances* 7:eabf7129. doi: 10.1126/sciadv.abf7129

Chen, P.-H. A., Jolly, E., Cheong, J. H., and Chang, L. J. (2020). Intersubject representational similarity analysis reveals individual variations in affective experience when watching erotic movies. *NeuroImage* 216:116851. doi: 10.1016/j.neuroimage.2020.116851

Choi, S., On, K.-W., Heo, Y.-J., Seo, A., Jang, Y., Lee, M., et al. (2021). DramaQA: character-centered video story understanding with hierarchical QA. *Proc. AAAI Conf. Artif. Intell.* 35, 1166–1174. doi: 10.48550/arXiv.2005.03356

Coronel, J. C., O'Donnell, M. B., Pandey, P., Delli Carpini, M. X., and Falk, E. B. (2021). Political humor, sharing, and remembering: insights from neuroimaging. *J. Commun.* 71, 129–161. doi: 10.1093/joc/jqaa041

Cremonini, M., and Casamassima, F. (2017). Controllability of social networks and the strategic use of random information. *Comput. Soc. Netw.* 4:10. doi: 10.1186/s40649-017-0046-2

Çukur, T., Nishimoto, S., Huth, A. G., and Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nat. Neurosci.* 16, 763–770. doi: 10.1038/nn.3381

Cummins, R. (2000). Two conceptions of psychological explanation. In Explanation and Cognition, eds. F. Keil and R. Wilson (Cambridge, MA, USA: MIT Press). 117–145.

De la Vega, A., Rocca, R., Blair, R. W., Markiewicz, C. J., Mentch, J., Kent, J. D., et al. (2022). Neuroscout, a unified platform for generalizable and reproducible fMRI research. *bioRxiv*:2022.04.05.487222. doi: 10.1101/2022.04.05.487222

Decety, J., and Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* 13, 580–593. doi: 10.1177/1073858407304

DeFleur, M. L. (2016). *Mass communication theories: Explaining origins, processes, and effects*, New York: Routledge.

Di, X., and Biswal, B. B. (2020). Intersubject consistent dynamic connectivity during natural vision revealed by functional MRI. *NeuroImage* 216:116698. doi: 10.1016/j.neuroimage.2020.116698

Dill, K. E. (2013). *The Oxford handbook of media psychology*, Oxford University Press.

Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., and Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nat. Commun.* 5:4567. doi: 10.1038/ncomms5567

Dmochowski, J. P., Ki, J. J., DeGuzman, P., Sajda, P., and Parra, L. C. (2018). Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity. *NeuroImage*, 180, 134–146. doi: 10.1016/j.neuroimage.2017.05.037

Dmochowski, J. P., Sajda, P., Dias, J., and Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Front. Hum. Neurosci.* 6:112. doi: 10.3389%2Ffnhum.2012.00112

Douchet, J. (1985). "Hitchcock and his audience" in *Cahiers du Cinéma, the 1950s: Neo-realism, hollywood, new wave*. ed. J. Hillier (Cambridge, MA , USA: Harvard University Press), 150–157.

Dudai, Y. (2012). The cinema-cognition dialogue: a match made in brain. *Front. Hum. Neurosci.* 6:248. doi: 10.3389/fnhum.2012.00248

Falk, E. B., Berkman, E. T., and Lieberman, M. D. (2012). From neural responses to population behavior: neural focus group predicts population-level media effects. *Psychol. Sci.* 23, 439–445. doi: 10.1177/095679761143496

Falk, E. B., Berkman, E. T., Mann, T., Harrison, B., and Lieberman, M. D. (2010). Predicting persuasion-induced behavior change from the brain. *J. Neurosci.* 30, 8421–8424. doi: 10.1523/JNEUROSCI.0063-10.2010

Federman, J. (1995). *National Television Violence Study: Executive summary* (Editor/University of California, Santa Barbara).

Finn, E. S., Glerean, E., Hasson, U., and Vanderwal, T. (2022). Naturalistic imaging: the use of ecologically valid conditions to study brain function. *NeuroImage* 247:118776. doi: 10.1016/j.neuroimage.2021.118776

Fuster, J. M. (2003). *Cortex and mind: Unifying cognition*, Oxford, UK: Oxford University Press.

Gantman, A. P., and Van Bavel, J. J. (2014). The moral pop-out effect: enhanced perceptual awareness of morally relevant stimuli. *Cognition* 132, 22–29. doi: 10.1016/j.cognition.2014.02.007

Gong, X., Huskey, R., Xue, H., Shen, C., and Frey, S. (2022). Broadcast information diffusion processes on social media networks: exogenous events lead to more integrated public discourse. *MediArXiv*. doi: 10.33767/osf.io/wautz

Grady, S. M., Schmälzle, R., and Baldwin, J. (2022). Examining the relationship between story structure and audience response. *PRO* 16, 1–28. doi: 10.3167/proj.2022.160301

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., et al. (2013). "Moral foundations theory: the pragmatic validity of moral pluralism" in *Advances in experimental social psychology*, vol. *47* (Amsterdam, The Netherlands: Elsevier), 55–130.

Greenberg, B. S., and Atkin, C. K. (1980). *Life on television: Content analyses of U.S. TV Drama*, New York City, New York, USA: Ablex Publishing Corporation.

Grizzard, M., and Eden, A. (2022). The character engagement and moral adjustment model (CEMAM): a synthesis of more than six decades of research. *J. Broadcast. Electron. Media* 66, 698–722. doi: 10.1080/08838151.2022.2146116

Grubbs, J. B., and Kraus, S. W. (2021). Pornography use and psychological science: a call for consideration. *Curr. Dir. Psychol. Sci.* 30, 68–75. doi: 10.1177/096372142097959

Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., et al. (2015). Controllability of structural brain networks. *Nat. Commun.* 6:8414. doi: 10.1038/ncomms9414

Hahn, L., Tamborini, R., Prabhu, S., Klebig, B., Grall, C., and Pei, D. (2017). The importance of altruistic versus egoistic motivations: a content analysis of conflicted motivations in children's television programming. *Commun. Rep.* 30, 67–79. doi: 10.1080/08934215.2016.1251602

Hamdan, A. M. A., and Nayfeh, A. H. (1989). Measures of modal controllability and observability for first- and second-order linear systems. *J. Guid. Control. Dyn.* 12, 421–428. doi: 10.2514/3.20424

Hasson, U., Furman, O., Clark, D., Dudai, Y., and Davachi, L. (2008a). Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron* 57, 452–462. doi: 10.1016/j.neuron.2007.12.009

Hasson, U., and Honey, C. J. (2012). Future trends in neuroimaging: neural processes as expressed within real-life contexts. *NeuroImage* 62, 1272–1278. doi: 10.1016/j.neuroimage.2012.02.004

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D. J. (2008b). Neurocinematics: the neuroscience of film. *PRO* 2, 1–26. doi: 10.3167/proj.2008.020102

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008c). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* 28, 2539–2550. doi: 10.1523/JNEUROSCI.5487-07.2008

Heider, F., and Simmel, M. (1944). An experimental study of apparent behavior. *Am. J. Psychol.* 57, 243–259. doi: 10.2307/1416950

Henson, R. (2006). Forward inference using functional neuroimaging: dissociations versus associations. *Trends Cogn. Sci.* 10, 64–69. doi: 10.1016/j.tics.2005.12.005

Hermans, E. J., van Marle, H. J. F., Ossewaarde, L., Henckens, M. J. A. G., Qin, S., van Kesteren, M. T. R., et al. (2011). Stress-related noradrenergic activity prompts large-scale neural network reconfiguration. *Science* 334, 1151–1153. doi: 10.1126/science.1209603

Heusser, A. C., Ziman, K., Owen, L. L. W., and Manning, J. R. (2018). HyperTools: a python toolbox for gaining geometric insights into high-dimensional data. *J. Mach. Learn. Res.* 18, 1–6.

Honey, C. J., Thompson, C. R., Lerner, Y., and Hasson, U. (2012). Not lost in translation: neural responses shared across languages. *J. Neurosci.* 32, 15277–15283. doi: 10.1523/JNEUROSCI.1800-12.2012

Hopp, F. R., Fisher, J. T., and Weber, R. (2020). A graph-learning approach for detecting moral conflict in movie scripts. *Media Commun.* 8, 164–179. doi: 10.17645/mac.v8i3.3155

Hovland, C. I., and Lumsdaine, A. A. (2017). *Experiments on mass communication*, Princeton, NJ: Princeton University Press.

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154. doi: 10.1113/jphysiol.1962.sp006837

Huskey, R., Bue, A. C., Eden, A., Grall, C., Meshi, D., Prena, K., et al. (2020). Marr's tri-level framework integrates biological explanation across communication subfields. *J. Commun.* 70, 356–378. doi: 10.1093/joc/jqaa007

Huskey, R., Craighead, B., Miller, M. B., and Weber, R. (2018b). Does intrinsic reward motivate cognitive control? A naturalistic-fMRI study based on the synchronization theory of flow. *Cogn. Affect. Behav. Neurosci.* 18, 902–924. doi: 10.3758/s13415-018-0612-6

Huskey, R., Keene, J. R., Wilcox, S., Gong, X., Adams, R., and Najera, C. J. (2022). Flexible and modular brain network dynamics characterize flow experiences during media use: a functional magnetic resonance imaging study. *J. Commun.*, 72, 6–32. doi: 10.1093/joc/jqab044

Huskey, R., Mangus, J. M., Turner, B. O., and Weber, R. (2017). The persuasion network is modulated by drug-use risk and predicts anti-drug message effectiveness. *Soc. Cogn. Affect. Neurosci.* 12, 1902–1915. doi: 10.1093/scan/nsx126

Huskey, R., Wilcox, S., and Weber, R. (2018a). Network neuroscience reveals distinct neuromarkers of flow during media use. *J. Commun.* 68, 872–895. doi: 10.1093/joc/jqy043

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., and Gallant, J. L. (2016a). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532, 453–458. doi: 10.1038/nature17637

Huth, A. G., Lee, T., Nishimoto, S., Bilenko, N. Y., Vu, A., and Gallant, J. L. (2016b). Decoding the semantic content of natural movies from human brain activity. *Front. Syst. Neurosci.* 10, 453–458. doi: 10.3389/fnsys.2016.00081

Jääskeläinen, I. P., Pajula, J., Tohka, J., Lee, H.-J., Kuo, W.-J., and Lin, F.-H. (2016). Brain hemodynamic activity during viewing and re-viewing of comedy movies explained by experienced humor. *Sci. Rep.* 6:27741. doi: 10.1038/srep27741

Kalman, R. E. (1962). Canonical structure of linear dynamical systems. *Proc. Natl. Acad. Sci. U. S. A.* 48, 596–600. doi: 10.1073%2Fpnas.48.4.596

Karpathy, A. (2012). The state of computer vision. Available at: http://karpathy.github.io/2012/10/22/state-of-computer-vision/ [Accessed December 3, 2022].

Kauttonen, J., Hlushchuk, Y., Jääskeläinen, I. P., and Tikka, P. (2018). Brain mechanisms underlying cue-based memorizing during free viewing of movie memento. *NeuroImage* 172, 313–325. doi: 10.1016/j.neuroimage.2018.01.068

Kauttonen, J., Hlushchuk, Y., and Tikka, P. (2015). Optimizing methods for linking cinematic features to fMRI data. *NeuroImage* 110, 136–148. doi: 10.1016/j.neuroimage.2015.01.063

Kepplinger, H. M. (1989). Content analysis and reception analysis. *Am. Behav. Sci.* 33, 175–182. doi: 10.1177/0002764289033002010

Ki, J. J., Parra, L. C., and Dmochowski, J. P. (2020). Visually evoked responses are enhanced when engaging in a video game. *Eur. J. Neurosci.* 52, 4695–4708. doi: 10.1111/ejn.14924

Kim, J. Z., and Bassett, D. S. (2020). "Linear dynamics and control of brain networks" in *Neural Engineering*. ed. B. He (Heidelberg, Germany: Springer International Publishing), 497–518.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.

Klimmt, C., Hartmann, T., and Schramm, H. (2006). "Parasocial interactions and relationships" in *Psychology of Entertainment*. ed. P. Vorderer, Lawrence Erlbaum Associates Publishers. 291–313.

Knutson, B., and Genevsky, A. (2018). Neuroforecasting aggregate choice. *Curr. Dir. Psychol. Sci.* 27, 110–115. doi: 10.1177/0963721417737877

Kosslyn, S. M. (1999). If neuroimaging is the answer, what is the question? *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 354, 1283–1294. doi: 10.1098%2Frstb.1999.0479

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*, Newbury Park, California: SAGE.

Lammers, S., Bente, G., Tepest, R., Jording, M., Roth, D., and Vogeley, K. (2019). Introducing ACASS: an annotated character animation stimulus set for controlled (e)motion perception studies. *Front. Robot. AI* 6:94. doi: 10.3389/frobt.2019.00094

Lang, A. (2013). Discipline in crisis? The shifting paradigm of mass communication research. *Commun. Theory* 23, 10–24. doi: 10.1111/comt.12000

Lang, A. (2014). *Measuring psychological responses to media messages*, New York: Routledge.

Lang, A., and Ewoldsen, D. (2010). Beyond effects: Conceptualizing communication as dynamic, complex, nonlinear, and fundamental. Rethinking communication: Keywords in communication research, 111–122.

Larzabal, C., Bacon-Macé, N., Muratot, S., and Thorpe, S. J. (2017). Waking up buried memories of old TV programs. *Front. Behav. Neurosci.* 11:60. doi: 10.3389/fnbeh.2017.00060

Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* 31, 2906–2915. doi: 10.1523/JNEUROSCI.3684-10.2011

Lippmann, W. (1922). *Public opinion*. New York: Transaction Publisher.

Littlejohn, S. W., and Foss, K. A. (2009). *Encyclopedia of communication theory*, Newbury Park, California: SAGE.

Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L. (2011). Controllability of complex networks. *Nature* 473, 167–173. doi: 10.1038/nature10011

Lydon-Staley, D. M., Cornblath, E. J., Blevins, A. S., and Bassett, D. S. (2020). Modeling brain, symptom, and behavior in the winds of change. *Neuropsychopharmacology* 46, 20–32. doi: 10.1038/s41386-020-00805-6

Malik, M., Hopp, F. R., and Weber, R. (2022). Representations of racial minorities in popular movies. Computational. *Commun. Res.* 4, 208–253. doi: 10.5117/ccr2022.1.006.mali

Mantini, D., Hasson, U., Betti, V., Perrucci, M. G., Romani, G. L., Corbetta, M., et al. (2012). Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. *Nat. Methods* 9, 277–282. doi: 10.1038/nmeth.1868

Marcus, G., and Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*, New: York Knopf Doubleday Publishing Group.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., et al. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12574–12579. doi: 10.1073/pnas.1608282113

Masters, R. D., Frey, S., and Bente, G. (1991). Dominance & attention: images of leaders in German, French, & American TV news. *Polity* 23, 373–394. doi: 10.2307/3235132

McClelland, J. L., Hill, F., Rudolph, M., Baldridge, J., and Schütze, H. (2020). Placing language in an integrated understanding system: next steps toward human-level performance in neural language models. *Proc. Natl. Acad. Sci. U. S. A.* 117, 25966–25974. doi: 10.1073/pnas.1910416117

McNamara, Q., de la Vega, A., and Yarkoni, T. (2017). Developing a comprehensive framework for multimodal feature extraction. *arXiv*

Medaglia, J. D., Lynall, M. E., and Bassett, D. S. (2015). Cognitive network neuroscience. *J. Cogn. Neurosci.* 27, 1471–1491. doi: 10.1162/jocn_a_00810

Medaglia, J. D., Pasqualetti, F., Hamilton, R. H., Thompson-Schill, S. L., and Bassett, D. S. (2017). Brain and cognitive reserve: translation via network control theory. *Neurosci. Biobehav. Rev.* 75, 53–64. doi: 10.1016/j.neubiorev.2017.01.016

Meer, J. N.Van Der, Breakspear, M., Chang, L. J., Sonkusare, S., and Cocchi, L. (2020). Movie viewing elicits rich and reliable brain state dynamics. *Nat. Commun.*, 11,:5004. doi:10.1038/s41467-020-18717-w

Meshi, D., Tamir, D. I., and Heekeren, H. R. (2015). The emerging neuroscience of social media. *Trends Cogn. Sci.* 19, 771–782. doi: 10.1016/j.tics.2015.09.004

Mesulam, M. M. (1998). From sensation to cognition. *Brain* 121, 1013–1052. doi: 10.1093/brain/121.6.1013

Meyer, C., Padmala, S., and Pessoa, L. (2019). Dynamic threat processing. *J. Cogn. Neurosci.* 31, 522–542. doi: 10.1162/jocn_a_01363

Mitchell, M. (2020). On crashing the barrier of meaning in artificial intelligence. *AI Mag.* 41, 86–92. doi: 10.1609/aimag.v41i2.5259

Muldoon, S. F., Pasqualetti, F., Gu, S., Cieslak, M., Grafton, S. T., Vettel, J. M., et al. (2016). Stimulation-based control of dynamic brain networks. *PLoS Comput. Biol.* 12:e1005076. doi: 10.1371/journal.pcbi.1005076

Nabi, R. L., and Oliver, M. B. (2009). *The SAGE handbook of media processes and effects*, Newbury Park, California: Sage.

Naci, L., Cusack, R., Anello, M., and Owen, A. M. (2014). A common neural code for similar conscious experiences in different individuals. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14277–14282. doi: 10.1073/pnas.1407007111

Neuendorf, K. A. (2017). *The content analysis guidebook*, ewbury Park, California: SAGE.

Newman, M. (2010). Networks. OUP.

Nguyen, M., Vanderwal, T., and Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage* 184, 161–170. doi: 10.1016/j.neuroimage.2018.09.010

Ni, W., and Coupé, C. (2023). Time-synchronic comments on video streaming website reveal core structures of audience engagement in movie viewing. *Front. Psychol.* 13:1040755. doi: 10.3389/fpsyg.2022.1040755

Noelle-Neumann, E. (1991). The theory of public opinion: the concept of the spiral of silence. *Ann. Int. Commun. Assoc.* 14, 256–287. doi: 10.1080/23808985.1991.11678790

O'Donnell, M. B., and Falk, E. B. (2015). Big data under the microscope and brains in social context: integrating methods from computational social science and neuroscience. *Ann. Am. Acad. Pol. Soc. Sci.* 659, 274–289. doi: 10.1177/0002716215569446

O'Sullivan, P. B., and Carr, C. T. (2018). Masspersonal communication: a model bridging the mass-interpersonal divide. *New Media Soc.* 20, 1161–1180. doi: 10.1177/1461444816686104

Oatley, K. (2002). "Emotions and the story worlds of fiction" in *Narrative Impact: Social and Cognitive Foundations*, M C. Green, J. Strange, and T. C. Brock (Psychology Press) vol. 39, 69.

Okdie, B. M., Ewoldsen, D. R., Muscanell, N. L., Guadagno, R. E., Eno, C. A., Velez, J. A., et al. (2014). Missed programs (you can't TiVo this one): why psychologists should study media. *Perspect. Psychol. Sci.* 9, 180–195. doi: 10.1177/1745691614521243

Penfield, W. (1950). The cerebral cortex of man: a clinical study of localization of function. JAMA: the. *J. Am. Med. Assoc.* 144:1412. doi: 10.1001/jama.1950.02920160086033

Perloff, R. M. (2002). "The third-person effect" in *Media effects*, eds. J. Bryant and M. B. Oliver (New York: Routledge), 499–516.

Poeppel, D., Mangun, G. R., and Gazzaniga, M. S. (2020). *The cognitive neurosciences*, Cambridge, MA , USA: MIT Press.

Popova, M., Fakhar, K., and Braun, W. (2022). "One region to control them all"- the surprising effectiveness of network control theory in predicting post-stroke recovery from aphasia. *Front. Comput. Neurosci.* 16:943396. doi: 10.3389/fncom.2022.943396

Potter, R. F., and Bolls, P. (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*, New: York Routledge.

Prause, N., Steele, V. R., Staley, C., and Sabatinelli, D. (2015). Late positive potential to explicit sexual images associated with the number of sexual intercourse partners. *Soc. Cogn. Affect. Neurosci.* 10, 93–100. doi: 10.1093/scan/nsu024

Proskurnikov, A. V., and Tempo, R. (2017). A tutorial on modeling and analysis of dynamic social networks. Part I. *Annu. Rev. Control.* 43, 65–79. doi: 10.1016/j.arcontrol.2017.03.002

Proskurnikov, A. V., and Tempo, R. (2018). A tutorial on modeling and analysis of dynamic social networks. Part II. *Annu. Rev. Control.* 45, 166–190. doi: 10.1016/j.arcontrol.2017.03.002

Radford, A., Kim, J. W., Xu, T., Brockman, G., Mcleavey, C., and Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. Available at: https://cdn.openai.com/papers/whisper.pdf [Accessed December 4, 2022].

Rains, S. A., Levine, T. R., and Weber, R. (2018). Sixty years of quantitative communication research summarized: lessons from 149 meta-analyses. *Ann. Int. Commun. Assoc.* 42, 105–124. doi: 10.1080/23808985.2018.1446350

Raney, A. A. (2004). Expanding disposition theory: reconsidering character liking, moral evaluations, and enjoyment. *Commun. Theory* 14, 348–369. doi: 10.1111/j.1468-2885.2004.tb00319.x

Raz, G., and Hendler, T. (2014). Forking cinematic paths to the self: neurocinematically informed model of empathy in motion pictures. *PRO* 8, 89–114. doi: 10.3167/proj.2014.080206

Raz, G., Jacob, Y., Gonen, T., Winetraub, Y., Flash, T., Soreq, E., et al. (2014). Cry for her or cry with her: context-dependent dissociation of two modes of cinematic empathy reflected in network cohesion dynamics. *Soc. Cogn. Affect. Neurosci.* 9, 30–38. doi: 10.1093/scan/nst052

Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., et al. (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cogn. Affect. Behav. Neurosci.* 16, 709–723. doi: 10.3758/s13415-016-0425-4

Regev, M., Simony, E., Lee, K., Tan, K. M., Chen, J., and Hasson, U. (2018). Propagation of information along the cortical hierarchy as a function of attention while reading and listening to stories. *Cereb. Cortex* 34, 33–50. doi: 10.1093/cercor/bhy282

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., and Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nat. Commun.* 9:1027. doi: 10.1038/s41467-018-03399-2

Riff, D., Lacy, S., and Fico, F. (2014). *Analyzing media messages: Using quantitative content analysis in research*, New: York Routledge.

Rohrbach, A., Torabi, A., Rohrbach, M., Tandon, N., Pal, C., Larochelle, H., et al. (2017). Movie description. *Int. J. Comput. Vis.* 123, 94–120. doi: 10.1007/s11263-016-0987-1

Rouse, T. C., Ni, A. M., Huang, C., and Cohen, M. R. (2021). Topological insights into the neural basis of flexible behavior. *bioRxiv*:2021.09.24.461717. doi: 10.1101/2021.09..461717

Saarimäki, H. (2021). Naturalistic stimuli in affective neuroimaging: a review. *Front. Hum. Neurosci.* 15:675068. doi: 10.3389/fnhum.2021.675068

Salmi, J., Roine, U., Glerean, E., Lahnakoski, J., Nieminen-von Wendt, T., Tani, P., et al. (2013). The brains of high functioning autistic individuals do not synchronize with those of others. *NeuroImage Clin.* 3, 489–497. doi: 10.1016/j.nicl.2013.10.011

Sawahata, Y., Komine, K., Morita, T., and Hiruma, N. (2013). Decoding humor experiences from brain activity of people viewing comedy movies. *PLoS One* 8:e81009. doi: 10.1371/journal.pone.0081009

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". *NeuroImage* 19, 1835–1842. doi: 10.1016/s1053-8119(03)00230-1

Schlicht-Schmälzle, R., Chykina, V., and Schmälzle, R. (2018). An attitude network analysis of post-national citizenship identities. *PLoS One* 13:e0208241. doi: 10.1371/journal.pone.0208241

Schmälzle, R. (2022). Theory and method for studying how media messages prompt shared brain responses along the sensation-to-cognition continuum. *Commun. Theory* 32, 450–460. doi: 10.1093/ct/qtac009

Schmälzle, R., and Grall, C. (2020a). "Mediated messages and synchronized brains" in *Hanbdbook of communication and biology*. eds. R. Weber and K. Floyd, Routledge. 109–122.

Schmälzle, R., and Grall, C. (2020b). The coupled brains of captivated audiences: an investigation of the collective brain dynamics of an audience watching a suspenseful film. *J. Media Psychol.*, 32, 1–13. doi: 10.1027/1864-1105/a000271

Schmälzle, R., Imhof, M. A., Grall, C., Flaisch, T., and Schupp, H. T. (2017). Reliability of fMRI time series: similarity of neural processing during movie viewing. *Biorxiv*. doi: 10.1101/158188

Schmälzle, R., Wilcox, S., and Jahn, N. T. (2022). Identifying moments of peak audience engagement from brain responses during story listening. *Commun. Monogr.*, 32, 1–24. doi: 10.1080/03637751.2022.2032229

Scholz, C., Jovanova, M., Baek, E. C., and Falk, E. B. (2019). Media content sharing as a value-based decision. *Curr. Opin. Psychol.*, 31, 83–88. doi: 10.1016/j.copsyc.2019.08.004

Shaker, H. R., and Tahavori, M. (2013). Optimal sensor and actuator location for unstable systems. *J. Vib. Control.* 19, 1915–1920. doi: 10.1177/107754631245130

Sliwa, J., and Freiwald, W. A. (2017). A dedicated network for social interaction processing in the primate brain. *Science* 356, 745–749. doi: 10.1126/science.aam6383

Sonkusare, S., Breakspear, M., and Guo, C. (2019). Naturalistic stimuli in neuroscience: critically acclaimed. *Trends Cogn. Sci.* 23, 699–714. doi: 10.1016/j.tics.2019.05.004

Spiers, H. J., and Maguire, E. A. (2007). Decoding human brain activity during real-world experiences. *Trends Cogn. Sci.* 11, 356–365. doi: 10.1016/j.tics.2007.06.002

Tamborini, R. (2011). Moral intuition and media entertainment. *J. Media Psychol.* 23, 39–45. doi: 10.1027/1864-1105/a000031

Tang, E., and Bassett, D. S. (2018). Colloquium: control of dynamics in brain networks. *Rev. Mod. Phys.*, London, England, United Kingdom. 90:031003. doi: 10.1103/RevModPhys.90.031003

Tannenbaum, P. H. (2014). *The entertainment functions of television*, London, England, United Kingdom: Psychology Press.

Tikka, P., Kauttonen, J., and Hlushchuk, Y. (2018). Narrative comprehension beyond language: common brain networks activated by a movie and its script. *PLoS One* 13:e0200134. doi: 10.1371/journal.pone.0200134

Tikka, P., Väljamäe, A., de Borst, A. W., Pugliese, R., Ravaja, N., Kaipainen, M., et al. (2012). Enactive cinema paves way for understanding complex real-time social interaction in neuroimaging experiments. *Front. Hum. Neurosci.* 6:298. doi: 10.3389/fnhum.2012.00298

Twenge, J. M., Martin, G. N., and Spitzberg, B. H. (2019). Trends in U.S. adolescents' media use, 1976–2016: the rise of digital media, the decline of TV, and the (near) demise of print. *Psychol. Pop. Media Cult.* 8, 329–345. doi: 10.1037/ppm0000203

Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., and Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychol. Sci.* 20, 1092–1099. doi: 10.1111/j.1467-9280.2009.02411.x

Vanderwal, T., Eilbott, J., and Castellanos, F. X. (2019). Movies in the magnet: naturalistic paradigms in developmental functional neuroimaging. *Dev. Cogn. Neurosci.* 36:100600. doi: 10.1016/j.dcn.2018.10.004

Vicol, P., Tapaswi, M., Castrejon, L., and Fidler, S. (2018). "MovieGraphs: towards understanding human-centric situations from videos." in *2018 IEEE/CVF conference on computer vision and pattern recognition. 2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR), Salt Lake City, UT*.

Vorderer, P. (1996). *Fernsehen als "Beziehungskiste"*. Parasoziale Interaktionen und Beziehungen mit Medienfiguren. Opladen.

Vorderer, P., Wulff, H. J., and Friedrichsen, M. , (2006). *Suspense: Conceptualizations, theoretical analyses, and empirical explorations*, New: York Routledge.

Weber, R. (2013). Brain, mind, and media: neuroscience meets media psychology. *J. Media Psychol.* 25:202. doi: 10.1027/1864-1105/a000162

Weber, R., Eden, A., Huskey, R., Mangus, J. M., and Falk, E. B. (2015b). Bridging media psychology and cognitive neuroscience: challenges and opportunities. *J. Media Psychol.* 27, 146–156. doi: 10.1027/1864-1105/a000163

Weber, R., Huskey, R., Mangus, J. M., Westcott-Baker, A., and Turner, B. O. (2015a). Neural predictors of message effectiveness during counterarguing in antidrug campaigns. *Commun. Monogr.* 82, 4–30. doi: 10.1080/03637751.2014.971414

Weber, R., Mangus, J. M., Huskey, R., Hopp, F. R., Amir, O., Swanson, R., et al. (2018). Extracting latent moral information from text barratives: relevance, challenges, and solutions. *Commun. Methods Meas.* 12, 119–139. doi: 10.1080/19312458.2018.1447656

Weber, R., Ritterfeld, U., and Mathiak, K. (2006). Does playing violent video games induce aggression? Empirical evidence of a functional magnetic resonance imaging study. *Media Psychol.* 8, 39–60. doi: 10.1207/S1532785XMEP0801_4

Weber, R., Sherry, J., and Mathiak, K. (2008). "The neurophysiological perspective in mass communication research" in *Beatty et al. biological dimensions of communication: Perspectives, methods, and research*, eds. M. J. Beatty, J. C. McCroskey, and K. Floyd, Cresskill, NJ: Hampton Press. 41–71.

Wen, H., Shi, J., Chen, W., and Liu, Z. (2017). Transferring and generalizing deep-learning-based neural encoding models across subjects. *bioRxiv*:171017. doi: 10.1101/171017

Wilensky, R. (1983). Story grammars versus story points. *Behav. Brain Sci.* 6, 579–591.

Willems, R. M., Nastase, S. A., and Milivojevic, B. (2020). Narratives for neuroscience. *Trends Neurosci.* 43, 271–273. doi: 10.1016/j.tins.2020.03.003

Yeshurun, Y., Nguyen, M., and Hasson, U. (2017). Amplification of local changes along the timescale processing hierarchy. *Proc. Natl. Acad. Sci.* 114, 9475–9480. doi: 10.1073/pnas.1701652114

Young, C. B., Raz, G., Everaerd, D., Beckmann, C. F., Tendolkar, I., Hendler, T., et al. (2017). Dynamic shifts in large-scale brain network balance as a function of arousal. *J. Neurosci.* 37, 281–290. doi: 10.1523/JNEUROSCI.1759-16.2016

Zellers, R., Lu, X., Hessel, J., Yu, Y., Park, J. S., Cao, J., et al. (2021). Merlot: multimodal neural script knowledge models. *Adv. Neural Inf. Proces. Syst.* 34, 23634–23651. doi: 10.48550/arXiv.2106.02636

Zillmann, D., and Cantor, J. R. (1972). Directionality of transitory dominance as a communication variable affecting humor appreciation. *J. Pers. Soc. Psychol.* 24, 191–198. doi: 10.1037/h0033384

Zillmann, D., and Vorderer, P. (2000). *Media entertainment: The psychology of its appeal*, New: York Routledge.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 12 May 2023
DOI 10.3389/fnins.2023.1156890

OPEN ACCESS

EDITED BY
Allen Ming Yan Cheong,
Hong Kong Polytechnic University,
Hong Kong SAR, China

REVIEWED BY
Qing He,
Peking University, China
Bin Zhang,
Nova Southeastern University, United States

*CORRESPONDENCE
Minpeng Xu
✉ xmp52637@tju.edu.cn
Dong Ming
✉ richardming@tju.edu.cn

RECEIVED 07 February 2023
ACCEPTED 25 April 2023
PUBLISHED 12 May 2023

CITATION
Li R, Xu M, You J, Zhou X, Meng J, Xiao X, Jung T-P and Ming D (2023) Modulation of rhythmic visual stimulation on left–right attentional asymmetry.
Front. Neurosci. 17:1156890.
doi: 10.3389/fnins.2023.1156890

Modulation of rhythmic visual stimulation on left–right attentional asymmetry

Rong Li[1], Minpeng Xu[1,2]*, Jia You[1], Xiaoyu Zhou[1], Jiayuan Meng[1,2], Xiaolin Xiao[1,2], Tzyy-Ping Jung[1,2,3] and Dong Ming[1,2]*

[1]School of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China, [2]Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China, [3]Swartz Center for Computational Neuroscience, University of California San Diego, San Diego, CA, United States

The rhythmic visual stimulation (RVS)-induced oscillatory brain responses, namely steady-state visual evoked potentials (SSVEPs), have been widely used as a biomarker in studies of neural processing based on the assumption that they would not affect cognition. However, recent studies have suggested that the generation of SSVEPs might be attributed to neural entrainment and thus could impact brain functions. But their neural and behavioral effects are yet to be explored. No study has reported the SSVEP influence on functional cerebral asymmetry (FCA). We propose a novel lateralized visual discrimination paradigm to test the SSVEP effects on visuospatial selective attention by FCA analyses. Thirty-eight participants covertly shifted their attention to a target triangle appearing in either the lower-left or -right visual field (LVF or RVF), and judged its orientation. Meanwhile, participants were exposed to a series of task-independent RVSs at different frequencies, including 0 (no RVS), 10, 15, and 40-Hz. As a result, it showed that target discrimination accuracy and reaction time (RT) varied significantly across RVS frequency. Furthermore, attentional asymmetries differed for the 40-Hz condition relative to the 10-Hz condition as indexed by enhanced RT bias to the right visual field, and larger Pd EEG component for attentional suppression. Our results demonstrated that RVSs had frequency-specific effects on left–right attentional asymmetries in both behavior and neural activities. These findings provided new insights into the functional role of SSVEP on FCAs.

KEYWORDS

rhythmic visual stimulation (RVS), steady-state visual evoked potentials (SSVEPs), selective attention, attentional suppression, functional cerebral asymmetry (FCA)

1. Introduction

When the brain is exposed to a series of rhythmic visual stimuli (RVS, or flicker), it produces oscillatory neural responses which share identical or harmonically-related frequencies with RVS, namely steady-state visual evoked potentials (SSVEPs) (Herrmann, 2001; Srinivasan et al., 2006; Moratti et al., 2007). In comparison with broadband electroencephalography (EEG) signals, SSVEP can provide more stable amplitude, frequency, and phase properties of brain activities in a controlled way. Therefore, it has been widely used in neuroscience research and brain-computer interfaces (BCIs) as a frequency-tagged biomarker that would not introduce functional changes. For example, SSVEPs at different frequencies have been quantified to track the temporal dynamics of attentional processes or probe the neural basis of sensory processing (Mueller et al., 2006; Gregori Grgič et al., 2016; Davidson et al., 2020; Stefanac et al., 2021; Kritzman et al., 2022).

However, the frequency-tagging approach has been gradually questioned for the functional effects of RVS-driven neural entrainment on cognitive processes. Neural entrainment is the process whereby intrinsic neural oscillations synchronize with external stimulus rhythms (Thut et al., 2011; Notbohm et al., 2016; Hanslmayr et al., 2019; Beliaeva and Polania, 2020; Cucu et al., 2022). Previous studies have demonstrated that intrinsic neural oscillations are functionally relevant to cognitive processes, which may serve as a fundamental mechanism to support neural processing but not a meaningless byproduct when the brain works (Buzsaki and Draguhn, 2004; Bueno-Junior et al., 2017; Wstmann et al., 2019; Zhang et al., 2019; Riddle et al., 2020; He et al., 2022). Therefore, neural entrainment to rhythmic inputs, such as electrical, magnetic, or sensory stimulation, has been proposed as a promising mechanism to modulate cognitive functions supported by distinct oscillatory patterns. RVS enables the alteration of oscillatory rhythmic activities, thus showing the potential to causally affect neural processing. For example, it was found that 10-Hz visual flicker could entrain endogenous alpha-band neural oscillations, which could predict periodic behavior modulation in visual perception (Spaak et al., 2014) or impaired detection performance in selective visuospatial attention (Gulbinaite et al., 2017). These findings indicated that the generation of SSVEPs might be accompanied by behavioral changes in cognitive processes. In other words, SSVEPs might be more than simple frequency-following responses but have functional roles in cognitive processes. Yet despite there having been many studies of RVS-driven neural entrainment recently, the understanding of SSVEP effects on cognitive processes remains incomplete.

Functional cerebral asymmetry (FCA), a ubiquitous feature of cerebral organization, has been widely documented in cognitive processes, especially for visuospatial attention (Stevens et al., 2012; Chen et al., 2018). Previous studies have found a diversity of left versus right visual processing asymmetries at the neuroanatomical and functional levels. For example, there are asymmetrical EEG activities in neural networks for attentional processing (Kertesz et al., 1986; Jones and Sliva, 2020; Mulligan et al., 2022) and visual field asymmetries of attention (Brederoo et al., 2019) or perception performance (Carrasco et al., 2022). From the perspective of brain development and evolution, FCA's emergence increases neural capacity and confers superior brain efficiency of neural processing (Corballis, 2009). Furthermore, the degree or direction of FCA shows a close relationship with the performance of visuospatial attention (Wang et al., 2016) and can be affected by many factors, such as normal aging (Hong et al., 2015), training experience (Deng and Rogers, 2002; Rogers, 2006), and experimental stimulation (Shalev et al., 2018; Chiandetti and Vallortigara, 2019). In these views, FCA provides a functional indicator to evaluate changes in cognitive functions.

Here, we aimed to test the functional effects of SSVEP from the perspective of FCAs in visuospatial attention. Previous studies have demonstrated that the time course of FCAs can be reflected by SSVEPs (Martens and Hübner, 2013). By virtue of stable spectrum and high signal-to-noise ratio, the frequency-tagged SSVEP is able to indicate the asymmetrical allocation of visual attention (Zhang et al., 2022). However, to date, the functional effects of SSVEPs on FCAs have been rarely noticed, which limits our understanding of the mechanism underpinning attentional asymmetry. To this end, we devised a novel EEG paradigm in which participants were required to perform a lateralized visual discrimination task while exposed to a series of task-independent RVS backgrounds. FCA analyses were performed on both behavior and EEG responses in target discrimination. Notably, the applied RVS background was evenly distributed in the left and right visual fields (LVF and RVF) and would theoretically induce SSVEP responses in bilateral cerebral hemispheres with scalp EEG recording (Herrmann, 2001; Zhang et al., 2022). Therefore, the extraction of left–right asymmetrical EEG could attenuate the interference of SSVEP on the observation of subtle activities related to lateralized attentional processes, which would provide a novel and concise way to reveal the neural effects of SSVEPs.

2. Materials and methods

2.1. Participants

Thirty-eight (16 males; 23 ~ 27 years old) healthy right-handed adults were recruited for the experiment, which was approved by the Institutional Review Board at Tianjin University. All had a normal or corrected-to-normal vision, and they gave written informed consent before the experiment.

2.2. Experimental paradigms and procedure

Figure 1 shows the time course of the experimental paradigm. In the beginning, a concentric fixation cross would be displayed in the screen centre, with a red (RGB: 255, 0, 0) or green (RGB: 0, 255, 0) dot on it (size: 0.56° × 0.56°), cueing the target color of each session (red or green cue with equal probability). Participants would end the cueing period by pressing the "space" key on the keyboard once they were ready to start the task. Then, each trial would begin with an RVS flickering period lasting for 2,900 ~ 3,200 ms randomly. In this period, RVS was displayed in a concentric square area in the screen (the square edge to the screen centre is 5° apart). Two small black squares (RGB: 0, 0, 0) were embedded in the bottom-left and right RVS flickering area (the square centre to the screen centre is 3.9° apart, dip angle is 45°). During the last 800 ms of the RVS flickering period, a visual search array consisting of a distractor and a target triangle (size: 0.56° × 0.56°) would appear in the centre of the two black squares, respectively. These two triangles were red and green. The target triangle had an upward or downward orientation in the left or right visual field (LVF or RVF), while the distracting triangle had an opposite orientation in the opposite visual field. Participants were required to gaze at the concentric fixation cross and covertly allocate attention to the lateralized target triangle to discriminate its orientation as quickly as possible by pressing the "upward" or "downward" button with the index or middle finger of their dominant hands. In each trial, RVS background would flash between white (RGB: 255, 255, 255) and black (RGB: 0, 0, 0) with a frequency selected from 0, 10, 15, and 40 Hz in a random ergodic sequence. For 0-Hz (the control condition), the RVS square would always remain white (RGB: 255, 255, 255). To eliminate the potential effect of RVS phase on target

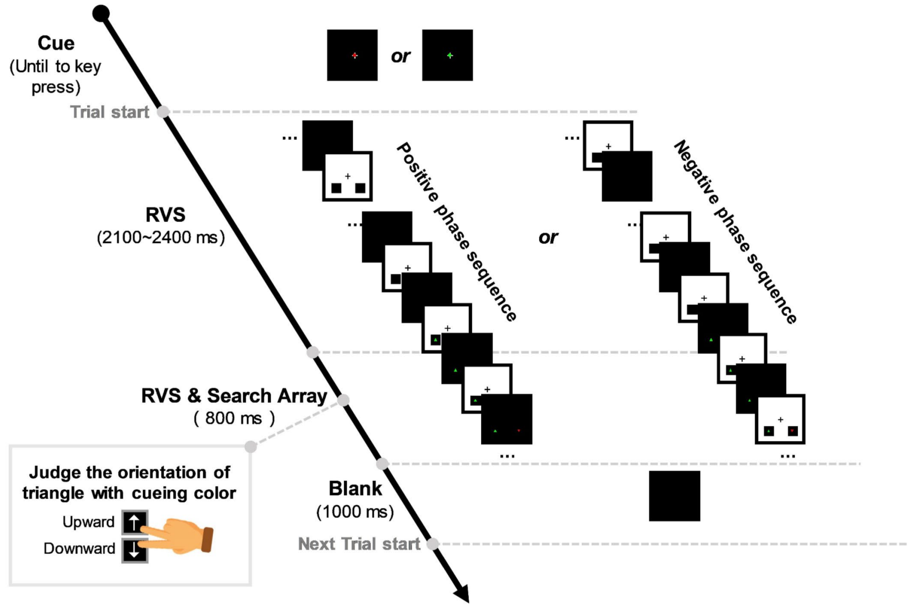


FIGURE 1
Experimental stimulation sequence. A red/green dot would appear in the centre of the black task background, cueing the target color of each session. Then, a single trial would begin with rhythmic visual stimulation (RVS) flickering in the concentric square area and lasting for 2,900~3,200ms randomly. During the last 800ms of the RVS flickering period, a visual search array consisting of a distractor and a target triangle would appear in the bottom-left and right black squares, respectively. Participants were required to judge the upward/downward orientation of the target triangle with cueing color as accurately and quickly as possible. In each trial, RVS had a frequency selected from 0, 10, 15, and 40Hz in a random ergodic sequence. To eliminate the effect of RVS phase on target discrimination, we designed positive and negative phases for the 10-Hz, 15-Hz, and 40-Hz RVS sequences. The positive RVS phase indicates that the visual search array appears when the RVS background flashes from black to white, while the negative phase indicates the opposite. There was a 1,000-ms intertrial interval for participants to blink and relax.

discrimination, we designed two initial phases, namely positive and negative phases, for the 10-Hz, 15-Hz, and 40-Hz RVS sequences. The positive RVS phase indicates that the visual search array appears when the RVS background flashes from black to white, while the negative phase indicates the opposite. Between trials, there would be a 1,000-ms rest period for participants to blink and relax.

The stimuli paradigm was designed with Psych-toolbox 3.0 package in MATLAB software. Participants sat 75 cm from a 27-inch monitor (spatial resolution: 1,920 × 1,080 pixels, refresh rate: 120 Hz) with a black background (mean luminance: 97.5 cd/m^2). Each participant completed 24 sessions of the lateralized visual discrimination task in the experiment. There were 28 trials in each session. As a result, we collected 672 trials across 2 visual field × 4 RVS frequency conditions from each participant. All participants' key-pressing behaviors, such as manual reaction time (RT) and discrimination accuracy, and their EEG signals were recorded for the following within-subject statistical analyses.

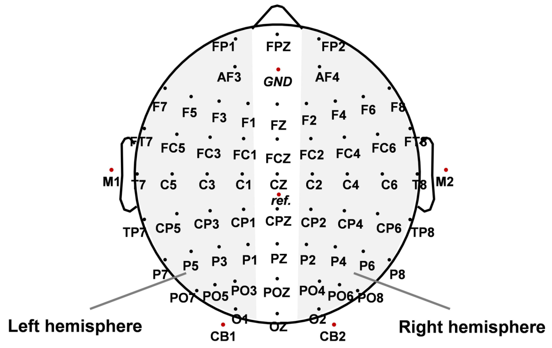


FIGURE 2
The spatial distribution of scalp electrode sites. *GND* indicates the location of the ground electrode and *ref.* indicates the location of the reference electrode. The paired electrodes in the gray area were used to analyze event-related lateralizations (ERLs).

2.3. Data acquisition and trials selection

The EEG data were acquired by a Neuroscan Synamps2 system with a 68-channel cap whose electrodes were placed according to the International 10/20 System (see Figure 2). All recording electrodes were grounded to the middle electrode site between FPZ and FZ. The reference electrode was placed between CZ and CPZ. Eye movements and blinks were monitored by a pair of horizontal electrooculogram (HEOG) channels placed at the outer canthi of both eyes and a pair of vertical electrooculogram (VEOG) channels above and below the left eye. The EEG signal was recorded with a bandpass filter ranging from 0.01 to 100 Hz and a notch filter centred at 50 Hz. The sampling rate of the EEG was set at 1,000 Hz.

The recorded EEG data were offline segmented into trials ranging from −500 ms to 1,500 ms after target-triangle onset. If a trial's response was wrong or RT was outside the time range of 200–800 ms, it would be excluded from the following analyses. Then, the remains of trials with the top 10% largest HEOG fluctuations would be identified and removed to minimize the impact of eye blinks and eye movements on EEG data. As a result, there were 572.9 ± 5.3 trials available for each participant.

2.4. Analyses of behavioral and EEG asymmetries

The behavioral measurements of the proposed task included RT and discrimination accuracy. We further calculated a laterality index of these measurements, allowing us to quantify behavioral asymmetry. The laterality indices for RT (LI_{RT}) and discrimination accuracy (LI_{acc}) were extracted according to the formula (Resnick et al., 1994; Beaton et al., 2007):

$$\mathrm{LI} = \left(L - R\right) / \left(L + R\right) \times 100. \tag{1}$$

where *L* represents the grand-averaged behavioral performance in the LVF condition and R in the RVF condition. The LI values ranged from −1 to 1. Taken LI_{RT} as an example, a positive LI value indicated an RVF advantage of response speed in the lateralized visual discrimination task, and negative LI indicated a LVF advantage. The higher the absolute LI value, the stronger the behavioral asymmetry.

For EEG data, each trial was first filtered by a third-order Butterworth bandpass filter ranging from 1 to 12 Hz. Then baseline correction was performed within the 100 ms time window before target onset. We extracted event-related lateralizations (ERLs) averaged from three symmetrical electrode pairs (P5/P6, PO5/PO6, O1/O2) to analyze hemispheric differences in EEG activities in selective attention. The ERL was calculated as the contralateral-minus-ipsilateral difference of EEG activities recorded in paired electrodes in the left and right hemispheres (LH and RH, see Figure 2) (Oostenveld et al., 2003):

$$\mathrm{ERL} = \begin{pmatrix} \mathrm{EEG}_{\mathrm{RH,LVF}} + \mathrm{EEG}_{\mathrm{LH,RVF}} \\ -\mathrm{EEG}_{\mathrm{RH,RVF}} - \mathrm{EEG}_{\mathrm{LH,LVF}} \end{pmatrix} / 2 \tag{2}$$

To some degree, the calculation process of ERL waveform suppresses symmetric EEG activity, which is identical to bilateral cerebral hemispheres and visual field conditions. We focused on two ERL components, i.e., N2 posterior contralateral (N2pc) (Luck and Hillyard, 1994; Fuggetta and Duke, 2017) and distractor positivity (Pd) components (Hickey et al., 2009; Suarez-Suarez et al., 2019). Amplitude and latency analyses of N2pc were limited to a time window of 210–270 ms after target onset, while Pd was limited to 300–360 ms. The amplitude of each ERL component was calculated as the mean amplitude within the respective time window. The latency was measured as the time point before which 50% of the total component area was observed in the respective time window.

2.5. Statistical analyses

A mixed 2 × 4 analysis of variance (ANOVA) was used to test RT and discrimination accuracy in behavioral performance, with visual field (LVF and RVF) and RVS frequency (control, 10-Hz, 15-Hz, and 40-Hz) as within-subject factors. Then, a one-way ANOVA with RVS frequency (control, 10-Hz, 15-Hz, and 40-Hz) as a within-subject factor was used to test RVS modulation on behavioral asymmetries as measured by LI_{RT} and LI_{acc}. The ERL amplitude and latency were separately submitted to a one-way ANOVA with RVS frequency (control, 10-Hz, 15-Hz, and 40-Hz) as a within-subject factor to examine the RVS modulation on EEG asymmetries. The above statistical analyses were subjected to Bonferroni-corrected *post hoc* tests with a significance level of $p < 0.05$.

3. Behavioral results

3.1. Discrimination accuracy

Figure 3A shows that the grand-averaged discrimination accuracy across all participants was 83.92% ± 0.84% (all results are presented as the mean ± standard error). Statistical analysis indicated that participants had similar accuracy between the LVF and RVF conditions (LVF: 83.60% ± 0.91%; RVF: 84.24% ± 0.89%; Visual field: $F\,(1, 37) = 1.108$, $p = 0.299$, $\eta^2 = 0.029$). However, it varied significantly with RVS frequency ($F\,(3, 111) = 5.793$, $p = 0.001$, $\eta^2 = 0.135$). Specifically, participants performed better in the 15-Hz and 40-Hz RVS than in the control condition (control: 82.36% ± 0.96% vs. 15-Hz: 85.17% ± 0.91% and 40-Hz: 84.52% ± 0.96%; both $p < 0.05$ after Bonferroni correction). There was no interaction of visual field × RVS frequency ($F\,(3, 111) = 0.634$, $p = 0.595$, $\eta^2 = 0.017$).

3.2. Reaction time (RT)

Figure 3B shows the grand-averaged RTs for all conditions. RVF's grand-averaged RT was found to be approximately 20 ms faster than LVF's (LVF: 569.03 ± 7.09 ms, RVF: 545.81 ± 8.18 ms). Statistical analyses indicated that the advantage of response speed for RVF was significant in the lateralized visual discrimination task (Visual field: F (1, 37) = 39.464, $p < 0.001$, $\eta^2 = 0.516$). Furthermore, RVS frequency had a significant effect on the grand-averaged RTs collapsed across visual fields (RVS frequency: $F\,(3, 111) = 23.376$, $p < 0.001$, $\eta^2 = 0.387$). Specifically, participants responded with a latency of 551.52 ± 7.14 ms for 40-Hz RVS, which was significantly faster than 555.85 ± 7.38 ms for control, 563.84 ± 7.79 ms for 10-Hz RVS, and 558.48 ± 7.59 ms for 15-Hz RVS (all $p < 0.05$ for paired comparisons after Bonferroni correction). Besides, 10-Hz RVS led to significantly slower responses than the control and 15-Hz (both $p < 0.01$ for paired comparisons after Bonferroni correction). Notably, there was a significant interaction of visual field X RVS frequency ($F\,(3, 111) = 5.437$, $p = 0.002$, $\eta^2 = 0.128$). For LVF, only 10-Hz RVS showed a significant modulation effect on RT (10-Hz vs. control and 40-Hz: both $p < 0.05$ for paired comparisons after Bonferroni correction), whereas for RVF, RTs were significantly modulated in both 10-Hz and 40-Hz RVS conditions (10-Hz vs. others, 40-Hz vs. others: all $p < 0.01$ for paired comparisons after Bonferroni correction).

3.3. Analyses of behavioral asymmetries

Figures 3C,D show the behavioral asymmetries as evaluated by LI_{acc} and LI_{RT}, respectively. As shown in Figure 3C, the grand-averaged LI_{acc} values were −0.92 ± 0.94, −0.66 ± 0.53, 0.08 ± 0.59 and

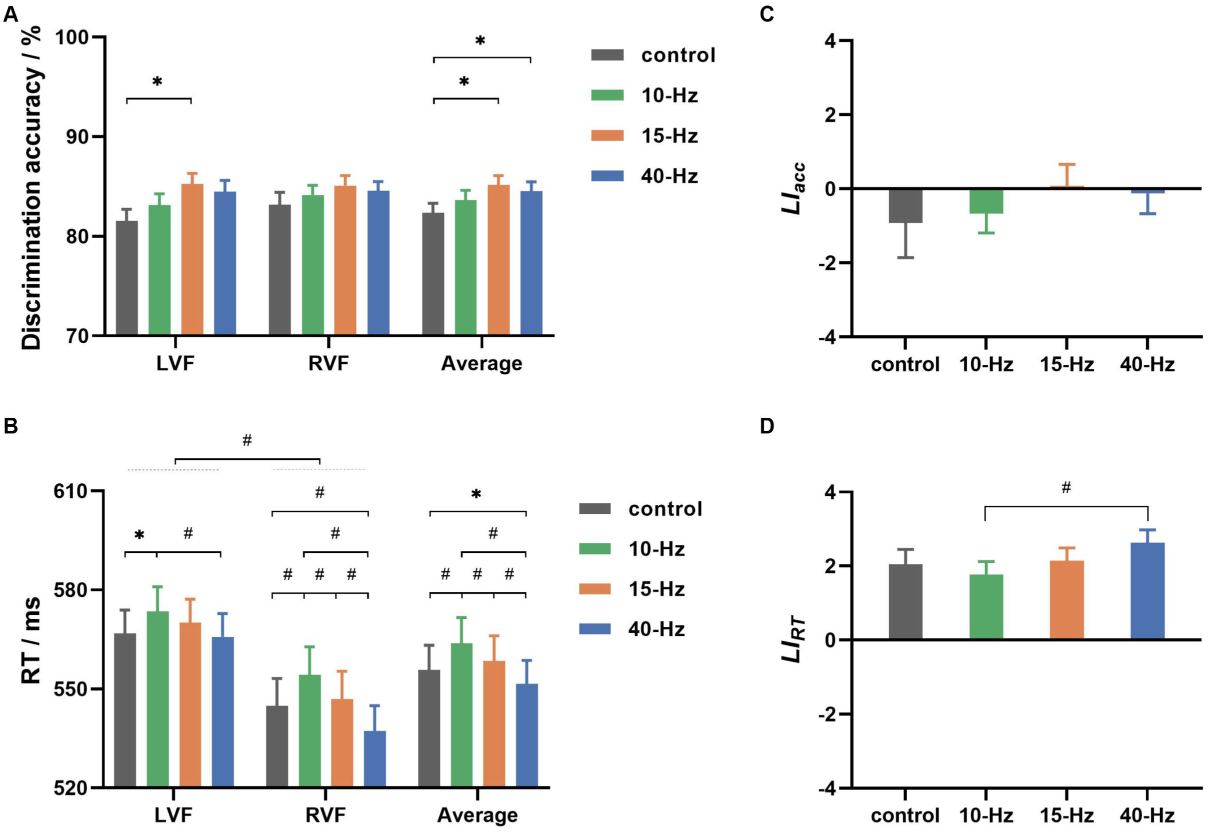


FIGURE 3
(A) The grand-averaged discrimination accuracies for left and right visual fields (LVF and RVF) in 4 RVS frequency conditions (control, 10, 15, and 40Hz). **(B)** The grand-averaged reaction time (RT) for 2 visual fields in 4 RVS frequency conditions. **(C)** The grand-averaged LI_{acc} values for 4 RVS frequency conditions. **(D)** The grand-averaged LI_{RT} values for 4 RVS frequency conditions. Error bars represent ± standard error; * indicates p<0.05, and # indicates p<0.01 for paired comparisons after Bonferroni correction.

−0.13 ± 0.54 for control, 10-Hz, 15-Hz, and 40-Hz RVS, respectively. There was no significant difference between them according to statistical analysis (RVS frequency: F (3, 111) = 0.523, $p = 0.667$, $\eta^2 = 0.014$). $\mathrm{LI_{RT}}$ had values of 2.04 ± 0.41, 1.77 ± 0.35, 2.14 ± 0.35, and 2.63 ± 0.37 for control, 10-Hz, 15-Hz, and 40-Hz, respectively, as shown in Figure 3D. These positive $\mathrm{LI_{RT}}$ values indicated that RVF had a group-level advantage in response speed. Statistical analyses revealed a significant RVS modulation effect on $\mathrm{LI_{RT}}$ (RVS frequency: F (3, 111) = 5.937, $p = 0.001$, $\eta^2 = 0.138$). Specifically, the grand-averaged $\mathrm{LI_{RT}}$ was significantly higher for 40-Hz than the other conditions before Bonferroni correction but only higher than the 10-Hz condition after Bonferroni correction.

4. ERL results

4.1. Correlation between ERL components and behaviors

Figure 4A shows the ERL waveforms. Two prominent components, i.e., N2pc and Pd, could be found in the grand-averaged ERL waveform marked red. To explore the relationship between ERL and behavioral asymmetry, regression analyses were performed between the two components and $\mathrm{LI_{RT}}$, as shown in Figures 4B,C. As a result, the Pd amplitude was found to have a significantly positive correlation with $\mathrm{LI_{RT}}$ ($R^2 = 0.036$, $p < 0.05$, see Figure 4B), indicating that a larger Pd corresponded to a larger RT bias toward the RVF. Besides, there was a significantly negative correlation between N2pc latency and $\mathrm{LI_{RT}}$ ($R^2 = 0.028$, $p < 0.05$, see Figure 4C), implying that an earlier N2pc had a larger LI_{RT}. N2pc amplitude and Pd latency had no significant correlations with $\mathrm{LI_{RT}}$ (both $p > 0.05$).

4.2. RVS modulation on ERL components

Figure 5A shows the grand-averaged ERL waveforms across all participants and the topographic maps of N2pc and Pd for each RVS condition. As can be seen, the Pd component for 40-Hz appeared to be larger than the others (see Figure 5B). Statistical analyses showed that RVS had a significant main effect on the Pd amplitude (RVS frequency: F (3, 111) = 2.282, $p = 0.042$, $\eta^2 = 0.071$). Moreover, such RVS modulation effects on the Pd amplitude were consistent with that on $\mathrm{LI_{RT}}$, i.e., 40-Hz had a significantly larger Pd amplitude than 10-Hz after Bonferroni correction ($p < 0.05$). No significant RVS modulation effects were found on the N2pc component.

Using Short-Time Fourier Transform (STFT), we analyzed the target-induced changes in inter-trial coherence (ITC) and event-related spectral perturbation (ERSP) of EEG trials filtered by a

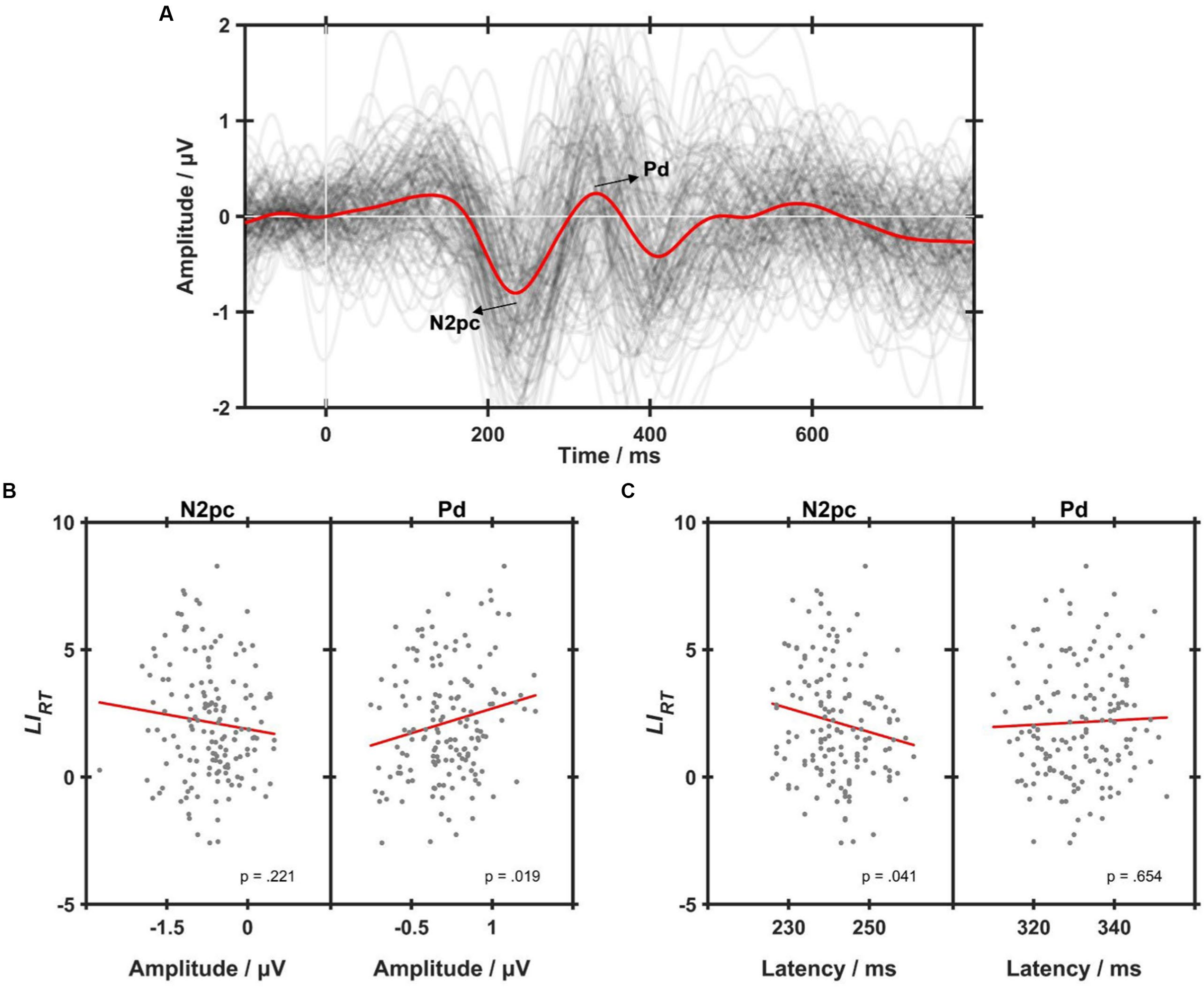


FIGURE 4
(A) The grand-averaged event-related lateralizations (ERLs) across all conditions (the red line). Each dark gray line represents an individual ERL waveform averaged within each condition, i.e., control, 10-Hz, 15-Hz, and 40-Hz. **(B)** Simple lineal regression analyses on the correlations between LI_{RT} values and N2pc (left) and Pd (right) amplitudes. **(C)** Simple lineal regression analyses on the correlations between LI_{RT} values and N2pc (left) and Pd (right) latencies. Each gray dot represents an individual of one RVS condition, and the red line represents the linear correlation fitting curve.

third-order Butterworth bandpass filter ranging from 1 to 95 Hz. The ITC measure has been widely estimated as the time-frequency representation of phase-locking strength to the time-locking events of EEG signals (Makeig et al., 2004). To determine whether RVSs could induce frequency-tagged SSVEP response at the corresponding frequency, we presented the two-dimensional (2D) images of ITC measure extracted from the Oz channel in Figure 6. As a result, it was found that 10, 15 and 40-Hz RVS could all exhibit apparent SSVEP responses at the stimulation frequency and its harmonics. Furthermore, consistent with the previous study (Herrmann, 2001), the ITC strength of SSVEP responses showed a decreasing trend with the frequency increased. ERSP measure reflects event-related changes in spectral power (in dB) of EEG data relative to baseline. To investigate RVS modulation on the asymmetrical pattern of EEG power, we extracted the ERL measure of ERSP baseline-corrected with the mean of 100-ms EEG activities before target onset (details in Supplementary material). A one-way ANOVA with RVS frequency as a within-subject factor was performed on the ERL measure of ERSP averaged from three symmetrical electrode pairs (P5/P6, PO5/PO6, O1/O2) and with a time-frequency window of interest (a 150–400 ms window at 1–12 Hz). As a result, no significant RVS modulation effects were found on the left–right asymmetry degree of ERSP measures.

5. Discussion

In the current study, we aimed to investigate the functional effects of SSVEP on visuospatial selective attention. To this end, we designed a lateralized visual discrimination task with RVS background and analyzed the left–right attentional asymmetries in behavior and EEG data to evaluate potential functional changes with RVS modulation. Our results indicated that RVSs could shift the fundamental RVF advantage of response speed with frequency-dependent effects. Furthermore, such behavioral changes were in accordance with EEG variations, i.e., the Pd measure of attentional suppression became significantly more prominent for the 40-Hz than the 10-Hz RVS condition, reflecting RVS-elicited enhancement in the suppression of distractor during visual search. These findings indicate that 40-Hz

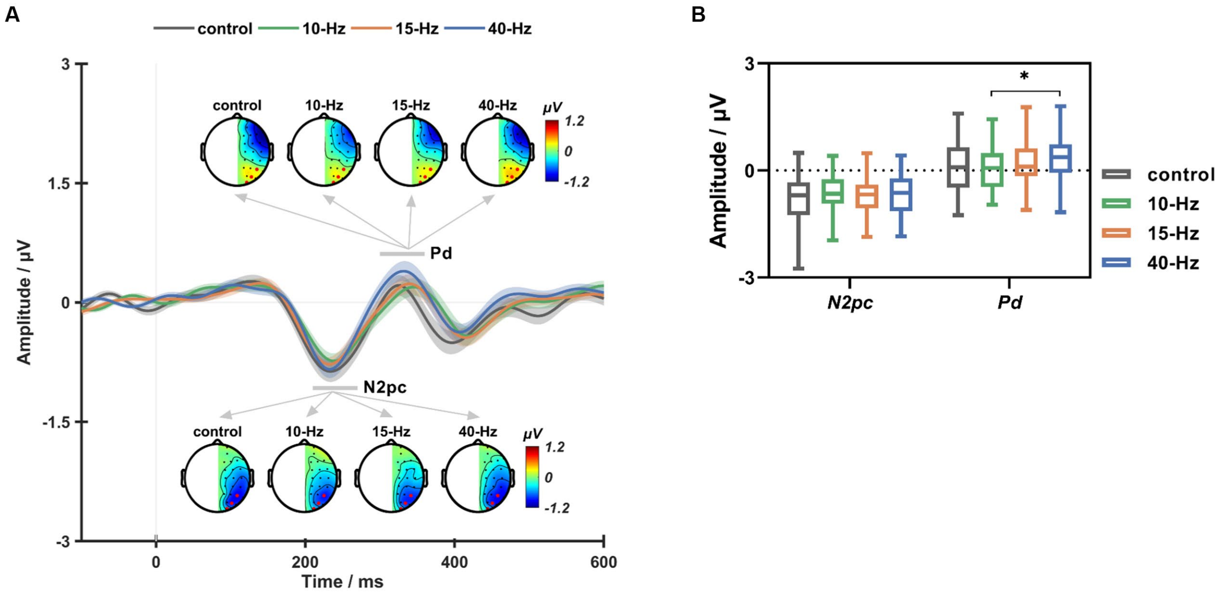


FIGURE 5
(A) The grand-averaged ERL waveforms and topographic maps of N2pc and Pd for the control, 10-Hz, 15-Hz, and 40-Hz conditions, respectively. **(B)** The box plots of ERL amplitude for the control, 10-Hz, 15-Hz, and 40-Hz RVS conditions. Error bars represent ± standard error; * indicates $p<0.05$ for paired comparisons after Bonferroni correction.

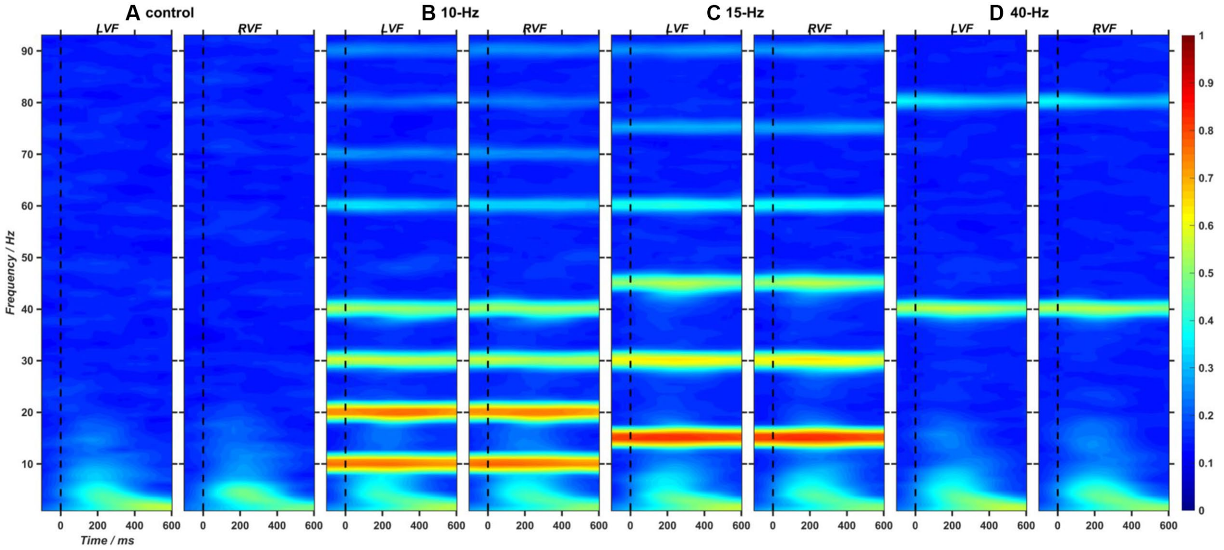


FIGURE 6
(A–D) The 2D images of inter-trial coherence (ITC) extracted from EEG trials at Oz channel and in the left and right visual fields (LVF and RVF) for the control, 10-Hz, 15-Hz, and 40-Hz conditions, respectively.

RVS can entrain neural oscillations relevant to attentional suppression and bring behavioral consequences.

5.1. Functional cerebral asymmetries in visuospatial selective attention

Both humans and animals exhibit hemispheric asymmetry, which has been documented for many cognitive processes, including visual foraging (Güntürkün et al., 2000), global–local shape processing (Gerlach and Poirel, 2020), and visual-word recognition (Van der Haegen and Brysbaert, 2011). The hemispheric asymmetry contributes a lot to cognitive performance from brain development and evolution (Corballis, 2009). The functional differences between hemispheres in selective attention have been shown to have a lateralized effect on behavioral outcomes (Van der Haegen and Brysbaert, 2011). Moreover, evidence from neuropathological studies indicates that the degree of behavioral asymmetry is closely linked to attentional ability

(Poynter et al., 2010). In this study, we found a rightward bias of behavior in the proposed selective attention paradigm. Specifically, the participants had a significant advantage of RVF over LVF in terms of response speed. Our findings align with previous studies on attentional bias, which have shown an RVF advantage in the lateralized recognition task (Vergilino Perez et al., 2012; Bergerbest et al., 2017).

The current task required participants to covertly orient their attention to the target, which appeared in pairs with a distractor. Thus, two distinct attentional processes were theoretically involved in shaping their behaviors, namely attentional selection and distractor suppression (Mazza et al., 2009). Indeed, we observed two relevant ERL components in the posterior brain region, i.e., N2pc and Pd. In light of previous studies, the current N2pc could be explained as a covert deployment of visual attention to the lateralized target during visual search, whereas the presence of Pd indicated an inhibitory process that prevented attention from being oriented to distracting items (Sawaki et al., 2012). The subsequent regression analyses revealed significant correlations between the posterior asymmetry patterns of EEG activities and the behavioral bias, i.e., both the N2pc latency and Pd amplitude were significantly correlated to $\mathrm{LI}_{\mathrm{RT}}$. These findings suggest that the N2pc and Pd components are effective in measuring the attentional bias, which is consistent with previous studies (Salahub and Emrich, 2020; Kappenman et al., 2021).

5.2. RVS modulation on left–right attentional asymmetries

An interesting and novel finding in this study was the shift of behavioral and EEG asymmetries in selective attention under different RVS modulations. This study used $\mathrm{LI}_{\mathrm{RT}}$ to assess behavioral asymmetry, which showed significant differences among RVS frequencies. Compared to the control condition without RVS, 10-Hz RVS had a slightly lower $\mathrm{LI}_{\mathrm{RT}}$, but 40-Hz RVS led to a higher $\mathrm{LI}_{\mathrm{RT}}$. As a result, the LI_{RT} differences between the 10-Hz and 40-Hz RVS conditions reached a significant level. Specifically, 10-Hz RVS significantly reduced the participant's response speed for both RVF and LVF, indicating that the alpha-band flicker could impair the task performance in selective attention (Gulbinaite et al., 2017). Whereas 40-Hz RVS significantly sped up the participant's response only for RVF but not for LVF. The effect of 40-Hz RVS was consistent with previous findings that the gamma-band flicker could facilitate target detection and discrimination (Bauer et al., 2009). Furthermore, it added new insights that 40-Hz RVS could enhance the rightward asymmetry of attention behavior. Consistently, RVS showed frequency-specific modulations on ERL components. Specifically, compared to the 10-Hz condition, 40-Hz RVS induced a significantly larger Pd, indicating increased EEG asymmetry. However, no significant differences were found for N2pc. Considering the functional relevance of these two ERL components, this study indicates that the RVS could influence attentional suppression but not attentional selection. We suggest that RVS might influence behavioral outputs by modulating the suppression process in selective attention.

Our observations, as discussed above, suggest that the applied RVS showed functional modulation on selective attention with behavioral consequences. However, it may be questioned from the following aspects. Firstly, it might be argued that the observed frequency-dependent behavioral effects are not due to the RVS-induced functional modulation but instead reflect the interference of unknown fatigue effects due to prolonged exposure to rhythmic visual input (Dreyer et al., 2017). Indeed, compared with the control condition, the applied RVSs at 10, 15, and 40-Hz can all impact task performance, with a promoting or inhibiting effect on RT or discrimination accuracy. Yet, the applied RVS was evenly distributed in the LVF and RVF. The resulting fatigue effect is expected not to interfere with evaluating behavioral asymmetry. Thus, we suggest that the degree of behavioral asymmetry provides a feasible option for assessing the behavioral effects of RVS modulation. Secondly, because the mental task in this study involved a key-pressing behavior, it is possible that motor-related EEG activity, which overlapped with Pd in the same time window, contributed to the phenomenon of posterior EEG asymmetry. To address this question, we computed ERL on the motor area to investigate the potential influences. Statistical analyses showed no significant differences in the motor-related EEG activity among the four RVS conditions in the Pd time window (details in Supplementary material). Therefore, the observed Pd changes could not be attributed to the contamination of motor-related EEG activity.

5.3. SSVEP entrainment from the view of functional cerebral asymmetry

Using time-frequency analyses, this study replicated previous finding that RVS could induce frequency-tagged SSVEP responses at the stimulation frequency of RVS and its harmonics (Herrmann, 2001). In previous studies, SSVEPs have been widely used for studying dynamic neural processes, such as attention control (Wittenhagen and Mattingley, 2019; Kritzman et al., 2022), working memory (Ellis et al., 2006), emotion recognition (Deng et al., 2020), visual information processing (Hansen et al., 2019; Montani et al., 2019) and visual perception (Chicherov and Herzog, 2015). In their views, the frequency-tagged SSVEP can reflect the dynamic neural processes but not influence them, thus making it a good candidate biomarker of brain functioning. However, the frequency-tagging approach is challenged by a contradictory view, i.e., neural entrainment, which argues SSVEP is at least partly generated by the endogenous entrainment of neural oscillations and thus can bring functional changes in cognitive processes (Keitel et al., 2019). It has been demonstrated that rhythmic sensory stimulation (RSS) can entrain endogenous neural oscillations and furtherly affect the subject's behaviors when they perform some mental tasks, such as attentional selection (Bauer et al., 2009; Gulbinaite et al., 2017), memory (Williams, 2001; Garcia-Argibay et al., 2019; Albouy et al., 2022) and spatial discrimination (Schlieper and Dinse, 2012; Ross et al., 2022). Such behavioral effects caused by RSS could be attributed to the rhythmic shifting of excitability in neuronal ensembles (Lakatos et al., 2008; Calderone et al., 2014) or the interaction between exogenous rhythmic input and endogenous neural rhythm (Spaak et al., 2014; Gulbinaite et al., 2017). Yet, Although there are mounting studies on the rhythmic entrainment phenomenon, the underlying neural mechanism of SSVEP effects remains to be explored.

Using FCA analyses, we demonstrated that the generation of SSVEP were accompanied by functional changes in visuospatial selective attention, which support that SSVEP is more than a biomarker of visual functioning, but also can bring about functional effects via rhythmic entrainment. As one of the most prominent features of cerebral organization, FCA has been widely documented in both human and non-human species (Güntürkün et al., 2000; Hong et al., 2015; Wang et al., 2016; Chen et al., 2018). Abundant converging evidence has revealed the close link between FCA and behavioral output in many cognitive processes, such as visual recognition and cognitive control (Ambrosini and Vallesi, 2016; Schnell et al., 2018). Notably, the degree of FCA is susceptible to many factors, especially environmental stimulation (Shalev et al., 2018; Chiandetti and Vallortigara, 2019). Therefore, we would like to examine whether RVS could impact attention behavior and alter the corresponding neural activities of asymmetry patterns. As a result, we found that RVSs modulated the left–right attentional asymmetries in RT and EEG activities in a consistent manner. These findings demonstrated that RVS could alter attentional processes with asymmetrical behavior and EEG consequences. Furthermore, since the event-related potential is closely related to neural oscillations (Klimesch et al., 2007), the changes in ERL in this study indicate that RVS can entrain functional-relevant neural oscillations. These RVS-induced changes reflect the functional effects of SSVEP, which modulates visuospatial selective attention by altering the left–right asymmetry degree of EEG activities and attention behavior.

In summary, we proposed a novel lateralized visual discrimination task with the background of RVS to explore the functional effects of SSVEP from the view of attention-related asymmetries. As a result, we show that the RVS can influence attentional performance and the degree of left–right attentional asymmetries in behavior and EEG activities. These findings support that SSVEPs play functional roles in neural processing. Furthermore, because FCA has been proposed to be prevalent and relevant to selective attention, we suggest that the observed Pd changes in EEG asymmetries provide new insights into the functional mechanism of SSVEP, which can modulate selective attention by regulating the attentional suppression of distractors during visual search.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board at Tianjin University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MX, RL, and DM designed research. RL and JY performed research. RL, MX, XX, and XZ analyzed data. RL, MX, T-PJ, JM, and DM wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the "STI 2030—Major Projects 2022ZD0208900," the National Natural Science Foundation of China (62122059, 61976152, 81925020, 62106170, and 62106173), and the Introduce Innovative Teams of 2021 "New High School 20 Items" Project (2021GXRC071).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1156890/full#supplementary-material

References

Albouy, P., Martinez-Moreno, Z. E., Hoyer, R. S., Zatorre, R. J., and Baillet, S. (2022). Supramodality of neural entrainment: rhythmic visual stimulation causally enhances auditory working memory performance. *Sci. Adv.* 8:eabj9782. doi: 10.1126/sciadv.abj9782

Ambrosini, E., and Vallesi, A. (2016). Asymmetry in prefrontal resting-state EEG spectral power underlies individual differences in phasic and sustained cognitive control. *NeuroImage* 124, 843–857. doi: 10.1016/j.neuroimage.2015.09.035

Bauer, F., Cheadle, S. W., Parton, A., Müller, H. J., and Usher, M. (2009). Gamma flicker triggers attentional selection without awareness. *Proc. Natl. Acad. Sci.* 106, 1666–1671. doi: 10.1073/pnas.0810496106

Beaton, A. A., Suller, S., and Workman, L. (2007). Visual laterality effects in readers of a deep and a shallow orthography. *Laterality* 12, 199–215. doi: 10.1080/13576500601112283

Beliaeva, V., and Polania, R. (2020). Can low-intensity tACS genuinely entrain neural activity in vivo? *Brain Stimulat.* 13, 1796–1799. doi: 10.1016/j.brs.2020.10.002

Bergerbest, D., Shilkrot, O., Joseph, M., and Salti, M. (2017). Right visual-field advantage in the attentional blink: asymmetry in attentional gating across time and space. *Atten. Percept. Psychophys.* 79, 1979–1992. doi: 10.3758/s13414-017-1356-z

Brederoo, S. G., Nieuwenstein, M. R., Cornelissen, F. W., and Lorist, M. M. (2019). Reproducibility of visual-field asymmetries: nine replication studies investigating lateralization of visual information processing. *Cortex* 111, 100–126. doi: 10.1016/j.cortex.2018.10.021

Bueno-Junior, L. S., Simon, N. W., Wegener, M. A., and Moghaddam, B. (2017). Repeated nicotine strengthens gamma oscillations in the prefrontal cortex and improves visual attention. *Neuropsychopharmacology* 42, 1590–1598. doi: 10.1038/npp.2017.15

Buzsaki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745

Calderone, D. J., Lakatos, P., Butler, P. D., and Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends Cogn. Sci.* 18, 300–309. doi: 10.1016/j.tics.2014.02.005

Carrasco, M., Roberts, M., Myers, C., and Shukla, L. (2022). Visual field asymmetries vary between children and adults. *Curr. Biol.* 32, R509–R510. doi: 10.1016/j.cub.2022.04.052

Chen, Z., Zhao, X., Fan, J., and Chen, A. (2018). Functional cerebral asymmetry analyses reveal how the control system implements its flexibility. *Hum. Brain Mapp.* 39, 4678–4688. doi: 10.1002/hbm.24313

Chiandetti, C., and Vallortigara, G. (2019). Distinct effect of early and late embryonic light-stimulation on chicks' lateralization. *Neuroscience* 414, 1–7. doi: 10.1016/j.neuroscience.2019.06.036

Chicherov, V., and Herzog, M. H. (2015). Targets but not flankers are suppressed in crowding as revealed by EEG frequency tagging. *NeuroImage* 119, 325–331. doi: 10.1016/j.neuroimage.2015.06.047

Corballis, M. C. (2009). The evolution and genetics of cerebral asymmetry. *Philos. Trans. R Soc. Lond. B Biol.* 364, 867–879. doi: 10.1098/rstb.2008.0232

Cucu, M. O., Kazanina, N., and Houghton, C. (2022). Syllable-initial phonemes affect neural entrainment to consonant-vowel syllables. *Front. Neurosci.* 16:826105. doi: 10.3389/fnins.2022.826105

Davidson, M. J., Mithen, W., Hogendoorn, H., Boxtel, J. J. A. V., and Tsuchiya, N. (2020). The SSVEP tracks attention, not consciousness, during perceptual filling-in. *eLife Sci.* 9:e60031. doi: 10.7554/eLife.60031

Deng, C., and Rogers, L. J. (2002). Prehatching visual experience and lateralization in the visual Wulst of the chick. *Behav. Brain Res.* 134, 375–385. doi: 10.1016/s0166-4328(02)00050-5

Deng, C., Tong, J., Deng, X., Zhang, Z., and Qin, Y. (2020). Emotion recognition positively correlates with steady-state visual evoked potential amplitude and alpha entrainment. *Neuroscience* 434, 191–199. doi: 10.1016/j.neuroscience.2020.01.045

Dreyer, A. M., Herrmann, C. S., and Rieger, J. W. (2017). Tradeoff between user experience and BCI classification accuracy with frequency modulated steady-state visual evoked potentials. *Front. Hum. Neurosci.* 11:391. doi: 10.3389/fnhum.2017.00391

Ellis, K. A., Silberstein, R. B., and Nathan, P. J. (2006). Exploring the temporal dynamics of the spatial working memory n-back task using steady state visual evoked potentials (SSVEP). *NeuroImage* 31, 1741–1751. doi: 10.1016/j.neuroimage.2006.02.014

Fuggetta, G., and Duke, P. A. (2017). Enhancing links between visual short term memory, visual attention and cognitive control processes through practice: an electrophysiological insight. *Biol. Psychol.* 126, 48–60. doi: 10.1016/j.biopsycho.2017.04.004

Garcia-Argibay, M., Santed, M. A., and Reales, J. M. (2019). Binaural auditory beats affect long-term memory. *Psychol. Res.* 83, 1124–1136. doi: 10.1007/s00426-017-0959-2

Gerlach, C., and Poirel, N. (2020). Who's got the global advantage? Visual field differences in processing of global and local shape. *Cognition* 195:104131. doi: 10.1016/j.cognition.2019.104131

Gregori Grgič, R., Calore, E., and de'Sperati, C. (2016). Covert enaction at work: recording the continuous movements of visuospatial attention to visible or imagined targets by means of steady-state visual evoked potentials (SSVEPs). *Cortex* 74, 31–52. doi: 10.1016/j.cortex.2015.10.008

Gulbinaite, R., van Viegen, T., Wieling, M., Cohen, M. X., and VanRullen, R. (2017). Individual alpha peak frequency predicts 10 Hz flicker effects on selective attention. *J. Neurosci.* 37, 10173–10184. doi: 10.1523/jneurosci.1163-17.2017

Güntürkün, O., Diekamp, B., Manns, M., Nottelmann, F., Prior, H., Schwarz, A., et al. (2000). Asymmetry pays: visual lateralization improves discrimination success in pigeons. *Curr. Biol.* 10, 1079–1081. doi: 10.1016/s0960-9822(00)00671-0

Hansen, B. C., Field, D. J., Greene, M. R., Olson, C., and Miskovic, V. (2019). Towards a state-space geometry of neural responses to natural scenes: a steady-state approach. *NeuroImage* 201:116027. doi: 10.1016/j.neuroimage.2019.116027

Hanslmayr, S., Axmacher, N., and Inman, C. S. (2019). Modulating human memory via entrainment of brain oscillations. *Trends Neurosci.* 42, 485–499. doi: 10.1016/j.tins.2019.04.004

He, Q., Yang, X. Y., Gong, B., Bi, K., and Fang, F. (2022). Boosting visual perceptual learning by transcranial alternating current stimulation over the visual cortex at alpha frequency. *Brain Stimulat.* 15, 546–553. doi: 10.1016/j.brs.2022.02.018

Herrmann, C. S. (2001). Human EEG responses to 1-100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp. Brain Res.* 137, 346–353. doi: 10.1007/s002210100682

Hickey, C., Di Lollo, V., and McDonald, J. J. (2009). Electrophysiological indices of target and distractor processing in visual search. *J. Cogn. Neurosci.* 21, 760–775. doi: 10.1162/jocn.2009.21039

Hong, X. F., Sun, J. F., Bengson, J. J., Mangun, G. R., and Tong, S. B. (2015). Normal aging selectively diminishes alpha lateralization in visual spatial attention. *NeuroImage* 106, 353–363. doi: 10.1016/j.neuroimage.2014.11.019

Jones, S. R., and Sliva, D. D. (2020). Is alpha asymmetry a byproduct or cause of spatial attention? New evidence alpha neurofeedback controls measures of spatial attention. *Neuron* 105, 404–406. doi: 10.1016/j.neuron.2019.12.033

Kappenman, E. S., Geddert, R., Farrens, J. L., McDonald, J. J., and Hajcak, G. (2021). Recoiling from threat: anxiety is related to heightened suppression of threat, not increased attention to threat. *Clin. Psychol. Sci.* 9, 434–448. doi: 10.1177/2167702620961074

Keitel, C., Keitel, A., Benwell, C. S., Daube, C., and Gross, J. (2019). Stimulus-driven brain rhythms within the alpha band: the attentional-modulation conundrum. *J. Neurosci.* 39, 3119–3129. doi: 10.1523/JNEUROSCI.1633-18.2019

Kertesz, A., Black, S. E., Polk, M., and Howell, J. (1986). Cerebral asymmetries on magnetic resonance imaging. *Cortex* 22, 117–127. doi: 10.1016/s0010-9452(86)80036-3

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., and Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neurosci. Biobehav. Rev.* 31, 1003–1016. doi: 10.1016/j.neubiorev.2007.03.005

Kritzman, L., Eidelman-Rothman, M., Keil, A., Freche, D., Sheppes, G., and Levit-Binnun, N. (2022). Steady-state visual evoked potentials differentiate between internally and externally directed attention. *NeuroImage* 254:119133. doi: 10.1016/j.neuroimage.2022.119133

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735

Luck, S. J., and Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31, 291–308. doi: 10.1111/j.1469-8986.1994.tb02218.x

Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends Cogn. Sci.* 8, 204–210. doi: 10.1016/j.tics.2004.03.008

Martens, U., and Hübner, R. (2013). Functional hemispheric asymmetries of global/local processing mirrored by the steady-state visual evoked potential. *Brain Cogn.* 81, 161–166. doi: 10.1016/j.bandc.2012.11.005

Mazza, V., Turatto, M., and Caramazza, A. (2009). Attention selection, distractor suppression and N2pc. *Cortex* 45, 879–890. doi: 10.1016/j.cortex.2008.10.009

Montani, V., Chanoine, V., Grainger, J., and Ziegler, J. C. (2019). Frequency-tagged visual evoked responses track syllable effects in visual word recognition. *Cortex* 121, 60–77. doi: 10.1016/j.cortex.2019.08.014

Moratti, S., Clementz, B. A., Gao, Y., Ortiz, T., and Keil, A. (2007). Neural mechanisms of evoked oscillations: stability and interaction with transient events. *Hum. Brain Mapp.* 28, 1318–1333. doi: 10.1002/hbm.20342

Mueller, M. M., Andersen, S., Trujillo, N. J., Valdes-Sosa, P., Malinowski, P., and Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14250–14254. doi: 10.1073/pnas.0606668103

Mulligan, D. J., Palopoli, A. C., van den Heuvel, M. I., Thomason, M. E., and Trentacosta, C. J. (2022). Frontal alpha asymmetry in response to stressor moderates the relation between parenting hassles and child externalizing problems. *Front. Neurosci.* 16:917300. doi: 10.3389/fnins.2022.917300

Notbohm, A., Kurths, J., and Herrmann, C. S. (2016). Modification of brain oscillations via rhythmic light stimulation provides evidence for entrainment but not for superposition of event-related responses. *Front. Hum. Neurosci.* 10:10. doi: 10.3389/fnhum.2016.00010

Oostenveld, R., Stegeman, D. F., Praamstra, P., and van Oosterom, A. (2003). Brain symmetry and topographic analysis of lateralized event-related potentials. *Clin. Neurophysiol.* 114, 1194–1202. doi: 10.1016/s1388-2457(03)00059-2

Poynter, W., Ingram, P., and Minor, S. (2010). Visual field asymmetries in attention vary with self-reported attention deficits. *Brain Cogn.* 72, 355–361. doi: 10.1016/j.bandc.2009.10.014

Resnick, S. M., Lazar, J., Gur, R. E., and Gur, R. C. (1994). The stability of tachistoscopic measures of hemispheric specialization. *Neuropsychologia* 32, 1419–1430. doi: 10.1016/0028-3932(94)00075-1

Riddle, J., Scimeca, J. M., Cellier, D., Dhanani, S., and D'Esposito, M. (2020). Causal evidence for a role of theta and alpha oscillations in the control of working memory. *Curr. Biol.* 30, 1748–1754.e4. doi: 10.1016/j.cub.2020.02.065

Rogers, L. J. (2006). Factors influencing development of lateralization. *Cortex* 42, 107–109. doi: 10.1016/s0010-9452(08)70332-0

Ross, B., Dobri, S., Jamali, S., and Bartel, L. (2022). Entrainment of somatosensory beta and gamma oscillations accompany improvement in tactile acuity after periodic and aperiodic repetitive sensory stimulation. *Int. J. Psychophysiol.* 177, 11–26. doi: 10.1016/j.ijpsycho.2022.04.007

Salahub, C., and Emrich, S. M. (2020). Fear not! Anxiety biases attentional enhancement of threat without impairing working memory filtering. *Cogn. Affect. Behav. Neurosci.* 20, 1248–1260. doi: 10.3758/s13415-020-00831-3

Sawaki, R., Geng, J. J., and Luck, S. J. (2012). A common neural mechanism for preventing and terminating the allocation of attention. *J. Neurosci.* 32, 10725–10736. doi: 10.1523/jneurosci.1864-12.2012

Schlieper, S., and Dinse, H. R. (2012). Perceptual improvement following repetitive sensory stimulation depends monotonically on stimulation intensity. *Brain Stimulat.* 5, 647–651. doi: 10.1016/j.brs.2011.07.002

Schnell, A. K., Bellanger, C., Vallortigara, G., and Jozet-Alves, C. (2018). Visual asymmetries in cuttlefish during brightness matching for camouflage. *Curr. Biol.* 28, R925–R926. doi: 10.1016/j.cub.2018.07.019

Shalev, N., De Wandel, L., Dockree, P., Demeyere, N., and Chechlacz, M. (2018). Beyond time and space: the effect of a lateralized sustained attention task and brain stimulation on spatial and selective attention. *Cortex* 107, 131–147. doi: 10.1016/j.cortex.2017.09.009

Spaak, E., de Lange, F. P., and Jensen, O. (2014). Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. *J. Neurosci.* 34, 3536–3544. doi: 10.1523/jneurosci.4385-13.2014

Srinivasan, R., Bibi, F. A., and Nunez, P. L. (2006). Steady-state visual evoked potentials: distributed local sources and wave-like dynamics are sensitive to flicker frequency. *Brain Topogr.* 18, 167–187. doi: 10.1007/s10548-006-0267-4

Stefanac, N. R., Zhou, S. H., Spencer-Smith, M. M., O'Connell, R., and Bellgrove, M. A. (2021). A neural index of inefficient evidence accumulation in dyslexia underlying slow perceptual decision making. *Cortex* 142, 122–137. doi: 10.1016/j.cortex.2021.05.021

Stevens, W. D., Kahn, I., Wig, G. S., and Schacter, D. L. (2012). Hemispheric asymmetry of visual scene processing in the human brain: evidence from repetition priming and intrinsic activity. *Cereb. Cortex* 22, 1935–1949. doi: 10.1093/cercor/bhr273

Suarez-Suarez, S., Holguin, S. R., Cadaueira, F., Nobre, A. C., and Doallo, S. (2019). Punishment-related memory-guided attention: neural dynamics of perceptual modulation. *Cortex* 115, 231–245. doi: 10.1016/j.cortex.2019.01.029

Thut, G., Veniero, D., Romei, V., Miniussi, C., Schyns, P., and Gross, J. (2011). Rhythmic TMS causes local entrainment of natural oscillatory signatures. *Curr. Biol.* 21, 1176–1185. doi: 10.1016/j.cub.2011.05.049

Van der Haegen, L., and Brysbaert, M. (2011). The mechanisms underlying the interhemispheric integration of information in foveal word recognition: evidence for transcortical inhibition. *Brain Lang.* 118, 81–89. doi: 10.1016/j.bandl.2010.03.006

Vergilino Perez, D., Lemoine, C., Siéroff, E., Ergis, A. M., Bouhired, R., Rigault, E., et al. (2012). The role of saccade preparation in lateralized word recognition: evidence for the attentional bias theory. *Neuropsychologia* 50, 2796–2804. doi: 10.1016/j.neuropsychologia.2012.07.039

Wang, J., Tian, Y., Wang, M., Cao, L., Wu, H., Zhang, Y., et al. (2016). A lateralized top-down network for visuospatial attention and neglect. *Brain Imaging Behav.* 10, 1029–1037. doi: 10.1007/s11682-015-9460-y

Williams, J. H. (2001). Frequency-specific effects of flicker on recognition memory. *Neuroscience* 104, 283–286. doi: 10.1016/s0306-4522(00)00579-0

Wittenhagen, L., and Mattingley, J. B. (2019). Steady-state visual evoked potentials reveal enhanced neural responses to illusory surfaces during a concurrent visual attention task. *Cortex* 117, 217–227. doi: 10.1016/j.cortex.2019.03.014

Wstmann, M., Alavash, M., and Obleser, J. (2019). Alpha oscillations in the human brain implement distractor suppression independent of target selection. *J. Neurosci.* 39, 9797–9805. doi: 10.1523/JNEUROSCI.1954-19.2019

Zhang, S., Chen, X., Wang, Y., Liu, B., and Gao, X. (2022). Visual field inhomogeneous in brain-computer interfaces based on rapid serial visual presentation. *J. Neural Eng.* 19:016015. doi: 10.1088/1741-2552/ac4a3e

Zhang, Y. Y., Zhang, Y. F., Cai, P., Luo, H., and Fang, F. (2019). The causal role of alpha-oscillations in feature binding. *Proc. Natl. Acad. Sci. U. S. A.* 116, 17023–17028. doi: 10.1073/pnas.1904160116

Glossary

2D	two-dimensional
ANOVA	analysis of variance
BCIs	brain-computer interfaces
EEG	electroencephalography
ERLs	event-related lateralizations
ERSP	event-related spectral perturbation
FCA	functional cerebral asymmetry
HEOG	horizontal electrooculogram
ITC	inter-trial coherence
LI_{acc}	laterality index of discrimination accuracy
LI_{RT}	laterality index of reaction time
LVF	left visual field
N2pc	N2 posterior contralateral
Pd	distractor positivity
RT	reaction time
RVF	right visual field
RVS	rhythmic visual stimulation
SSVEPs	steady-state visual evoked potentials
STFT	Short-Time Fourier Transform
VEOG	vertical electrooculogram

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 12 June 2023
DOI 10.3389/fnins.2023.1148205

OPEN ACCESS

EDITED BY
Miguel Ángel Martín-Pascual,
Spanish Public Television, Spain

REVIEWED BY
Maximilian Nentwich,
City College of New York (CUNY), United States
Monica Perusquia-Hernandez,
Nara Institute of Science and Technology (NAIST), Japan

*CORRESPONDENCE
Vladimir Kosonogov
✉ vkosonogov@hse.ru

RECEIVED 19 January 2023
ACCEPTED 17 May 2023
PUBLISHED 12 June 2023

CITATION
Kosonogov V, Shelepenkov D and Rudenkiy N (2023) EEG and peripheral markers of viewer ratings: a study of short films. *Front. Neurosci.* 17:1148205.
doi: 10.3389/fnins.2023.1148205

EEG and peripheral markers of viewer ratings: a study of short films

Vladimir Kosonogov[1]*, Danila Shelepenkov[1] and Nikita Rudenkiy[2]

[1]Institute for Cognitive Neuroscience, HSE University, Moscow, Russia, [2]National Research Nuclear University MEPhI, Moscow, Russia

Introduction: Cinema is an important part of modern culture, influencing millions of viewers. Research suggested many models for the prediction of film success, one of them being the use of neuroscientific tools. The aim of our study was to find physiological markers of viewer perception and correlate them to short film ratings given by our subjects. Short films are used as a test case for directors and screenwriters and can be created to raise funding for future projects; however, they have not been studied properly with physiological methods.

Methods: We recorded electroencephalography (18 sensors), facial electromyography (*corrugator supercilii* and *zygomaticus major*), photoplethysmography, and skin conductance in 21 participants while watching and evaluating 8 short films (4 dramas and 4 comedies). Also, we used machine learning (CatBoost, SVR) to predict the exact rating of each film (from 1 to 10), based on all physiological indicators. In addition, we classified each film as low or high rated by our subjects (with Logistic Regression, KNN, decision tree, CatBoost, and SVC).

Results: The results showed that ratings did not differ between genres. *Corrugator supercilii* activity ("frowning" muscle) was larger when watching dramas; whereas *zygomaticus major* ("smiling" muscle) activity was larger during the watching of comedies. Of all somatic and vegetative markers, only *zygomaticus major* activity, PNN50, SD1/SD2 (heart rate variability parameters) positively correlated to the film ratings. The EEG engagement indices, beta/(alpha+theta) and beta/alpha correlated positively with the film ratings in the majority of sensors. Arousal $(\text{beta}_{F3}+\text{beta}_{F4})/(\text{alpha}_{F3}+\text{alpha}_{F4})$, and valence $(\text{alpha}_{F4}/\text{beta}_{F4})$ - $(\text{alpha}_{F3}/\text{beta}_{F3})$ indices also correlated positively to film ratings. When we attempted to predict exact ratings, MAPE was 0.55. As for the binary classification, logistic regression yielded the best values (area under the ROC curve=0.62) than other methods (0.51–0.60).

Discussion: Overall, we revealed EEG and peripheral markers, which reflect viewer ratings and can predict them to a certain extent. In general, high film ratings can reflect a fusion of high arousal and different valence, positive valence being more important. These findings broaden our knowledge about the physiological basis of viewer perception and can be potentially used at the stage of film production.

KEYWORDS
short films, EEG, heart rate, skin conductance, valence, arousal, *zygomaticus major*, *corrugator supercilii*

Introduction

Cinema is an important part of modern popular culture, influencing millions of viewers. In 2021, according to the Theatrical and Home/Mobile Entertainment Market Environment report, 403 films were released in Canada and the US alone, and the combined global theatrical and home/mobile entertainment market reached $99.7 billion (Motion Picture Association, 2022).

However, in addition to great budgets, the film industry is characterised by competition and risk. Due to high production costs and marketing budgets, even well-known films may not break even at the box office (income from ticket sales). For example, in 2021, famous films with big marketing budgets like *The Suicide Squad* and *The Last Duel* failed to make enough money at the box office to recoup their budget. Some authors point out that more than 75% of new film releases face a net loss during their run in theatres (Boksem and Smidts, 2015). Because of the risk of losing money, companies in the film industry are turning to various methods of film promotion and prediction of success.

To mitigate such risks, directors, screenwriters, and film studios need to know how audiences would react to films beforehand. Therefore, research suggested many models for the prediction of film success, which is commonly understood as high box office revenue. Typically, such models include information about the actors, locations, budget, release time, polls, etc. (for more detailed information, see Lash and Zhao, 2016; Ahmad et al., 2017). Besides box office receipts, another measure of a film's success is its rating on websites like *RottenTamatoes*, *Metacritic*, *Kinopoisk,* and *Internet Movie Database (IMDb)*.Scores on websites received from all users and critics are related to viewing revenue and viewing satisfaction (Moon et al., 2010). It was shown that viewer ratings are the most effective predictors of financial income (Dellarocas et al., 2005). Recently, researchers also began to to apply machine-learning algorithms to predict film scores, since this have a very promising significance to the film industry. Thus, Dhir and Raj (2018) found the importance of the number of *Facebook* likes, film genres, the number of critics in reviews in the film score prediction. Latif and Afzal (2016) derived ratings from the number of votes on Oscar awards. In addition, film ratings can be predicted from the facial expression of viewers (Shetty et al., 2021).

This area of knowledge is actively developing and looking for new approaches, one of them being the use of neuroscientific tools and data. This area of research offers a wide range of techniques such as functional magnetic resonance imaging (fMRI; Hasson et al., 2008) and electroencephalography (EEG; Heimann et al., 2014). The most common tool for such studies is the EEG. This is due to the fact that EEG studies are easier to reproduce, EEG is relatively cheap, which is why it can be more applicable to marketing research (Nilashi et al., 2020). Peripheral physiological measures, such as automated facial emotion recognition, heart rate, respiration rate and electrodermal activity (EDA) have been used less often (Aznar et al., 2019). But recent studies show the presence of inter-subject correlation in physiological data during watching movies, such as heart rate (Madsen and Parra, 2022) and galvanic skin response (Palumbo et al., 2017). Research also shows that physiological data such as corrugator and zygomaticus EMG correlate with valence of films (Sato et al., 2020). Overall, these studies show that physiological measures can potentially contribute to predict a film rating and commercial success.

The use of neuroscientific methods has become popular following research showing that they can improve self-reported data (Boksem and Smidts, 2015). Thus, Boksem and Smidts (2015) measured the EEG activity while watching 18 film trailers, and collected behavioural information about liking movies and willingness to pay for tickets. They showed that an important predictor of box office performance was EEG activity around the fronto-central regions in the gamma-band, while the results of the willingness-to-pay poll were not a significant predictor of box office success. This conclusion was confirmed in a following study (Barnett and Cerf, 2017). They measured brain activity of fifty-eight participants with EEG and peripheral data in a commercial theatre while watching film trailers. Then, they calculated the relative levels of neural similarity, which they called cross-brain correlation (CBC). The level of CBC could predict film recall and box office revenue, at the same time the autonomic nervous system data, like EDA, cardiac or respiratory rate, were not associated with recall or box office success (Barnett and Cerf, 2017). However, box office receipts depend on many factors, including distribution related factors (e.g., budget or franchise), brand and star effects (e.g., top actors or directors), and evaluation sources (e.g., critics and audience rating) and region-specific variables (Gaenssle et al., 2018). We focused on the neural correlates of audience rating and user rating prediction, which had rarely been considered in film neuroscience before. However, a recent study has shown that EEG recordings while watching movie trailers can also be successfully used to predict ranks of subjects' preferences using machine learning methods (Shetty et al., 2021).

The found physiological markers, capable of predicting ratings or commercial success, are usually linked with cognitive states of focused attention, the brain reward network, emotional response, engagement levels, and enjoyment. For example, Christoforou et al. (2017) have found that the gamma-band of EEG while watching trailers significantly predicted box office success on the first weekend and in the following few weeks. They associated gamma-activity while watching, with the trailer or film's ability to capture the viewer's attention. Other EEG components related to engagement and enjoyment can be used to predict other aspects of film success such as rating. For example, activity in the beta-band of EEG is usually associated with an individual preference for short-term rewards (Cohen et al., 2007). In this vein, Boksem and Smidts (2015) showed that beta-activity was related to a high viewer preference for films.

We hypothesised that such well-spread EEG markers as engagement indices, beta/(alpha+theta) and beta/alpha (Pope et al., 1995) and valence and arousal indices (Giraldo and Ramirez, 2013) may be other accurate markers for predicting film ratings. Although these indices have been widely used for engagement and workload measurement (Berka et al., 2007; Lelis-Torres et al., 2017; Apicella et al., 2022), to our knowledge, they have not been applied to study film perception and film ratings. The frequency of the beta-band may be related to the activation of the visual system and also to the state of attention (Molteni et al., 2007). An increase in alpha- and theta-activity is usually associated with a decrease of attention and vigilance (Coelli et al., 2015), so a complex index can give a more accurate result. In other words, engagement indices, widely used in different psychophysiological studies, could enrich the field of neurocinematics, since they take into account different EEG bands, related to both activation and deactivation. Supposedly, complex or composite indices could capture differences or relationships, invisible when the EEG rhythms are studied independently (Shestyuk et al., 2019).

We also applied machine-learning methods to predict the rating of films based on EEG data and peripheral indicators, expecting that the engagement indices would be among the significant predictors of film ratings. By now, researchers predicted film scores from *Facebook* likes and number of critics (Dhir and

Raj, 2018), Oscar votes (Latif and Afzal, 2016), or even facial expression (Shetty et al., 2021). Previously, machine learning was also shown to be very fruitful to recognize, via EEG signals, such mental states as, for example, engagement, workload (Berka et al., 2007; Walter et al., 2017) and emotions (Soleymani et al., 2015; Rayatdoost and Soleymani, 2018; Rayatdoost et al., 2020). This motivated us to apply machine-learning techniques to predict film rationg from EEG and peripheral signals.

In addition, previous studies have mostly used film trailers as stimuli (Boksem and Smidts, 2015; Liu et al., 2016; Christoforou et al., 2017; Wu et al., 2017, 2018; Dushantha et al., 2020). Trailers are convenient stimuli for studying as they are actively used in marketing research. However, they display a number of drawbacks, as they consist of disparate scenes of the film and rarely present a coherent narrative. In the current study, we focused on short films. According to the Academy of Motion Picture Arts and Sciences, short films are "original films that are less than 40 min long…" Short films are usually used as a test case for directors and screenwriters and can be created to raise funding for future projects because short films are much cheaper to produce. We suppose that due to their shorter duration and at the same time the integrity of the narrative, short films could be a suitable object for a psychophysiological study.

Hence, the aim of our study was to find physiological markers of viewer perception and correlate them to short film ratings. For this purpose, we recorded electroencephalography, facial electromyography (*corrugator supercilii* and *zygomaticus major*), photoplethysmography, and skin conductance that were supposed to reflect viewer engagement (Palumbo et al., 2017) and emotions (Lundqvist, 1995; Li et al., 2018; Sato et al., 2020; Madsen and Parra, 2022). As for the cardiac activity, along with heart rate, we also extracted different features of heart rate variability (HRV), which are frequently considered to reflect emotional states (Kuoppa et al., 2016; Shi et al., 2017; for a review see Zhu et al., 2019). Finally, we used machine learning to predict both the exact rating of each film and classify them as low or high rated.

Materials and methods

Sample

Twenty-one healthy volunteers (76.19% females) participated in the experiment in exchange for a monetary reward (an equivalent of 20 USD at purchasing parity power in 2021). Their mean age was 22.5, SD = 4.0. The study was carried out in accordance with the Declaration of Helsinki and was approved by the local research ethics committee (#52, 14.01.2019). Each participant provided written consent for his or her participation in the study.

Stimuli

We selected eight short films with different levels of scores on the *Kinopoisk*, a film rating database. The mean score was 7.5 (SD = 0.5, min = 6.8, max = 8.1) on a scale of 1 to 10. The number of ratings for each video was over 800, with an average of 5,400 ratings. We removed the titles, so the subjects could watch only the films. The mean duration was 6 min 4 s; the range was from 4 min 21 s to 7 min 25 s (Table 1).

Procedure

Participants were informed that they would have to watch and evaluate a number of short films. Each participant watched short films on a 31.5-inch computer screen in a random order. After watching each film, participants were asked "to rate the film" on a scale of 1 to 10 (where 10 meant the best grade, following *Kinopoisk* or *IMDb* scales). There was a rest period for 60 s between the evaluation offset and a new film onset. Participants were asked not to move and blink much, because the rest periods were used in the analyses as well (see below). The procedure was programmed in *PsychoPy* (Peirce, 2008).

Data collection

To record and amplify physiological signals, we used ActiChamp equipment (Brain Products, Germany). The signal recording frequency was 1,000 Hz. EEG signals were recorded from 18 active electrodes (F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P8, P4, Pz, P3, P7, O1, Oz, O2), according to the 10–20 system (Jasper, 1958). The impedance of each electrode was kept below 15 kΩ. Tp9 and Tp10 electrodes were used as an online reference. Vertical eye movements were recorded with one electrode, which was placed on the orbicularis oculi muscle under the right eye. A photoplethysmograph was put on the middle finger of the left hand. Electromyographic activity of *zygomaticus major* (the "smiling" muscle) and *corrugator supercilii* (the "frowning" muscle) was recorded by placing 4-mm Ag/AgCl surface electrodes (Fridlund and Cacioppo, 1986). Although it was shown that the left side of the face is more sensitive to corresponding emotions (Dimberg and Petterson, 2000), due to technical problems we recorded EMG from the right side. Skin conductance was measured by placing two Ag/AgCl surface electrodes on the index and ring fingers of the left hand (non-dominant for all participants).

Data reduction

The EEG preprocessing was conducted in MNE Python (Gramfort, 2013). The raw EEG was downsampled from 1,000 Hz to 125 Hz to reduce computational complexity and filtered with lower-pass edge of 0.05 and upper-pass of 30 Hz. We decided to exclude the gamma-activity, since recent studies show that it is not possible to completely clear the signal from muscle activity in the gamma-band. And also that data contamination from muscle activity in the gamma band >30 Hz over the entire scalp is higher than in the beta band. Moreover, the analysis of independent components provides effective clipping of EMG in EEG beta activity in almost all leads, but not in gamma (Pope et al., 2022). After that, we interpolated bad channels by fitting PyPrep Pipeline with RANSAC method (Bigdely-Shamlo et al., 2015). To correct EEG for eye blinks, we ran ICA decomposition from MNE.ICA module. To reject muscle artefacts, we deleted intervals (about 2% of the data) where the *z*-score was greater than 10. We estimated power spectral density using Welch's method in Yasa SciPy welch (Vallat and Walker, 2021). We computed the median power of the EEG in theta (4–8 Hz), alpha (8–12 Hz), and beta (12–30 Hz) bands in the one-second window (125 samples) with a 50% overlap.

TABLE 1 The description of short films used in the study.

Title	Year	Duration, s	Genre	Director	Rating page (consulted 1/3/2021)
One-Minute Time Machine	2014	318	Comedy	D. Avery	https://www.kinopoisk.ru/film/864243/
Star	2001	395	Comedy	G. Ritchie	https://www.kinopoisk.ru/film/12201/
The Flying Car	2002	364	Comedy	K. Smith	https://www.kinopoisk.ru/film/328034/
The Expert	2014	445	Comedy	L. Beinerts	https://www.kinopoisk.ru/film/838922/
One Hundredth of a Second	2006	261	Drama	S. Jacobson	https://www.kinopoisk.ru/film/272683/
The Gift	2010	391	Drama	I. Petukhov	https://www.kinopoisk.ru/film/841213/
Aningaaq	2013	346	Drama	J. Cuarón	https://www.kinopoisk.ru/film/788239/
Cargo	2013	394	Drama	B. Howling, Y. Ramke	https://www.kinopoisk.ru/film/756665/

We also got ratios of theta, alpha, and beta bands within each trial (one film, one subject). In addition, we calculated the engagement index as beta/(alpha+theta) and beta/alpha (Pope et al., 1995; Freeman et al., 1999), arousal index, $(\text{beta}_{F3}+\text{beta}_{F4})/(\text{alpha}_{F3}+\text{alpha}_{F4})$, and valence index, $(\text{alpha}_{F4}/\text{beta}_{F4})$ - $(\text{alpha}_{F3}/\text{beta}_{F3})$, where positive values mean positive emotions (Giraldo and Ramirez, 2013). For each variable, we subtracted the baseline (the mean value of 60-s rest period before each film) from the value of each trial. For machine-learning purposes, following an increasing trend towards the use of complexity analysis in quantifying neural activity, we additionally calculated brain entropy and complexity measures (Lau et al., 2022). Using neurokit2 (Makowski et al., 2021), we extracted Petrosian fractal dimension (PFD), differential entropy (DE), Katz's fractal dimension (KFD), Sevcik fractal dimension (SFD), permutation entropy (PEn), Shannon entropy (ShanEn), spectral entropy, singular value decomposition entropy (SVDEn), Fisher information (FI), Hjorth's complexity (Hjorth), relative roughness (RR) for each EEG channel and each film (Rahman et al., 2021).

Photoplethysmograms were processed with HeartPy, Python heart rate analysis toolkit (Van Gent et al., 2019). They were filtered using a Hampel filter with filter size parameter set on 6, that means that three data points on each side were taken to detect outliers and correct the signal. For each film, we extracted variables thought to reflect emotional states (Kuoppa et al., 2016; Shi et al., 2017; Zhu et al., 2019): mean heart rate (HR) and different characteristics of heart rate variability (HRV): the standard deviation of NN intervals (SDNN), the root mean square of the successive differences (RMSSD), the standard deviation between successive differences (SDSD), the proportion of NN20 and NN50 (pNN20 and pNN50), the median absolute deviation of RR intervals (MAD), and SD1, SD2, SD1/SD2 of Poincaré plot. For each variable, we subtracted the baseline (the mean value of 60-s rest period before each film) from the value of each trial.

Zygomaticus and corrugator EMG activity was processed in MNE-Python. We applied the FIR filter with a lower–pass of 10 Hz and upper-pass of 350 Hz, took absolute values of the signal and averaged it within each trial. Then we subtracted the baseline (the mean value of 60-s rest period before each film) from the EMG value of each trial.

Skin conductance was processed with Neurokit2 (Makowski et al., 2021). We excluded the tonic component, detected skin conductance responses, and extracted amplitudes of all peaks. For each trial, we summed all the peak amplitudes and divided by time in order to correct for different epoch length.

Data analysis

We compared film ratings and somatic and vegetative variables between genres (comedies/dramas), using *t*-tests for paired samples with Cohen's *d* as an effect size measure. Then we correlated film ratings with all physiological variables. In the analysis of EEG, Benjamini-Hochberg correction was applied for multiple correlations for each individual channel.

To predict film ratings, based on physiological data, we used all above-mentioned features. After this, we removed outliers defined as >3 SD or < -3 SD by column in our matrix. After this, we imputed missed values with a multivariate imputation by chained equations in which the specification occurs on at the variable level, excluding artificial correlations between them (mice; Van Buuren and Groothuis-Oudshoorn, 2011). The final matrix consisted of 168 cases (8 films × 21 subjects) and 522 columns (519 physiological features (see Supplementary materials), film, subject and rating). For each prediction, we extracted 15 most important physiological features, which then were used for training and prediction. The importance of a feature was computed as the reduction of the criterion brought by that feature. It is also known as the Gini importance (Nembrini et al., 2018). Thus, the final matrices were always 168 × 18 (15 physiological features, film, subject and rating). To predict continuous ratings (from 1 to 10), we applied CatBoost (CatBoost Python package; Prokhorenkova et al., 2018) and support vector regression (with Scikit-learn Python package). Additionally, we ranked all the ratings within each subject as low or high rated to apply binary classification with logistic regression, KNN, decision tree, CatBoost, support vector classification (with Scikit-learn Python package). To compute metrics on regression and classification, we used *k*-fold cross-validation. We applied the leave-one-film-out strategy; hence, for each of the eight films, the data collected from all participants who watched seven of them were utilized for training. Meanwhile, the other film was designated for testing purposes, and to forecast the ratings of all subjects for the eighth film. This process was carried out independently for each film, resulting in eight separate predictions (Kramer, 2016). In other words, we expected to predict the film ratings (from 1 to 10) or ranks (low/high) of a film, based on the EEG and peripheral signals data of this film and EEG and peripheral signals and ratings or ranks of seven other films. This might have an applied significance in future attempts to predict ratings at the film production stage. In simple words, a studio could collect physiological databases during film perception and after having shown some films and asked for ratings to a sample of subjects in a neuroscientific laboratory, they could predict subjects' ratings of new films.

Results

First, we analysed whether genres provoked different reactions. Film ratings did not differ depending on genres, $t\ (20) = 0.67$, $p = 0.50$, $M \pm SD_{\text{dramas}} = 6.12 \pm 2.43$, $M \pm SD_{\text{comedies}} = 6.37 \pm 2.40$. Corrugator supercilii activity was larger while watching dramas, $t\ (20) = 3.25$, $p = 0.004$, $d = 0.71$. Zygomaticus major activity was larger while watching comedies, $t\ (20) = 2.12$, $p = 0.047$, $d = 0.46$. All other somatic and vegetative variables did not show differences between comedies and dramas, $ts < 1.6$, $ps > 0.08$.

Second, we subjected all physiological variables to correlation analysis with the film ratings. Of all somatic and vegetative markers, only three were related to film ratings. *Zygomaticus major* activity positively correlated to film ratings, $r = 0.26$, $p = 0.001$. Also, PNN50 and SD1/SD2 (indices of HRV) as well positively correlated to film ratings, $r = 0.18$, $p = 0.019$ and $r = 0.16$, $p = 0.043$. Correlation analysis between EEG rhythms and film ratings yielded some significant results. Thus, the engagement index, beta/(alpha+theta), correlated positively with film ratings in the majority of sensors (Figure 1A). The higher the index was, the larger self-reported value was. We received a similar pattern when we calculated the engagement index as beta/alpha. It also correlated positively with film ratings in the majority of sensors (Figure 1B). Valence and arousal indices also positively correlated to film ratings ($r = 0.21$, $p = 0.010$ and $r = 0.24$, $p = 0.003$, respectively). Of note, the engagement index beta/(alpha+theta) positively correlated to the arousal index in all sensors ($0.31 < rs < 0.80$, all $ps < 0.001$), besides O1, Oz and O2, but to the valence index only in Cz, negatively ($r = -0.28$, $p = 0.007$). The engagement index beta/alpha positively correlated to the arousal index in all sensors ($0.29 < rs < 0.88$, all $ps < 0.001$), and negatively to the valence index only in Cz ($r = -0.23$, $p = 0.004$).

We also separately correlated the power of each band to film ratings and did not find any correlation (all $ps > 0.05$). Finally, within each subject we split films into low and high rated halves. Mean ratings were 4.55 and 8.02, $t = 13.34$, $p < 0.001$; but the interaction genre × half (low/high) was not significant, $p = 0.57$. However, we found no significant difference in any physiological measure between halves ($ts < 1.79$, $ps > 0.08$).

Third, we attempted to predict (using cross-validation) film ratings based on physiological data. We predicted the rating of each film based on the physiological data of all subjects recorded while watching the other films. As an example, the 15 variables with their importance in CatBoost regression for Film 1 can be found in Figure 2. For both models applied, the mean absolute prediction error (MAPE) was 0.53. The statistics of predictions for each film can be found in Table 2.

We then tried the binary classification of ratings. That is, within each subject we split films into low and high rated halves. After this, for each film we predicted whether it would receive a low or high rating by each subject, based on the binary ratings of all other films (Table 3). As an example, the 15 variables with their importance in CatBoost classification for Film 6 can be found in Figure 2. Logistic regression turned out to be the best predictive model. The mean area under the curve was 0.62 (with 0.50 being the random value). The best area under the curve (0.71) was found for Film 6. Other models yielded lower values of the area under the curve (0.51–0.59).

Discussion

The aim of our study was to explore physiological markers of short film perception and correlate them to film ratings. For this purpose, we measured electroencephalography, facial electromyography, photoplethysmography and skin conductance in 21 participants, while watching 8 short films (4 dramas and 4 comedies). We also used machine learning to predict the exact rating of each film and to classify each film as low or high rated.

First, we simply compared ratings and all physiological variables between two genres. *Corrugator supercilii* (the "frowning" muscle) activity was larger in response to dramas, while comedies provoked an increase in *zygomaticus major* (the "smiling" muscle) EMG. These findings are in accordance with the previous studies, which showed that unpleasant stimuli evoke an increase in *corrugator supercilii* activity, whereas *zygomaticus major* is activated by pleasant stimuli (Dimberg and Karlsson, 1997; Bullack et al., 2018). At the same time, heart rate and skin conductance did not differ. This may mean that both comedies and dramas evoked the same level of arousal, but with the opposite valence (Bradley et al., 2001). In other words, comedies were perceived positively, while dramas evoked negative emotions, but the same level of physiological arousal.

Curiously, ratings between genres did not differ, which may mean that they might reflect arousal rather than valence. We then split films into low and high rated halves (within each subject), but found no significant difference in any physiological measure between the halves. Also, the interaction between halves and genres did not have a significant effect on the ratings, which would mean that comedies and dramas were distributed equally between halves. In other words, both comedies and dramas were in low and high rated halves. This may explain that many films and genres are not perceived as pleasant (like horror), but attract the attention of viewers, presumably, due to the level of arousal. This is consistent with the suggestion that for many viewers, arousal itself may be an important reason for watching, like in the case of horrors or tragedies (Martin, 2019). Thus, in a study by Vecchiato et al. (2009), skin conductance (a marker of arousal) was not sensitive to differentiate TV commercials, seemingly because of the equally high level of arousal.

Second, correlation analysis between EEG rhythms and film ratings showed that the engagement index (calculated as beta/(alpha+theta) or beta/alpha), correlated positively with the ratings in the majority of sensors. The higher the index was, the larger self-reported value (film rating) was. We admit that the correlations were weak, however consistent throughout the scalp. At the same time, EEG indices of valence and arousal also correlated to film ratings. In other words, positive and arousing films (based on EEG) were evaluated with a higher rating. Also, engagement indices correlated to the arousal index. These findings confirm previous studies that revealed that engagement index reflects arousal rather than valence. Thus, Chaouachi et al. (2010) found that engagement index correlated with arousal, but not with valence in an educational process. McMahan et al. (2015) showed that the engagement index differentiated low and high intensity video games. It was positively related to arousal, and, at the same time, negatively to valence. In other words, the engagement

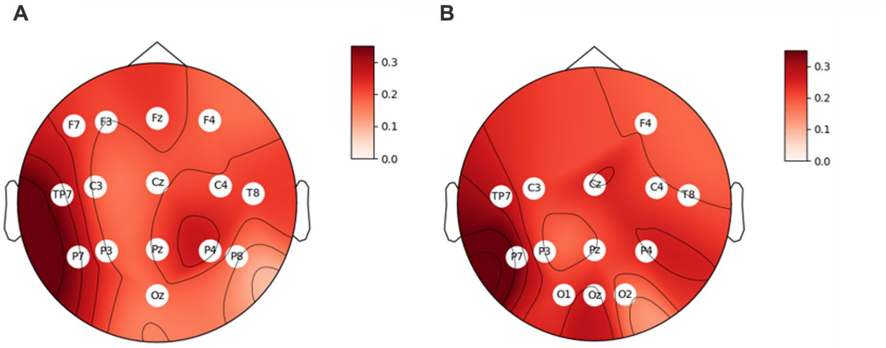


FIGURE 1
Correlation coefficients between film ratings and the engagement index, **(A)** beta/(alpha+theta), and **(B)** beta/alpha. The sensors where the correlation is significant (*ps*<0.05) are indicated with their titles.

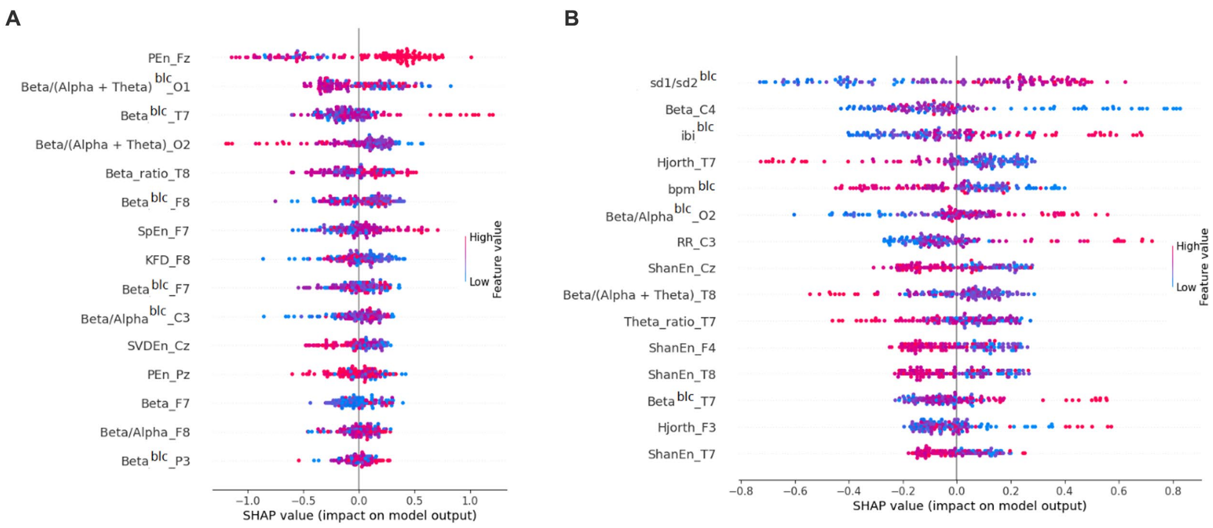


FIGURE 2
Importance of features for prediction in **(A)** CatBoost regression for Film 1 and **(B)** CatBoost classification for Film 6. Shapley Additive Explanation (SHAP) values attribute to each feature the change in the expected model prediction when conditioning on that feature (Lundberg and Lee, 2017). blc, baseline-corrected.

index increased in response to a more arousing and unpleasant video game event. Maran et al. (2017) also concluded that engaging and effective educational material should induce arousal states of different valence, both aversive and appetitive.

However, besides arousal, the EEG index of valence and *zygomaticus major* activity positively correlated to film ratings, as well as PNN50 and SD1/SD2 (HRV variables). This is in accordance with the previous findings on the frontal EEG asymmetry (Davidson, 2004), indicating the role of the left hemisphere in positive emotions. As for HRV, although being a controversial marker of affective states, it was found to reflect valence as well. Thus, Shi et al. (2017) revealed its increase during happy states in comparison to sad ones. On the contrary, Kuoppa et al. (2016) found a lower HRV in response to positive food, compared to negative food, but no difference for non-edible stimuli. Nevertheless, our results coincide with the data of Vecchiato et al. (2011) who showed that spectral EEG frontal asymmetries correlate with the experienced pleasantness of TV commercial advertisements. In this vein, Shestyuk et al. (2019), found a correlation between frontal asymmetry (pleasantness) and TV viewership (number of viewers). In addition, our data partly replicated the findings of Sato et al. (2020) who showed correlations between *corrugator* and *zygomaticus* EMG and the valence of films. Thus, at the same time, film ratings reflect valence and arousal experienced during perception. It is worth noting, that we deliberately avoided the typical affective self-report scales of valence and arousal. We wanted to replicate the scales used in the film industry of mere "rating," which, as we understand, does not equal valence, because even for films in the same genres and equal user ratings, the emotions of the audience can be radically different (Topal and Ozsoyoglu, 2016). We had a concern that the usage of all three scales (film rating, valence, arousal) could impact the perception of the "film rating" scale, that is, this could have suggested subjects to consider the nature of "film rating."

Third, we tested several machine learning models in order to predict ratings, based on the physiological data. We predicted the

TABLE 2 Statistics of rating predictions for each of 8 films, based on seven other films.

Film	Observed rating	CatBoost			SVR		
		MAPE	MAE	Predicted	MAPE	MAE	Predicted
1	7.52	0.26	1.96	6.13	0.30	2.04	6.45
2	6.50	0.37	1.85	6.30	0.34	1.82	6.29
3	4.86	0.62	2.05	6.41	0.72	2.25	6.76
4	6.62	0.49	2.52	6.10	0.61	2.44	6.60
5	6.52	0.86	2.39	6.21	0.73	2.06	6.28
6	6.10	0.39	1.84	6.15	0.33	1.37	6.25
7	6.52	0.38	2.17	6.19	0.36	1.98	6.01
8	5.33	0.84	2.36	6.64	0.85	2.17	6.70
Mean of 8 films		0.53	2.14		0.53	2.02	

TABLE 3 Statistics of binary predictions for each of 8 films, based on seven other films, and the mean values for different models.

	Logistic regression			CatBoost			KNN			Decision trees			SVC		
Film	Acc.	F1	AUC	Acc.	F1	AUC	Acc.	F1	AUC	Acc.	F1	AUC	Acc.	F1	AUC
1	0.52	0.55	0.62	0.52	0.58	0.57	0.57	0.67	0.55	0.38	0.48	0.37	0.29	0.35	0.30
2	0.60	0.60	0.61	0.70	0.67	0.70	0.40	0.33	0.39	0.65	0.59	0.64	0.60	0.56	0.60
3	0.62	0.43	0.67	0.52	0.38	0.61	0.57	0.31	0.54	0.52	0.38	0.61	0.62	0.33	0.57
4	0.52	0.55	0.53	0.52	0.62	0.50	0.52	0.50	0.54	0.52	0.50	0.54	0.48	0.56	0.46
5	0.71	0.80	0.67	0.57	0.73	0.50	0.52	0.69	0.46	0.57	0.64	0.56	0.52	0.50	0.54
6	0.70	0.67	0.71	0.70	0.57	0.67	0.55	0.53	0.56	0.55	0.53	0.56	0.55	0.61	0.60
7	0.48	0.52	0.48	0.57	0.61	0.58	0.38	0.38	0.40	0.33	0.36	0.34	0.43	0.40	0.47
8	0.62	0.64	0.67	0.62	0.56	0.62	0.62	0.56	0.62	0.52	0.58	0.59	0.62	0.50	0.60
Mean	0.60	0.60	0.62	0.59	0.59	0.59	0.52	0.50	0.51	0.51	0.51	0.53	0.51	0.47	0.51

acc. – accuracy, AUC – area under the curve.

exact rating of each film and classified each film as low or high rated. As for prediction of exact ratings, the MAPE was 0.55 both for CatBoost and SVR. When we classified films as low or high rated, the best area under the curve equalled 0.62 in the case of logistic regression. Interestingly, in a study by Dhir and Raj (2018), where they predicted film ratings, based on *Facebook* likes and number of critics, the prediction quality was low (F1 = 0.59), although their sample consisted of 5,043 films. Much better results on 2000 films were obtained by Latif and Afzal (2016), who used budget, genre, critics, Oscar votes to predict film ratings (ROC area = 0.93). This discrepancy can lie in the difference of films used in the study. Due to the restrictions of psychophysiological laboratory and our experimental plan, we could present only eight films, in comparison to the studies relied only on open data from large samples.

A future study could involve much more films and evaluated models with and without physiological data. To resolve this and present more films, future studies could be organised so that different subjects could watch some overlapping subsamples of films. Another possibility to expand this line of research would be to ask subjects to indicate a "dynamic valence/arousal rating" during the whole viewing (Nummenmaa et al., 2012). These time-series could then be correlated to physiological markers within each trial or on average. This would allow finding crucial scenes in films. We had a concern that the usage of different self-report scales (film rating, valence, arousal) could have impacted viewer perception. Therefore, future large-scale studies could go deeper into the question of the relationship between these scales in order to understand what psychological phenomena lie behind "film rating." Our physiological exploration implies that it can be a fusion of high arousal and different valence, positive valence being more important on average. Nevertheless, this would depend on genre, since some films inducing negative emotions, like horror (Zillmann, 1996) or sadness (Oliver, 1993), also obtain very high ratings.

Another limitation of our study lies in the recording of EMG from the right side the face. Previously, it was shown that the left side of the face is more sensitive to basic emotions. Thus, Dimberg and Petterson (2000) showed that *corrugator supercilii* and *zygomaticus major* were more activated on the left side of face, while expressing anger and happiness, respectively. This pattern was then confirmed by Zhou and Hu (2004). This difference could reflect the right hemisphere dominance in emotional expression. Supposedly, future studies in the field of neurocinematics could apply EMG sensors to both hemifaces for a more detailed analysis.

To conclude, we revealed that the engagement, valence and arousal indices of EEG, as well as the *zygomaticus* activity and some HRV variables, positively correlated to short film ratings given by our subjects. Central and peripheral markers, thus, reflect viewer ratings and can predict them to a certain extent, as we showed using machine learning.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. All the code for processing and machine learning has been uploaded to https://github.com/avenator/EEG-FILMS.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board of HSE University (No. 52, 14.01.2019). The patients/participants provided their written informed consent to participate in this study.

Author contributions

VK: design, analysis, and writing. DS: recording and writing. NR: analysis. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the International Laboratory of Social Neurobiology ICN HSE RF Government Grant Ag. No. 075–15–2022-1037 and was carried out using Unique Scientific Installation of the National Research University Higher School of Economics “Automated System for Non-Invasive Brain Stimulation with the Possibility of Synchronous Registration of Brain Biocurrents and Eye Movement Tracking”.

Acknowledgments

The authors thank the research assistant Kirill Makas for stimulus selection and running pilot subjects.

Conflict of interest

Publisher’s note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1148205/full#supplementary-material

References

Ahmad, J., Duraisamy, P., Yousef, A., and Buckles, B. (2017). “Movie success prediction using data mining” in *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (IEEE), 1–4.

Apicella, A., Arpaia, P., Frosolone, M., Improta, G., Moccaldi, N., and Pollastro, A. (2022). EEG-based measurement system for monitoring student engagement in learning 4.0. *Sci. Rep.* 12, 5857–5813. doi: 10.1038/s41598-022-09578-y

Aznar, J. S., Molina, J. J. C., Paredes, C. A., and Gómez, L. S. (2019). “The spectator facing the cut: a Neurocinematics review” in *The European Conference on Media, Communication & Film*

Barnett, S. B., and Cerf, M. (2017). A ticket for your thoughts: method for predicting movie trailer recall and future ticket sales using neural similarity among moviegoers. *J. Consum. Res.* 44, 160–181. doi: 10.1093/jcr/ucw083

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., et al. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* 78, B231–B244.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K., and Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.* 9:16. doi: 10.3389/fninf.2015.00016

Boksem, M. A., and Smidts, A. (2015). Brain responses to movie trailers predict individual preferences for movies and their population-wide commercial success. *J. Mark. Res.* 52, 482–492. doi: 10.1509/jmr.13.0572

Bradley, M. M., Codispoti, M., Cuthbert, B. N., and Lang, P. J. (2001). Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion* 1, 276–298. doi: 10.1037/1528-3542.1.3.276

Bullack, A., Büdenbender, N., Roden, I., and Kreutz, G. (2018). Psychophysiological responses to “happy” and “sad” music. *Music. Percept.* 35, 502–517. doi: 10.1525/mp.2018.35.4.502

Chaouachi, M., Pierre, C., Jraidi, I., and Frasson, C. (2010). “Affect and mental engagement: towards adaptability for intelligent” in *Proceedings of the 23rd international FLAIRS conference* (Daytona Beach: AAAI Press).

Christoforou, C., Papadopoulos, T. C., Constantinidou, F., and Theodorou, M. (2017). Your brain on the movies: a computational approach for predicting box-office performance from viewer’s brain responses to movie trailers. *Front. Neuroinform.* 11:72. doi: 10.3389/fninf.2017.00072

Coelli, S., Sclocco, R., Barbieri, R., Reni, G., Zucca, C., and Bianchi, A. M. (2015). EEG-based index for engagement level monitoring during sustained attention. *Ann. Int. Conf. EEE Eng. Med. Biol. Soc.*, 1512–1515. doi: 10.1109/embc.2015.7318658

Cohen, M. X., Elger, C. E., and Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *NeuroImage* 35, 968–978. doi: 10.1016/j.neuroimage.2006.11.056

Davidson, R. J. (2004). What does the prefrontal cortex “do” in affect: perspectives on frontal EEG asymmetry research. *Biol. Psychol.* 67, 219–234. doi: 10.1016/j.biopsycho.2004.03.008

Dellarocas, C., Awad, N., and Zhang, M. (2005). *Using online ratings as a proxy of word-of-mouth in motion picture revenue forecasting* Smith School of Business, University of Maryland.

Dhir, R., and Raj, A. (2018). “Movie success prediction using machine learning algorithms and their comparison” in *2018 first International Conference on Secure Cyber Computing and Communication (ICSCCC)* (IEEE).

Dimberg, U., and Karlsson, B. (1997). Facial reactions to different emotionally relevant stimuli. *Scand. J. Psychol.* 38, 297–303. doi: 10.1111/1467-9450.00039

Dimberg, U. L. F., and Petterson, M. (2000). Facial reactions to happy and angry facial expressions: evidence for right hemisphere dominance. *Psychophysiology* 37, 693–696. doi: 10.1111/1469-8986.3750693

Dushantha, A., Akalanka, R., Gayan, H., Siriwardhana, K. C., Haddela, P. S., and Wickramasinghe, L. (2020). “Evaluation method for video Advertisetments using EEG signals.” in *2020 2nd International Conference on Advancements in Computing (ICAC)* (IEEE)

Freeman, F. G., Mikulka, P. J., Prinzel, L. J., and Scerbo, M. W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biol. Psychol.* 50, 61–76. doi: 10.1016/s0301-0511(99)00002-2

Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for human Electromyographic research. *Psychophysiology* 23, 567–589. doi: 10.1111/j.1469-8986.1986.tb00676.x

Gaenssle, S., Budzinski, O., and Astakhova, D. (2018). Conquering the box office: factors influencing success of international movies in Russia. *Rev. Netw. Econ.* 17, 245–266. doi: 10.1515/rne-2019-0017

Giraldo, S., and Ramirez, R. (2013). "Brain-activity-driven real-time music emotive control" in *The 3rd International Conference on Music & Emotion, Jyväskylä, Finland, June 11–15, 2013* (Department of Music: University of Jyväskylä).

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D. J. (2008). Neurocinematics: the neuroscience of film. *PRO* 2, 1–26. doi: 10.3167/proj.2008.020102

Heimann, K., Umiltà, M. A., Guerra, M., and Gallese, V. (2014). Moving mirrors: a high-density EEG study investigating the effect of camera movements on motor cortex activation during action observation. *J. Cognit. Neurosci.* 26, 2087–2101. doi: 10.1162/jocn_a_00602

Jasper, H. H. (1958). Report of the committee on methods of clinical examination in electroencephalography: 1957. *Electroencephalogr. Clin. Neurophysiol.* 10, 370–375.

Kramer, O. (2016). Scikit-learn. *Stud. Big Data* 20, 45–53. doi: 10.1007/978-3-319-33383-0_5

Kuoppa, P., Tarvainen, M. P., Karhunen, L., and Narvainen, J. (2016). "Heart rate reactivity associated to positive and negative food and non-food visual stimuli" in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE)

Lash, M. T., and Zhao, K. (2016). Early predictions of movie success: the who, what, and when of profitability. *J. Manag. Inf. Syst.* 33, 874–903. doi: 10.1080/07421222.2016.1243969

Latif, M. H., and Afzal, H. (2016). Prediction of movies popularity using machine learning techniques. *Int. J. Comp. Sci. Netw. Secur.* 16:127.

Lau, Z. J., Pham, T., Chen, S. A., and Makowski, D. (2022). Brain entropy, fractal dimensions and predictability: A review of complexity measures for EEG in healthy and neuropsychiatric populations. *Eur. J. Neurosci.* 56, 5047–5069. doi: 10.1111/ejn.15800

Lelis-Torres, N., Ugrinowitsch, H., Apolinário-Souza, T., Benda, R. N., and Lage, G. M. (2017). Task engagement and mental workload involved in variation and repetition of a motor skill. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-15343-3

Li, S., Walters, G., Packer, J., and Scott, N. (2018). Using skin conductance and facial electromyography to measure emotional responses to tourism advertising. *Curr. Issues Tour.* 21, 1761–1783. doi: 10.1080/13683500.2016.1223023

Liu, S., Lv, J., Hou, Y., Shoemaker, T., Dong, Q., Li, K., et al. (2016). "What makes a good movie trailer?" in *Proceedings of the 24th ACM international conference on Multimedia* (ACM)

Lundberg, S. M., and Lee, S. I. (2017). A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 30, 1–10.

Lundqvist, L. O. (1995). Facial EMG reactions to facial expressions: a case of facial emotional contagion? *Scand. J. Psychol.* 36, 130–141. doi: 10.1111/j.1467-9450.1995.tb00974.x

Madsen, J., and Parra, L. C. (2022). Cognitive processing of a common stimulus synchronizes brains, hearts, and eyes. *PNAS Nexus* 1:pgac020. doi: 10.1093/pnasnexus/pgac020

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., et al. (2021). NeuroKit2: a python toolbox for neurophysiological signal processing. *Behav. Res. Methods* 53, 1689–1696. doi: 10.3758/s13428-020-01516-y

Maran, T., Sachse, P., Martini, M., Weber, B., Pinggera, J., Zuggal, S., et al. (2017). Lost in time and space: states of high arousal disrupt implicit acquisition of spatial and sequential context information. *Front. Behav. Neurosci.* 11:206. doi: 10.3389/fnbeh.2017.00206

Martin, G. N. (2019). (why) do you like scary movies? A review of the empirical research on psychological responses to horror films. *Front. Psychol.* 10:2298. doi: 10.3389/fpsyg.2019.02298

McMahan, T., Parberry, I., and Parsons, T. D. (2015). Evaluating player task engagement and arousal using electroencephalography. *Procedia Manuf.* 3, 2303–2310. doi: 10.1016/j.promfg.2015.07.376

Molteni, E., Bianchi, A. M., Butti, M., Reni, G., and Zucca, C. (2007). Analysis of the dynamical behaviour of the EEG rhythms during a test of sustained attention. *Ann. Int. Conf. IEEE Eng. Med Biol. Soc.* 2007, 1298–1301. doi: 10.1109/iembs.2007.4352535

Moon, S., Bergey, P. K., and Iacobucci, D. (2010). Dynamic effects among movie ratings, movie revenues, and viewer satisfaction. *J. Mark.* 74, 108–121. doi: 10.1509/jmkg.74.1.108

Motion Picture Association (2022). *2021 THEME report - motion picture association*. Available at: https://www.motionpictures.org/research-docs/2021-theme-report/

Nembrini, S., König, I. R., and Wright, M. N. (2018). The revival of the Gini importance? *Bioinformatics* 34, 3711–3718. doi: 10.1093/bioinformatics/bty373

Nilashi, M., Samad, S., Ahmadi, N., Ahani, A., Abumalloh, R. A., Asadi, S., et al. (2020). Neuromarketing: a review of research and implications for marketing. *J. Soft Comput. Decis. Support Syst.* 7, 23–31.

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., and Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proc. Natl. Acad. Sci.* 109, 9599–9604. doi: 10.1073/pnas.1206095109

Oliver, M. B. (1993). Exploring the paradox of the enjoyment of sad films. *Hum. Commun. Res.* 19, 315–342. doi: 10.1111/j.1468-2958.1993.tb00304.x

Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., et al. (2017). Interpersonal autonomic physiology: a systematic review of the literature. *Personal. Soc. Psychol. Rev.* 21, 99–141. doi: 10.1177/1088868316628405

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* 2:10. doi: 10.3389/neuro.11.010.2008

Pope, A. T., Bogart, E. H., and Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biol. Psychol.* 40, 187–195. doi: 10.1016/0301-0511(95)05116-3

Pope, K. J., Lewis, T. W., Fitzgibbon, S. P., Janani, A. S., Grummett, T. S., Williams, P. A., et al. (2022). Managing electromyogram contamination in scalp recordings: an approach identifying reliable beta and gamma EEG features of psychoses or other disorders. *Brain Behav.* 12:e2721. doi: 10.1002/brb3.2721

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* 31, 1–11.

Rahman, M. M., Sarkar, A. K., Hossain, M. A., Hossain, M. S., Islam, M. R., Hossain, M. B., et al. (2021). Recognition of human emotions using EEG signals: a review. *Comput. Biol. Med.* 136:104696. doi: 10.1016/j.compbiomed.2021.104696

Rayatdoost, S., Rudrauf, D., and Soleymani, M. (2020). "Multimodal gated information fusion for emotion recognition from EEG signals and facial behaviors" in *Proceedings of the 2020 International Conference on Multimodal Interaction*, 655–659.

Rayatdoost, S., and Soleymani, M. (2018). "Cross-corpus EEG-based emotion recognition" in *2018 IEEE 28th international workshop on machine learning for signal processing (MLSP)* (IEEE), 1–6.

Sato, W., Kochiyama, T., and Yoshikawa, S. (2020). Physiological correlates of subjective emotional valence and arousal dynamics while viewing films. *Biol. Psychol.* 157:107974. doi: 10.1016/j.biopsycho.2020.107974

Shestyuk, A. Y., Kasinathan, K., Karapoondinott, V., Knight, R. T., and Gurumoorthy, R. (2019). Individual EEG measures of attention, memory, and motivation predict population level TV viewership and twitter engagement. *PLoS One* 14:e0214507. doi: 10.1371/journal.pone.0214507

Shetty, C., Khan, A., Singh, T., and Kharatmol, K. (2021). "Movie review prediction system by real time analysis of facial expression" in *2021 6th International Conference on Communication and Electronics Systems (ICCES)* (IEEE)

Shi, H., Yang, L., Zhao, L., Su, Z., Mao, X., Zhang, L., et al. (2017). Differences of heart rate variability between happiness and sadness emotion states: a pilot study. *J. Med. Biol. Eng.* 37, 527–539. doi: 10.1007/s40846-017-0238-0

Soleymani, M., Asghari-Esfeden, S., Fu, Y., and Pantic, M. (2015). Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Trans. Affect. Comput.* 7, 17–28. doi: 10.1109/TAFFC.2015.2436926

Topal, K., and Ozsoyoglu, G. (2016). "Movie review analysis: emotion analysis of IMDB movie reviews" in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (IEEE).

Vallat, R., and Walker, M. P. (2021). An open-source, high-performance tool for automated sleep staging. *Eife* 10:e70092. doi: 10.7554/elife.70092

Van Buuren, S., and Groothuis-Oudshoorn, K. (2011). Mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* 45, 1–67.

Van Gent, P., Farah, H., Van Nes, N., and Van Arem, B. (2019). Analysing noisy driver physiology real-time using off-the-shelf sensors: heart rate analysis software from the taking the fast lane project. *J. Open Res. Soft.* 7:32. doi: 10.5334/jors.241

Vecchiato, G., Astolfi, L., De Vico Fallani, F., Cincotti, F., Mattia, D., Salinari, S., et al. (2009). Changes in brain activity during the observation of TV commercials by using EEG, GSR and HR measurements. *Brain Topogr.* 23, 165–179. doi: 10.1007/s10548-009-0127-0

Vecchiato, G., Toppi, J., Astolfi, L., De Vico Fallani, F., Cincotti, F., Mattia, D., et al. (2011). Spectral EEG frontal asymmetries correlate with the experienced pleasantness of TV commercial advertisements. *Med. Biol. Eng. Comput.* 49, 579–583. doi: 10.1007/s11517-011-0747-x

Walter, C., Rosenstiel, W., Bogdan, M., Gerjets, P., and Spüler, M. (2017). Online EEG-based workload adaptation of an arithmetic learning environment. *Front. Hum. Neurosci.* 11:286. doi: 10.3389/fnhum.2017.00286

Wu, Q., Zhao, W., and Jacopo, T. (2018). "Towards objective assessment of movie trailer quality using human electroencephalogram and facial recognition" in *2018 IEEE International Conference on Electro/Information Technology (EIT)* (IEEE), 0449–0452.

Wu, Q., Zhao, W., Sharma, H., and Qiu, T. (2017). "Movie trailer quality evaluation using real-time human electroencephalogram" in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (IEEE), 3553–3557.

Zhou, R., and Hu, S. (2004). Effects of viewing pleasant and unpleasant photographs on facial EMG asymmetry. *Percept. Mot. Skills* 99, 1157–1167. doi: 10.2466/PMS.99.7.1157-1167

Zhu, J., Ji, L., and Liu, C. (2019). Heart rate variability monitoring for emotion and disorders of emotion. *Physiol. Meas.* 40:064004. doi: 10.1088/1361-6579/ab1887

Zillmann, D. (1996). "The psychology of suspense in dramatic exposition," in *Suspense: conceptualizations, theoretical analyses, and empirical explorations*. eds. P. Vorderer, H. J. Wulff and M. Friedrichsen (New York: Routledge).

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 26 June 2023
DOI 10.3389/fnins.2023.1204809

OPEN ACCESS

EDITED BY
Xin Li,
West Virginia University, United States

REVIEWED BY
Annalisa Tosoni,
University of Studies G. d'Annunzio Chieti and Pescara, Italy
Guy Cheron,
Universitélibre de Bruxelles, Belgium

*CORRESPONDENCE
Celia Andreu-Sánchez
✉ celia.andreu@uab.cat

RECEIVED 12 April 2023
ACCEPTED 12 June 2023
PUBLISHED 26 June 2023

CITATION
Andreu-Sánchez C, Martín-Pascual MÁ , Gruart A and Delgado-García JM (2023) Beta-band differences in primary motor cortex between media and non-media professionals when watching motor actions in movies. *Front. Neurosci.* 17:1204809.
doi: 10.3389/fnins.2023.1204809

Beta-band differences in primary motor cortex between media and non-media professionals when watching motor actions in movies

Celia Andreu-Sánchez[1]*, Miguel Ángel Martín-Pascual[1,2], Agnès Gruart[3] and José María Delgado-García[3]

[1]Neuro-Com Research Group, Department of Audiovisual Communication and Advertising, Universitat Autònoma de Barcelona, Barcelona, Spain, [2]Research and Innovation, Institute of Spanish Public Television (RTVE), Corporación Radio Televisión Española, Barcelona, Spain, [3]Division of Neurosciences, University Pablo de Olavide, Sevilla, Spain

To watch a person doing an activity has an impact on the viewer. In fact, the film industry hinges on viewers looking at characters doing all sorts of narrative activities. From previous works, we know that media and non-media professionals perceive differently audiovisuals with cuts. Media professionals present a lower eye-blink rate, a lower activity in frontal and central cortical areas, and a more organized functional brain connectivity when watching audiovisual cuts. Here, we aimed to determine how audiovisuals with no formal interruptions such as cuts were perceived by media and non-media professionals. Moreover, we wondered how motor actions of characters in films would have an impact on the brain activities of the two groups of observers. We presented a narrative with 24 motor actions in a one-shot movie in wide shot with no cuts to 40 participants. We recorded the electroencephalographic (EEG) activity of the participants and analyzed it for the periods corresponding to the 24 motor actions (24 actions×40 participants=960 potential trials). In accordance with collected results, we observed differences in the EEG activity of the left primary motor cortex. A spectral analysis of recorded EEG traces indicated the presence of significant differences in the beta band between the two groups after the onset of the motor activities, while no such differences were found in the alpha band. We concluded that media expertise is related with the beta band identified in the EEG activity of the left primary motor cortex and the observation of motor actions in videos.

KEYWORDS

movies, motor action, beta band, brain activity, neurocinematics

1. Introduction

1.1. Looking at motor actions

We perceive various narrative contents in plenty of motor actions through the day, and the observation of those actions has an impact on our brain activity (Muthukumaraswamy et al., 2004). It was more than half a century ago that an "arch rhythm" (with spectral peaks at 10 and 20 Hz) was found when looking at motor actions in movies (Gastaut et al., 1952; Cohen-Séat et al., 1954; Gastaut and Bert, 1954). One of the motor actions most studied in recent neuroscience has been grasping (Castiello, 2005). In this regard, it has been reported that human electroencephalographic mu rhythm (~8–13 Hz) changes while observing other people doing motor activities such as grasping, holding, and tearing (di Pellegrino et al., 1992; Gallese et al., 1996; Muthukumaraswamy et al., 2004). Moreover, previous experience in performing specific tasks seems to be more influential on this mu rhythm than observation of the task itself (Cannon et al., 2014). Other brain rhythms have also been studied in relation to grasping and action movements, such as alpha (~8–12 Hz; Perry and Bentin, 2010) or beta (~13–35 Hz) rhythms (Zaepffel et al., 2013; Khanna and Carmena, 2015). Several studies have worked on revealing neural correlates of grasping and other hand movements when doing them (Chavarriaga et al., 2018), when imaging them (Neuper et al., 2005; Ying et al., 2017), or when looking at someone doing them (Perry and Bentin, 2009). There are also studies that have compared activity modulation of brain EEG recordings while producing or observing social actions (Liao et al., 2015). Most of the studies about visual perception of motor activities paid attention to the primary motor cortex, suggesting a research interest in how a perceiver's brain activity is modulated in this area by the motor activity developed by the actor of the content. In the present study, we recorded the EEG activity in the primary motor cortex of media and non-media professionals watching a one-shot video showing different motor actions of an actor participating in the movie.

1.2. Professional expertise

Professionalization has been shown to be a relevant element when motor brain activity is the subject of study. For instance, professional athletes learn complex dynamic visual scenes better than non-athletes do (Faubert, 2013), and professional racing-car drivers show an increased neural efficiency in brain circuits as compared with naïve drivers (Bernardi et al., 2013). Furthermore, evidence has been found regarding music expertise, including that brain structures differ between musicians and non-musicians (Gaser and Schlaug, 2003), piano players seem to need more-reduced neuronal networks than control subjects to activate the same movements (Krings et al., 2000), musical training has been associated with an altered processing of negative emotions (Park et al., 2014), and professional musicians show more-focused cerebral activations in the contralateral primary sensorimotor cortex (Lotze et al., 2003). There are also studies regarding the impact, in terms of brain activity, of esthetics expertise (Kirk et al., 2008), baseball expertise (Muraskin et al., 2016), football professionalization (Brockhoff et al., 2011), or dancing professionalization (Calvo-Merino et al., 2005), among others. These studies are focused on the neuroscience of expertise, in a context where the performance of an activity becomes more efficient and automatic, proving a perceptual expertise that relies on information from the senses (Bilalić, 2017), and showing how it is of interest to keep researching on cognitive and motor expertise in order to learn how we can improve motor actions.

We have previously studied the impact of media professionalization in visual perception of movies. First, we studied the spontaneous blink rate (SBR) in media professionals and non-media professionals while they were watching movies and looking at theatrical narrative performances, and we found a significant inhibition of SBR in the professional group (Andreu-Sánchez et al., 2017). We found that—since SBR is inversely linked to attention (lower SBR correlates with higher attentional level)—media expertise evokes a higher attention to narratives in both on-screen and live performances. In addition, we checked that media professionalization impacts cognitive neurodynamics during audiovisual cuts: while cuts in movies trigger similar activation of visual cortex, differences are found in central and frontal cortical areas, with a lower activity among media professionals (Andreu-Sánchez et al., 2021). Moreover, after the new visual information that cuts bring to spectators, effective brain connectivity is more organized in media professionals than in non-media professionals (Andreu-Sánchez et al., 2021).

In the present study, we wondered how media professionals' brain activity in primary motor cortex would differ from that of non-media professionals when they were watching motor actions on screens, in a video without cuts that could interfere in the perceptive process.

2. Materials and methods

2.1. Participants

Forty participants aged 28–56 (43.75 ± 7.837) took part in this study. The group of media professionals ($N = 20$) comprised 15 males and five females. Their mean age was 44.25 ± 7.196 years. The time spent in their media professions was 20.2 ± 8.637 years. The group of non-media professionals ($N = 20$) consisted of 16 males and four females. Their mean age was 43.25 ± 8.589 years. The time spent in their non-media professions was 18.85 ± 9.422 years. Inclusion in the media professional group required participants to use video edition and to take decisions related to media editing in their everyday work. Non-media professionals were chosen outside of this criterion. All had normal or corrected-to-normal visual acuity. Subjects did not receive any economic compensation for their participation in this study.

2.2. Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Commission for Research with Animals and Humans (CEEAH) of the Universitat Autònoma de Barcelona (Barcelona, Spain). The participants provided their written informed consent to participate in this study.

2.3. Stimuli and procedure

We created four stimuli with the same narrative but different formats: (1) a one-shot movie in wide shot with no camera movements; (2) a movie edited according to classical rules of edition with smooth transitions; (3) a movie edited breaking classical rules of edition, having a chaotic style with sharp and illogical transitions between shots; and (4) a live performance. All four stimuli were randomly presented to all participants, but, for the purpose of this study, we only analyze the one-shot movie with no cuts. Thus, we isolate the effect of viewing video content without breaks due to editing cuts. The selected movie had a duration of 198 s. The narrative included 24 motor actions that were used as triggers to analyze participants' brain activities. The selected motor actions included objects being grasped, caught, and gripped.

Video stimuli were presented on a 42-in HD Led display (Panasonic TH-42PZ70EA) and participants were placed at 150 cm from the screen. Stimuli were presented with Paradigm Stimulus Presentation (Perception Research System Inc.). Participants were asked to attend to the stimuli, with no further information being given that a follow-up questionnaire would be presented. The questionary presented was actually a distractor without interest for the research.

2.4. Data acquisition

Continuous EEG data were acquired with the help of a wireless system (Enobio, Neuroelectrics), with 20 electrodes placed according to the International 10–20 system [O1, O2, P7, P3, Pz, P4, P8, T7, C3, Cz, C4, T8, F7, F3, Fz, F4, F8, Fp1, Fp2, and an external electrode used for electrooculogram (EOG) recording] referenced to electronically linked mastoid electrodes (see Martín-Pascual et al., 2018 for details). Data were sampled at 500 Hz. We recorded facial expressions of participants with an HD video-camera for contrasting participants' behavior during the sessions, and additionally to detect and to avoid artifacts and unwanted muscle movements.

2.5. Data analysis

Electroencephalographic data were processed using EEGLAB (Delorme and Makeig, 2004) software version 2022.1 running on MATLAB R2022b (The MathWorks Inc.) under a macOS Ventura 13.2.1 (Apple Inc.). We band-passed filtered the data between 0.5 and 40 Hz. We removed the EOG electrode and bad channels when needed. A common average reference was applied. We decomposed data with an ICA analysis (infomax algorithm) and got rid of artifactual components, including eye and muscle activity (Delorme and Makeig, 2004). We made 3-s epochs of 1 s before and 2 s after the onset of the motor activity, marked with triggers at the onset of the motor actions in the video. In total, we had 40 participants who attended 24 motor actions, making 40 × 24, i.e., 960 potential trials of 3 s each. We rejected bad epochs through visual inspection.

We analyzed data recorded from the primary motor cortex of both hemispheres. The left hemisphere was studied with activity from the C3 electrode, and the right hemisphere was studied with activity from the C4 electrode. We computed spectral activity in alpha (8–12 Hz) and beta (13–30 Hz) bands. We also distinguished between activity before (−500–0 ms) and after (0–1,000 ms) the onset of the motor activity. Statistical analysis was performed offline using JASP software (Version 0.17.1, Apple Silicon). We computed unpaired *t*-test or non-parametric Mann–Whitney Rank Sum Tests for each situation. Effect size was computed with rank-biserial correlation. For the normality test, we used the Shapiro–Wilk. We also computed event-related spectral perturbation (ERSP) and power spectrum density (PSD) for those C3 and C4 electrodes and compared them among the groups. We also computed event-related spectral perturbation (ERSP) and power spectrum density (PSD) for those C3 and C4 electrodes and compared them among the groups. ERSP (Delorme and Makeig, 2004) is used to visualize mean event-related changes in spectral power over time in a broad frequency range and generalize the narrow-band event-related desynchronization (ERD) and synchronization (ERS) measures introduced by Pfurtscheller and Aranibar (1979). And PSD helps to know how the strength of a signal is distributed in the frequency domain and its unit is energy per frequency, demonstrating the strength of the variations of a signal as a function of frequency (Valipour et al., 2014). We also plot the inter-trial coherence (ITC). We computed t-tests with a significance of $p < 0.05$ with EEGLAB statistics running on MATLAB.

3. Results

In the case of beta band (13–30 Hz) comparison between media and non-media professionals, although we did not find significant differences before the onset of the motor actions in the left hemisphere (C3 electrode; U = 264, $p = 0.086$, Mann–Whitney test, rank-biserial correlation: 0.320), we did find significant differences after the onset of the motor activities in C3: U = 273, $p = 0.049$, Mann–Whitney test, rank-biserial correlation: 0.365. In contrast, the right hemisphere (C4 electrode), did not show significant differences before (U = 191, $p = 0.820$, Mann–Whitney test, rank-biserial correlation: −0.045) or after the onset of the motor activities (U = 191, $p = 0.820$, Mann–Whitney test, rank-biserial correlation: −0.045; see Figure 1). In addition, we did not find significant differences in the alpha band (8–12 Hz) between groups in either the left hemisphere (C3 electrode) before (U = 214, $p = 0.718$, Mann–Whitney test, rank-biserial correlation: 0.070) and after (U = 219, $p = 0.620$, Mann–Whitney test, rank-biserial correlation: 0.095) the onset of the motor actions, or the right hemisphere (C4 electrode) before (U = 164, $p = 0.341$, Mann–Whitney test, rank-biserial correlation: −0.180) and after (U = 171, $p = 0.341$, Mann–Whitney test, rank-biserial correlation: −0.145) the onset of the motor actions.

When analyzing and comparing the ERSPs in C3 and C4 between the two groups, we found significant differences in the left hemisphere. Note that all trials included not a change of visual content but the onset of a motor action of the hand(s) of the character in scene (such as grasping an object), meaning that the visual presentation was in an organic flow, without any formal visual interruption, such as cut or flash. Media professionals decreased their activity around C3 significantly as compared with non-media professionals. Those differences were not found in the right hemisphere (see Figure 2). We also computed ERSPs in media and non-media professional groups, with an average of all electrodes, in alpha (8–12 Hz) and beta (13–30 Hz) bands, and compared them using a *t*-test to look for differences. No significant differences ($p < 0.05$) between groups were found in the alpha band. In the beta band, we found significant differences in motor cortex areas in the left hemisphere (around C3) but none in the right hemisphere (Figures 3, 4). When looking at differences in the PSD [Log Power $10*\log_{10}(\mu V^2)$] at C3 and C4, we also found statistically significant differences in the left hemisphere in the beta band—the group of non-media professionals showing a higher spectral power (Figure 5).

4. Discussion

Watching other people performing different motor actions modulates our brain processes. In part, films are based on the impact that artists' and creators' actions have on spectators. In recent years, neuroscientists have studied how looking at someone performing a

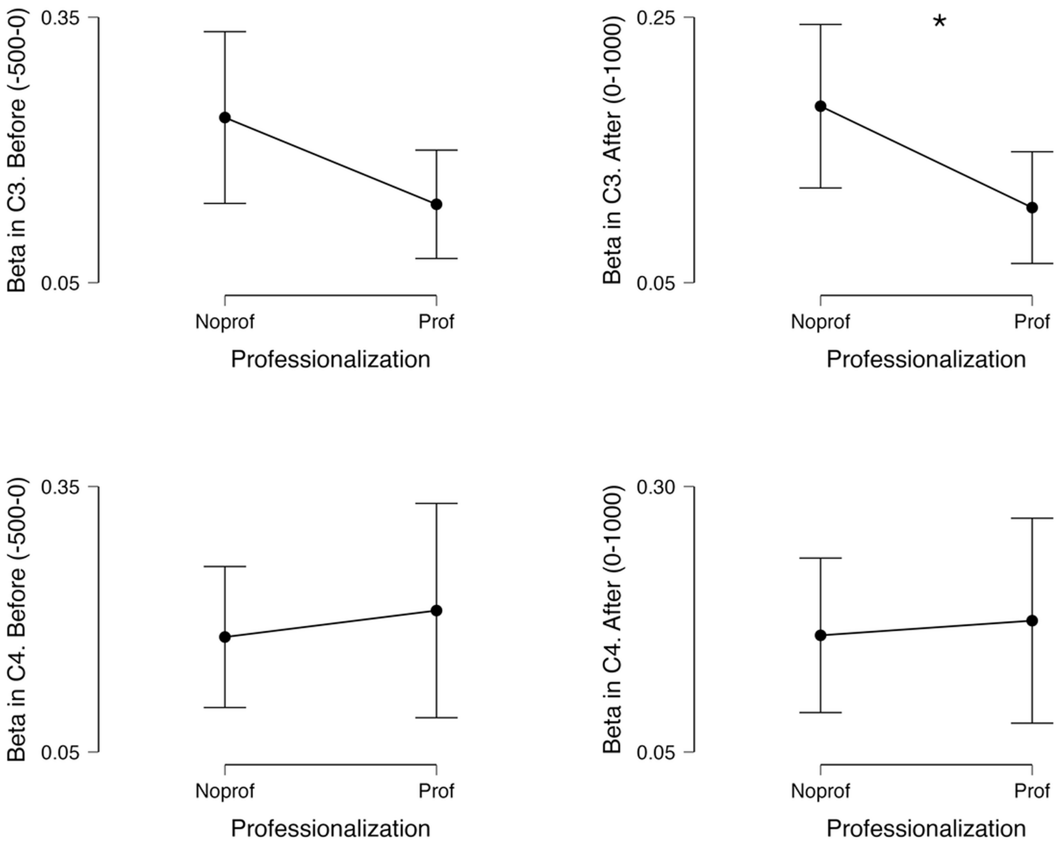


FIGURE 1
Mean spectral power in the beta band in media and non-media professionals, before the onset of the motor activities within the video (–500 to 0ms) and after the onset of the motor activities (0–1,000ms), in left (C3) and right (C4) hemispheres. *indicates $p<0.05$, non-parametric Mann–Whitney test.

motor action, such as grasping an object, modifies the brain activity of the observer (Babiloni et al., 2002). In fact, these changes in EEG activities have been linked to the mirror neuron system on several occasions (Muthukumaraswamy et al., 2004; Rizzolatti and Craighero, 2004; Perry and Bentin, 2009, 2010; Rizzolatti and Cattaneo, 2009; Marshall and Meltzoff, 2011; Ikeda et al., 2019). Moreover, some studies have proven that the imagination of a motor action also activates specific, complex brain networks (Cebolla et al., 2015, 2017). So far, these studies have been carried out in research centers around the world without, to our knowledge, paying any further attention to the media expertise of the perceiver. Here we proposed to analyze differences in brain activity when seeing someone doing motor activities in movies, based on whether the viewer had or did not have media professional expertise. Overall, we found significant differences in the left primary motor cortex (C3) in beta rhythm between the two groups, with a lower activity present in media professionals. In previous studies, beta oscillations have been correlated with planning and execution of grasping movements (Jasper and Penfield, 1949; Zaepffel et al., 2013; Khanna and Carmena, 2015), with a decrease in the spectral power of the oscillations during the preparation and the execution of voluntary movements. Beta-band desynchronization during motor preparation has been connected with the degree of uncertainty about the task (Tzagarakis et al., 2010). Here, we found that professional expertise of viewers when looking at someone doing motor actions impacts on that viewing with a decrease in the spectral power of the beta band as compared with the case of non-media professionals. It is difficult to understand the neural processes underlying these functional differences, but we think that perhaps media professionals are more impacted by the grasping actions since they might be more attentive to the narrative content, as we previously found (Andreu-Sánchez et al., 2017). It is true that the notable desynchronization in the motor area of the professionals (Figure 3) could be due to a more real sensation in this group in the narrative events of the movie. However, such desynchronization is also perceived prior to the onset of grasping actions. In fact, we have previously found that cuts in movies also provoke differences therein among both groups, showing a higher desynchronization in motor cortex in media professionals between 7 and 11 Hz at 200–300 ms after the cut (Andreu-Sánchez et al., 2021). In this regard, the reported results here would point to some important functional differences in the viewing of screens by media professionals. Although we initially found significant differences between media and non-media professionals in eye-blink rate when watching audiovisuals (Andreu-Sánchez et al., 2017), while studying differences between the two groups when they were looking at new visual information presented after audiovisual cuts, we found significant differences in frontal and parietal brain areas, but not in the occipital one (Andreu-Sánchez et al., 2021), suggesting that differences between these two groups might be more linked to the processing of the narrative content than to the actual visual processing of the formal visual information. Since

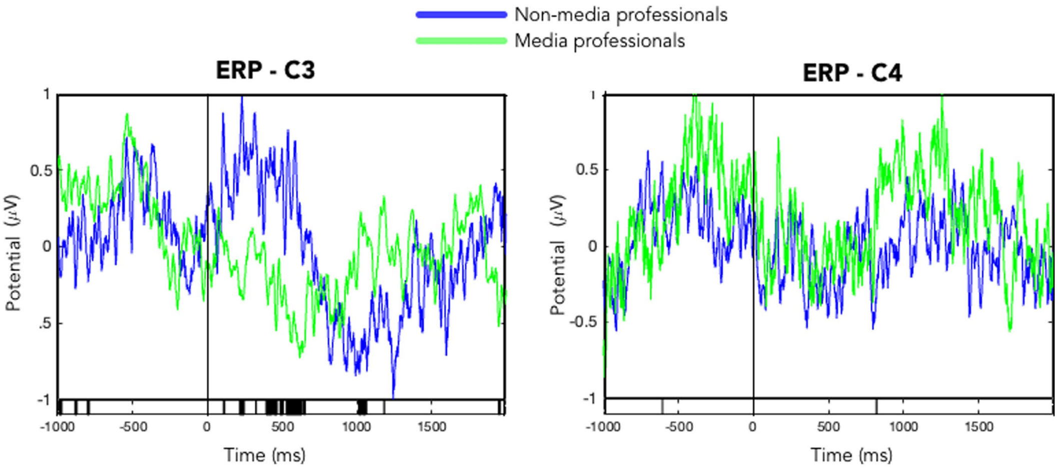


FIGURE 2
Average ERSPs of C3 and C4 in media (green) and non-media (blue) professionals, while watching 24 motor actions in the video. Vertical black lines (at Time 0) indicate the onset of the motor action of the character in the video. The lower vertical black bars show significant differences between groups across time (*t*-test, $p<0.05$).

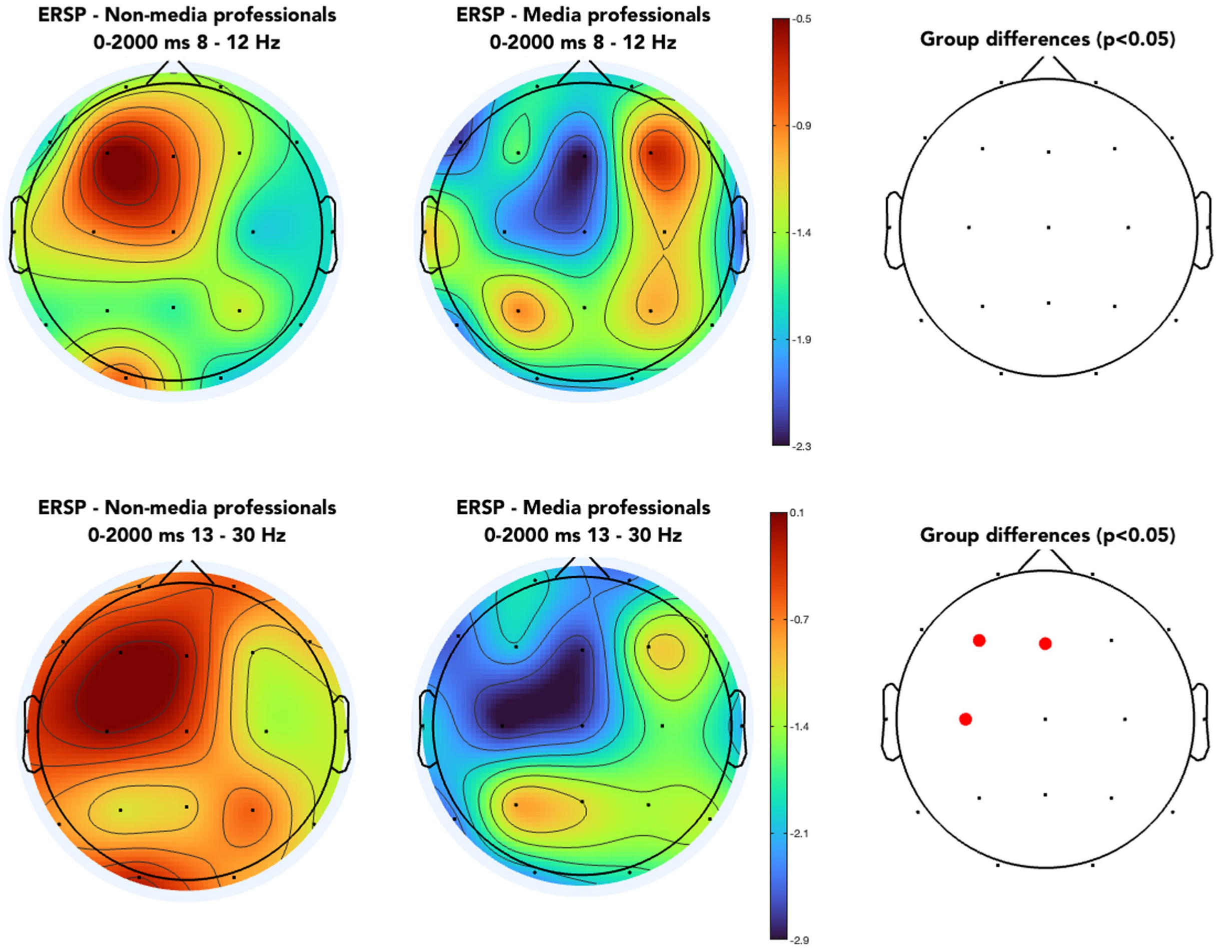


FIGURE 3
Average ERSPs of all electrodes in media and non-media professionals, from the onset of the motor activity to 2,000ms after, in alpha (upper) and beta (lower). Red dots indicate significant differences found between groups ($p<0.05$, *t*-test).

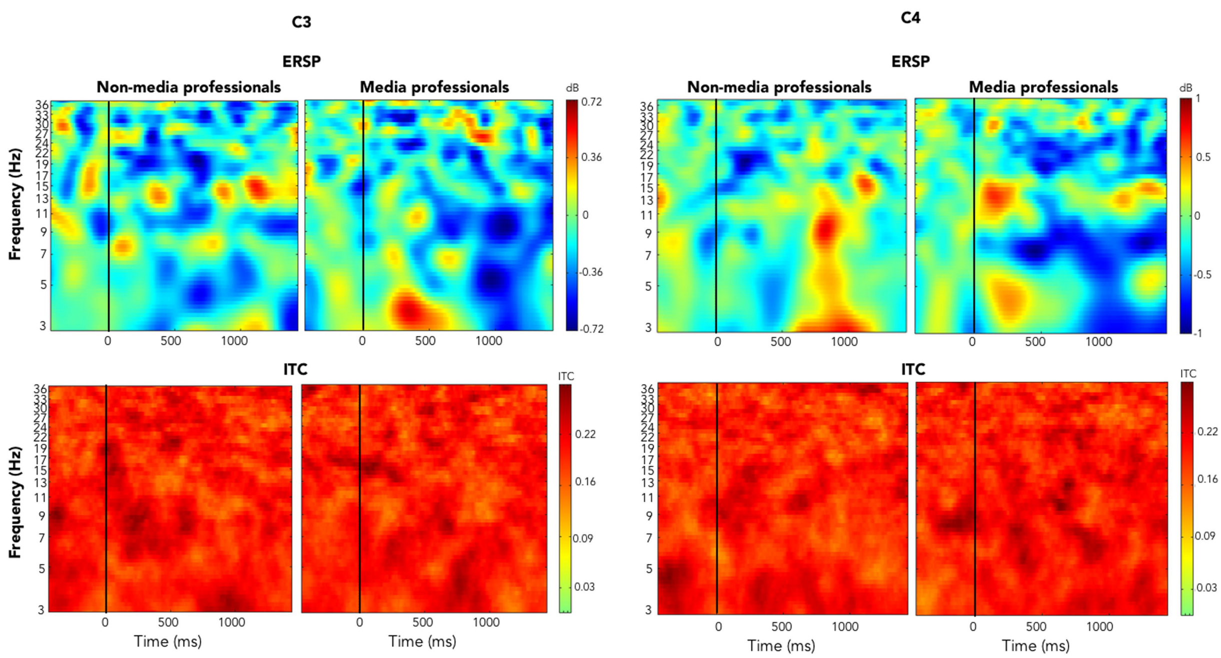


FIGURE 4
ERSPs (above) and ITC (below) in C3 and C4 in media and non-media professionals showing the temporal evolution. Vertical lines indicate the onset of the motor activity.

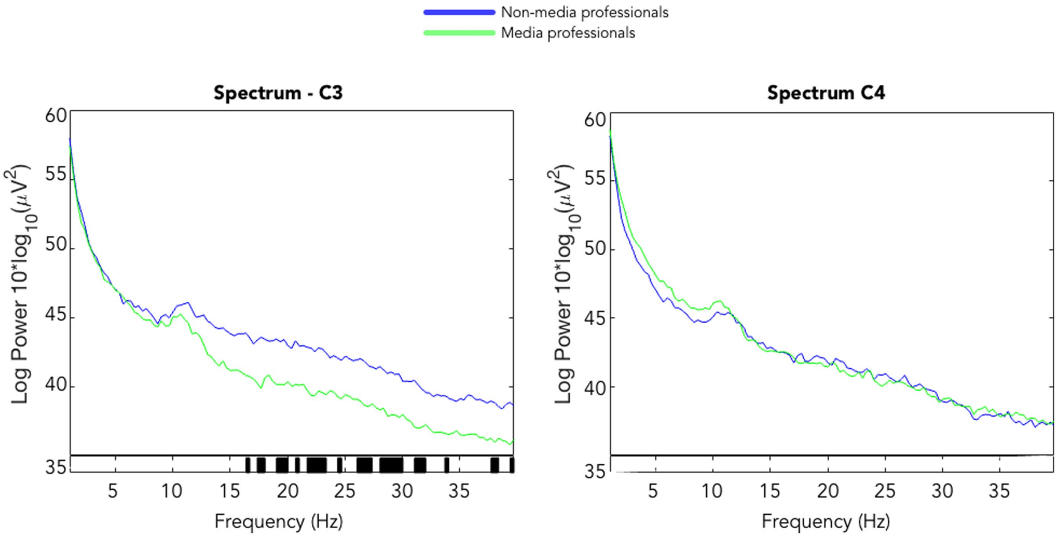


FIGURE 5
PSD of C3 and C4 activity in media (green) and non-media (blue) professionals in the whole 3-s epochs around the motor actions within the video stimulus. Significant differences (p<0.05, t-test) between groups are shown in black on the x-axis.

everything points to a professionalizing effect in media professionals, it would be interesting to replicate previous studies regarding motor imagery paying attention to media professionalization as a variable (Cebolla et al., 2015) as it could have a big impact in brain-computer interface (BCI) training contexts. We also found sharp asymmetry (Figure 3) in alpha and beta bands, regardless the group. It coincides with previous works that suggest contralateral activity in human motor cortex correlated with the hand dominance, specialization, and activation (Hund-Georgiadis and von Cramon, 1999; Bai et al., 2005). Unfortunately, we did not ask participants regarding their hand dominance, which is a limitation of this work and something that could have improved the analysis of our results. Another limitation of our study is the unbalanced male–female sample which prevents us from analyzing sex as a solid variable here.

Our results could also be seen from a perception-action perspective. Perception-action approaches suggest that one of the most important aspects of motor control is predictive control (von Hofsten and Rosander, 2012) and it is based on experience

(Ridderinkhof, 2014). Somehow, our brains use stored memories to constantly make predictions about what we see, feel, and hear (Hawkins and Blakeslee, 2004) and perception and actions would be unified through common principles (Ridderinkhof, 2014). In this context, the skill level has been previously correlated with perception processes linked with anticipatory tasks (Farrow and Abernethy, 2003). Here, we found that the professional experience (or expertise) in audiovisuals affects the brain activity in motor cortex while looking at motor actions. This suggests that the predictive control while viewing actions within the narrative contents could be trained by using audiovisuals as media professionals do on their daily basis.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Commission for Research with Animals and Humans (CEEAH) of the Universitat Autònoma de Barcelona (Barcelona, Spain). The patients/participants provided their written informed consent to participate in this study.

Author contributions

CA-S, MM-P, AG, and JD-G conducted the experimental design and wrote the article. CA-S and MM-P carried out experiments and data analyses. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by grant PID2021-122446NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe" to AG and JD-G.

Acknowledgments

We thank participants in this study. We also acknowledge the Institute of Spanish Public Television (IRTVE) for support and facilities; Bruno Tagliaferri for technical help during the design of stimulus presentation; Andrés Mora for experimental support when testing the protocols; Juan Antonio Gamero for support in the edition of movie stimulus; and Santiago Gómez for constructing the experimental set-up for the theatrical performance. We also thank Roger Churchill for his help in the final edition of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Andreu-Sánchez, C., Martín-Pascual, M. Á., Gruart, A., and Delgado-García, J. M. (2017). Looking at reality versus watching screens: Media professionalization effects on the spontaneous eyeblink rate. *PLoS One* 12:e0176030. doi: 10.1371/journal.pone.0176030

Andreu-Sánchez, C., Martín-Pascual, M. Á., Gruart, A., and Delgado-García, J. M. (2021). The Effect of Media Professionalization on Cognitive Neurodynamics During Audiovisual Cuts. *Front. Syst. Neurosci.* 1:598383. doi: 10.3389/fnsys.2021.598383

Babiloni, C., Babiloni, F., Carducci, F., Cincotti, F., Cocozza, G., Del Percio, C., et al. (2002). Human cortical electroencephalography (EEG) rhythms during the observation of simple aimless movements: A high-resolution EEG study. *NeuroImage* 17, 559–572. doi: 10.1016/S1053-8119(02)91192-4

Bai, O., Mari, Z., Vorbach, S., and Hallett, M. (2005). Asymmetric spatiotemporal patterns of event-related desynchronization preceding voluntary sequential finger movements: A high-resolution EEG study. *Clin. Neurophysiol.* 116, 1213–1221. doi: 10.1016/j.clinph.2005.01.006

Bernardi, G., Ricciardi, E., Sani, L., Gaglianese, A., Papasogli, A., Ceccarelli, R., et al. (2013). How Skill Expertise Shapes the Brain Functional Architecture: An fMRI Study of Visuo-Spatial and Motor Processing in Professional Racing-Car and Naïve Drivers. *PLoS One* 8, 1–11. doi: 10.1371/journal.pone.0077764

Bilalić, M. (2017). *The Neuroscience of Expertise*. Padstow Cornwall: Cambridge University Press

Brockhoff, A., Huff, M., Maurer, A., and Papenmeier, F. (2011). Seeing the unseen? Illusory causal filling in FIFA referees, players, and novices. *Cogn. Res. Princ. Implic.* 1, 1–12. doi: 10.1186/s41235-016-0008-5

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., and Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cereb. Cortex* 15, 1243–1249. doi: 10.1093/cercor/bhi007

Cannon, E. N., Yoo, K. H., Vanderwert, R. E., Ferrari, P. F., Woodward, A. L., and Fox, N. A. (2014). Action experience, more than observation, influences mu rhythm desynchronization. *PLoS One* 9:e92002. doi: 10.1371/journal.pone.0092002

Castiello, U. (2005). The neuroscience of grasping. *Nat. Rev. Neurosci.* 6, 726–736. doi: 10.1038/nrn1744

Cebolla, A. M., Palmero-Soler, E., Leroy, A., and Cheron, G. (2017). EEG spectral generators involved in motor imagery: A swLORETA study. *Front. Psychol.* 8, 1–16. doi: 10.3389/fpsyg.2017.02133

Cebolla, A. M., Petieau, M., Cevallos, C., Leroy, A., Dan, B., and Cheron, G. (2015). Long-lasting cortical reorganization as the result of motor imagery of throwing a ball in a virtual tennis court. *Front. Psychol.* 6, 1–10. doi: 10.3389/fpsyg.2015.01869

Chavarriaga, R., Pereira, M., Zhang, H., Corbet, T., Leeb, R., and Mill, R. (2018). Human EEG reveals distinct neural correlates of power and precision grasping types. *NeuroImage* 181, 635–644. doi: 10.1016/j.neuroimage.2018.07.055

Cohen-Séat, G., Gastaut, H., and Bert, J. (1954). Modification de l'E.E.G. pendant la projection cinématographique. *Rev. Int. Film.* 5, 7–25.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180. doi: 10.1007/BF00230027

Farrow, D., and Abernethy, B. (2003). Do expertise and the degree of perception—action coupling affect natural anticipatory performance? *Perception* 32, 1127–1139. doi: 10.1068/p3323

Faubert, J. (2013). Professional athletes have extraordinary skills for rapidly learning complex and neutral dynamic visual scenes. *Sci. Rep.* 3:1154. doi: 10.1038/srep01154

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609. doi: 10.1093/brain/119.2.593

Gaser, C., and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *J. Neurosci.* 23, 9240–9245. doi: 10.1523/JNEUROSCI.23-27-09240.2003

Gastaut, H., and Bert, J. (1954). EEG changes during cinematographic presentation (moving picture activation of the EEG). *Electroencephalogr. Clin. Neurophysiol.* 6, 433–444. doi: 10.1016/0013-4694(54)90058-9

Gastaut, H., Terzian, H., and Gastaut, Y. (1952). Etude d'une activité électroencéphalographique méconnue: le rythme rolandique en arceau. *Mars. Med.* 89, 296–310.

Hawkins, J., and Blakeslee, S. (2004). *On Intelligence*. New York: Henry Holt and Company

Hund-Georgiadis, M., and von Cramon, D. Y. (1999). Motor-learning-related changes in piano players and non-musicians revealed by functional magnetic-resonance signals. *Exp. Brain Res.* 125, 417–425. doi: 10.1007/s002210050698

Ikeda, Y., Nishimura, Y., and Higuchi, S. (2019). Effects of the differences in mental states on the mirror system activities when observing hand actions. *J. Physiol. Anthropol.* 38, 1–7. doi: 10.1186/s40101-018-0192-8

Jasper, H., and Penfield, W. (1949). Electrocorticograms in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus. *Arch. Psychiatr. Nervenkr.* 183, 163–174. doi: 10.1007/BF01062488

Khanna, P., and Carmena, J. M. (2015). Neural oscillations: beta band activity across motor networks This review comes from a themed issue on Large-scale recording technology. *Curr. Opin. Neurobiol.* 32, 60–67. doi: 10.1016/j.conb.2014.11.010

Kirk, U., Skov, M., Schram Christensen, M., and Nygaard, N. (2008). Brain correlates of aesthetic expertise: A parametric fMRI study. *Brain Cogn.* 69, 306–315. doi: 10.1016/j.bandc.2008.08.004

Krings, T., Töpper, R., Foltys, H., Erberich, S., Sparing, R., Willmes, K., et al. (2000). Cortical activation patterns during complex motor tasks in piano players and control subjects. A functional magnetic resonance imaging study. *Neurosci. Lett.* 278, 189–193. doi: 10.1016/S0304-3940(99)00930-1

Liao, Y., Akalin, Z., Makeig, S., and Deak, G. (2015). EEG imaging of toddlers during dyadic turn-taking: Mu-rhythm modulation while producing or observing social actions. *NeuroImage* 112, 52–60. doi: 10.1016/j.neuroimage.2015.02.055

Lotze, M., Scheler, G., Tan, H.-R., Braun, C., and Birbaumer, N. (2003). The musician's brain: functional imaging of amateurs and professionals during performance and imagery. *NeuroImage* 20, 1817–1829. doi: 10.1016/j.neuroimage.2003.07.018

Marshall, P. J., and Meltzoff, A. N. (2011). Neural mirroring systems: Exploring the EEG mu rhythm in human infancy. *Dev. Cogn. Neurosci.* 1, 110–123. doi: 10.1016/j.dcn.2010.09.001

Martín-Pascual, M. Á., Andreu-Sánchez, C., Delgado-García, J. M., and Gruart, A. (2018). Using electroencephalography measurements and high-quality video recording for analyzing visual perception of media content. *JoVE* e57321, 1–7. doi: 10.3791/57321

Muraskin, J., Dodhia, S., Lieberman, G., Garcia, J. O., Verstynen, T., Vettel, J. M., et al. (2016). Brain dynamics of post-task resting state are influenced by expertise: Insights from baseball players. *Hum. Brain Mapp.* 37, 4454–4471. doi: 10.1002/hbm.23321

Muthukumaraswamy, S. D., Johnson, B. W., and McNair, N. A. (2004). Mu rhythm modulation during observation of an object-directed grasp. *Cogn. Brain Res.* 19, 195–201. doi: 10.1016/j.cogbrainres.2003.12.001

Neuper, C., Scherer, R., Reiner, M., and Pfurtscheller, G. (2005). Imagery of motor actions: differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. *Brain Res. Cogn. Brain Res.* 25, 668–677. doi: 10.1016/j.cogbrainres.2005.08.014

Park, M., Gutyrchik, E., Bao, Y., Zaytseva, Y., Carl, P., Welker, L., et al. (2014). Differences between musicians and non-musicians in neuro-affective processing of sadness and fear expressed in music. *Neurosci. Lett.* 566, 120–124. doi: 10.1016/j.neulet.2014.02.041

Perry, A., and Bentin, S. (2009). Mirror activity in the human brain while observing hand movements: A comparison between EEG desynchronization in the mu-range and previous fMRI results. *Brain Res.* 1282, 126–132. doi: 10.1016/j.brainres.2009.05.059

Perry, A., and Bentin, S. (2010). Does focusing on hand-grasping intentions modulate electroencephalogram mu and alpha suppressions? *Neuroreport* 21, 1050–1054. doi: 10.1097/WNR.0b013e32833fcb71

Pfurtscheller, G., and Aranibar, A. (1979). Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. *Electroencephalogr. Clin. Neurophysiol.* 46, 138–146. doi: 10.1016/0013-4694(79)90063-4

Ridderinkhof, K. R. (2014). Neurocognitive mechanisms of perception-action coordination: A review and theoretical integration. *Neurosci. Biobehav. Rev.* 46, 3–29. doi: 10.1016/j.neubiorev.2014.05.008

Rizzolatti, G., and Cattaneo, L. (2009). The mirror-neuron system. *Arch. Neurol.* 66, 557–560. doi: 10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230

Tzagarakis, C., Ince, N. F., Leuthold, A. C., and Pellizzer, G. (2010). Beta-band activity during motor planning reflects response uncertainty. *J. Neurosci.* 30, 11270–11277. doi: 10.1523/JNEUROSCI.6026-09.2010

Valipour, S., Shaligram, A. D., and Kulkarni, G. R. (2014). Detection of an alpha rhythm of EEG signal based on EEGLAB. *J. Eng. Res. Appl.* 4, 154–159.

von Hofsten, C., and Rosander, K. (2012). Perception-action in children with ASD. *Front. Integr. Neurosci.* 6, 1–6. doi: 10.3389/fnint.2012.00115

Ying, R., Weisz, J., and Allen, P. K. (2017). Grasping with your brain: a brain-computer interface for fast grasp selection. *Robot. Res.* 2, 325–340. doi: 10.1007/978-3-319-51532-8_20

Zaepffel, M., Trachel, R., Kilavik, B. E., and Brochier, T. (2013). Modulations of EEG beta power during planning and execution of grasping movements. *PLoS One* 8:e60060. doi: 10.1371/journal.pone.0060060

frontiers | Frontiers in Neuroscience

TYPE Hypothesis and Theory
PUBLISHED 03 July 2023
DOI 10.3389/fnins.2023.1160843

Check for updates

OPEN ACCESS

EDITED BY
Miguel Ángel Martín-Pascual,
Spanish Public Television, Spain

REVIEWED BY
Jarkko Toikkanen,
University of Oulu, Finland
Ralf Schmälzle,
Michigan State University, United States
Jane Stadler,
The University of Queensland, Australia

*CORRESPONDENCE
Mehmet Burak Yilmaz
✉ mburak@tlu.ee

RECEIVED 07 February 2023
ACCEPTED 06 June 2023
PUBLISHED 03 July 2023

CITATION
Yilmaz MB, Lotman E, Karjus A and Tikka P (2023) An embodiment of the cinematographer: emotional and perceptual responses to different camera movement techniques. *Front. Neurosci.* 17:1160843.
doi: 10.3389/fnins.2023.1160843

An embodiment of the cinematographer: emotional and perceptual responses to different camera movement techniques

Mehmet Burak Yilmaz[1]*, Elen Lotman[1], Andres Karjus[2] and Pia Tikka[1]

[1]Baltic Film, Media, Arts and Communication School, Tallinn University, Tallinn, Estonia, [2]School of Humanities, Tallinn University, Tallinn, Estonia

We investigate the relationship between camera movement techniques and cognitive responses in audiences, reporting on an experiment exploring the effects of different camera movement methods on viewers' degree of immersion and emotional response. This follows directly from preceding experimental literature and is further motivated by accounts and experiences of practicing cinematographers (authors included), which indicates a correspondence between the two. We designed three different cinematic scenes with indifferent moods, and shot each one time with Steadicam, dolly, handheld, and static camera, resulting in 12 different clips. A total of 44 non-professional participants watched the clips and rated their reactions in terms of arousal and degree of involvement. Experimental results are mixed: movement affects the sense of involvement but not necessarily emotional response. We present and discuss some further explorative results and possible future directions to improve the design. We argue in this contribution that there is value in experimental approaches to cinematography, enabling the systematic study of creative intuitions and audience responses in controlled settings.

KEYWORDS

camera movement, embodied cognition, embodied simulation, cinematography, cinematographer, dolly, Steadicam, handheld

1. Introduction

The cognitive and emotional significance of cinematography is well acknowledged in film studies. However, different techniques of camera movements (e.g., dolly, Steadicam, and handheld) in narrative storytelling are surprisingly sparsely studied topics. Intuitively, audience experiences may vary depending on *how* the camera is moved, and what type of viewpoint it generates for the viewers. Drawing from the tacit knowledge accumulated by cinematographers (including three of the authors), we argue that when producing specific movements, cinematographers may be viewed as *extending* their perceptual bodily senses *via* the camera onto the screen, and in this sense, these movements may be further seen to become *embodied* in the viewer's experience.

Throughout the article, the term *immersion* refers to the viewer's intense mental involvement in the filmic world. In other words, it represents the degree of the feeling of "being there". *Involvement*, in return, is the perceptual and sensational (e.g., valence/arousal) association with onscreen events. *Engagement* describes the act of involvement or immersion itself. We utilize the concept of *embodiment* from Varela et al.'s (1991) perspective which highlights the vital role of the body with its sensorimotor capabilities in the process of cognition. Finally, *experience*, in our case, is an umbrella notion that encompasses the aforementioned terms and can be defined as the totality of one's sensory and motor reactions during movie-watching.

Our study has three objectives. First, to explore how different camera movement techniques impact the audience's emotional engagement with onscreen events in three varying mood situations. We explicitly test the hypothesis that a moving camera elicits stronger responses than a static camera and explore and visualize the outcomes of all the combinations. More broadly, we aim to understand how these techniques influence the degree of immersion (embodiment) of the viewer, and to what extent can theories, such as embodiment, be tested using experimental paradigms. Finally, we also seek to explore improvements in preceding experimental literature by creating more naturalistic, filmic stimuli and discuss their implications. We begin by introducing the theoretical background and cinematographic accounts that motivated the study.

1.1. Theories of embodied camera

It is widely acknowledged that the stature of camera movements is one of the key elements in film production that conveys the story, creates moods of the scenes, and emphasizes the internal struggles of the characters. Revealing the lifelikeness of events on the screen and absorbing the audience into the fictional reality, the moving camera is one of the most potent cinematographic tools that reduce the gap between viewers and the mediated world of fiction (Morgan, 2016).

The first moviegoers in the early 20th century enjoyed the commonly called phantom rides, the effect provided by the moving camera as it traveled on the train tracks and the canals of Venice as if being moved by an invisible phantom (Salt, 2009). By the 1930s, filmmakers explored the possibilities of moving the camera to an extent that a 1932 issue of *American Cinematographer* notes a meeting between directors and cinematographers over redundant execution and the abuse of camera movements (Hall, 1932). In 1930s Hollywood, the moving camera of the cinematographer Karl Freund (American Society of Cinematographers [ASC]) in *The Last Laugh* (dir. F. W. Murnau) "opens the film by descending in an elevator and gliding across a hotel lobby", the scene that Luci Marzola acknowledges in the *Journal of American Cinematographers* as "perhaps the single most talked-about camera technique in the history of motion pictures to that point" (Marzola, 2020, see also in Müller, 2014; Keating, 2019). Another European émigré to Hollywood, Serbian filmmaker Slavko Vorkapich argues that similar to the innate human appreciation of motion, which is already present in newborn babies, adult film viewers must be drawn to and feel "pleasure" by motion suggested by the moving images (Vorkapich, 1930, p. 30). He also linked the perception of motion on the screen directly to the bodily sensory-motor experiences of the spectators: "By merely seeing a motion on the screen, our minds, conscious or subconscious, may be made to react in a similar manner as in active participation" (Vorkapich, 1930, p. 30). Half a century later, Vivian Sobchack appraised the moving camera as perceived "as always meaningfully-directed, as *intentional*: the unifying embodied activity of a human consciousness as it is situated in and inhabits the world" (Sobchack, 1982, p. 317). Christian Metz wrote "because movement is never material but is always visual, to reproduce its appearance is to duplicate its reality" (Metz, 1991, p. 9).

These arguments bear compelling resemblances to the cognitive theories of *embodied mind* (Varela et al.'s, 1991), *embodied simulation* (ES; Gallese, 2007), and their application to the field of cinema (Tikka, 2008; Gallese and Guerra, 2012; Tikka and Kaipainen, 2014). Several functional neuroimaging studies have shown that different film viewers' brain activations may correlate in a time-locked manner when they watch the same films (see, Jääskeläinen et al., 2020, 2021, for reviews). In this regard, cinema has the ability to convert the act of seeing that each individual experiences idiosyncratically into an "intersubjectively extroverted" vision (Sobchack, 2016, p. 64). Many of these studies also provide evidence that the observed events on the screen are simulated, or mirrored, in the viewer's mind (Gallese and Guerra, 2014). According to Sobchack, "cinematic 'language' is grounded in the language of embodied existence" (Sobchack, 1992, p. 13), and the "moving image is not only perceptible, it is also perceptive" (Sobchack, 2016, p. 75).

Recent neurocinematics research on human mental projections of cinematic techniques has proliferated and produced promising results (Zacks, 2010; Heimann et al., 2014, 2019; Kauttonen et al., 2014, 2018; Raz and Hendler, 2014; Kovács et al., 2019). However, to the best of our knowledge, there are only a few cognitive or neuroscientific studies about the effects of camera movements in the context of embodied cognition. Heimann et al. used electroencephalography (EEG) to measure the neural motor responses of viewers who watched camera movements produced by Zoom, dolly, Steadicam, and stationary cameras (Heimann et al., 2014, 2019). Results demonstrated that Steadicam stimulated the corresponding areas of the brain more than the other techniques, meaning that it afforded the strongest bodily engagement. This supports the view that "film's intentionality and subjectivity are also grounded on the viewers' embodied simulation of camera movements, suggesting that the immanence of cinematic subjectivity largely relies on the bodily nature and understanding of film" (Guerra, 2015, p. 153). These first studies bear significance as they produced experimental evidence supporting the argument that different techniques of camera movements alter the degree of embodiment. Guerra suggests following up with "an ES-based approach, deepened with specific experiments on different stylistic solutions, could also explain stylistic changes caused by the constant evolution of viewers' ability to play a role in a virtual world" (Guerra, 2015, p. 151).

However, from our point of view as filmmakers, the video stimuli presented to participants were not aesthetically appealing in terms of cinematographic imagery and lacked certain aspects usually present in fiction films, such as dramatic lighting and narrative qualities. Hence, it might be debatable to which extent the results could satisfactorily explain effects related to viewing camera movements in cinema. We aimed to improve on this ecological validity aspect in this study.

Cognitive–perceptual theories applied to film studies often emphasize the analogy between a moving camera and the human eye (e.g., Bordwell, 1977; Sobchack, 1982; Barker, 2009; Guerra, 2015; Schonig, 2017). The key arguments are 2-fold. First, the viewpoint that certain camera movement techniques (e.g., Steadicam and handheld) depictions of the world are arguably analogous to the dynamic movements of the human eye (Schonig, 2017). Second, such camera movement on the screen appears to be "the closest approximation of muscular movement of the human body" (Barker, 2009, p. 110). Hence, the perceptual implications of the moving camera on the projected images resemble those of the bodily movement of the humans in the actual space. Similar to our innately mobile bodies, moving camera's frames' "articulated and finite boundaries orient and organize the viewing view's perceptual and motor access to—and in—the film's world." (Sobchack, 2016, p. 72). Therefore, "the frame's spatiotemporal coherence and relative constancy (even when the viewing view/viewed view within its bounds is moving) significantly synthesizes the viewing view's sense perception and movement into a particular and unified 'place"' (Sobchack, 2016, p. 72–73).

In this sense, if considered from the cinematographer's point of view, the human eye in the viewfinder of a camera extends its cognitive–perceptual abilities to the camera (technology). When the camera starts moving, this bodily extension not only engages the visual system but also the whole body of the cinematographer. The camera may be said to become an embodied extension of the cinematographer's body and mind. This can be understood in line with the embodied mind views to the body–brain–world interplay, in general, and to human creation in arts, in particular (Vesey, 1965; Gibson, 1966; Lakoff and Johnson, 1999; Grodal, 2002; Tikka, 2008; Coëgnarts, 2017; Coëgnarts and Slugan, 2022). However, it is important to acknowledge that certain mobile frames, such as cranes, drones, and vehicle-mounted shots, do not precisely correspond to human biological and muscular movements. Such movements can be assumed to be exempt from the eye–camera analogy. On the other hand, our gaze or kinesthetic movement does not always have to perfectly match that of the camera to perceive and comprehend the movement. Whether conscious or subconscious, we can still identify the cinematic world and orient ourselves in it by viewing even inhuman movements according to or with the non-anthropocentric dynamics of alternative filmic reality (Sobchack, 2016, p. 86).

We propose that the creative cognition of cinematographers in the process of image-making, as well as that of the viewers, may embody, and psycho-physiologically simulate the movements of the camera as if they were moving themselves (Coëgnarts, 2017; Tikka, 2022). In Henderson's words, "cinema overcomes two-dimensionality through its 'walk-around' capability" (Henderson, 1970). Such a relationship between movement (either by mobile frame or other similar means) and space is so essential that without it, Sobchack claims, there would be no cinema (Sobchack, 2016, p. 66). Garrett Brown, the inventor of Steadicam, puts it simply: "We are there"—similar to a human eye, the moving camera explores moment-by-moment missing information related to the physical story space (Brown, 2003). For the viewer, camera movement provides an experience of "subjective movement through an objective world" (Bordwell, 1977, p. 23).

1.2. Camera movement from ecological perspective

According to the ecological view, exploration of one's surroundings by means of a coordinated interplay of vision and sensorimotor locomotion allows for detecting affordances for goal-directed actions in the natural environment (Gibson, 1966). The Gibsonian perception–action loop assumes the dominance of the visual system over the motor system. Contemporary views on human cognitive inference abilities to explore space assume continuous prediction coding in the brain for immediate updates of the optic information (Tivadar et al., 2021). Prediction errors describe the neural bottom-up processes involved when unexpected events instead of expected events occur in the optic field (Alefantis et al., 2022).

Gibson's ecological views on perception have been adopted and widely applied in the field of cognitive film studies (e.g., Detenber and Reeves, 1996; Anderson, 1998; Anderson and Anderson, 2005; Cutting, 2005; Smith, 2012; Tan, 2018). The neural prediction coding and prediction error may be assumed intrinsic to the affective–cognitive sense-making processes that take place when perceiving continuously unfolding narrative information. Indeed, prediction coding and error can be considered as a systemic counterpart in the brain for the viewer's perception of false cues and false information which at some point deliver surprise for the viewers. Camera movements play an important part in both hiding something from the viewer as well as directing the viewer's attention to that something.

In the context of watching a movie, it can be assumed that the flow of optical arrays (light) generated by the moving camera and then projected on the screen are to some extent perceptually similar to physically moving in space. If a visual system is dominant over the muscular, according to Gibson, then the viewers could easily surrender to the visual illusion that they are moving with the camera in the narrative space, although they are physically stationarily seated in the cinema chair. In a similar vein, Detenber and Reeves (1996), regarding the perceptual responses to the film, claim that "there is no switch in the brain that deactivates them just because the stimulus is mediated rather than real" (Detenber and Reeves, 1996, p. 78). The perception of the animated screen operates akin to the perception of real life, but viewers also learn how to comprehend film and television footage through repeated experiences (Salt, 2009, p. 32). Visual perception alone, with or without locomotion, does not enable the cognitive mapping of space. Constructing an understanding of perceived space further involves experience and memories of moving around in one's environment (Neisser, 1980).

A significant component of human engagement with the arts and sciences relies on the human ability to mentally simulate situations, actions, and the consequences of those actions without moving a muscle in their body. Cinematographers can imagine their camera movements in a given space based on the memory of the space, without any locomotion. Furthermore, the seated viewer is constrained in their bodily movements. Drawing from Bolens' (2012) account, Müller (2014) suggests that perceived bodily movement onscreen stimulates the sensorimotor system, and the viewer reconstructs the same locomotion mentally. In return, although partially illusory, the analogous movements on screen and viewers' perceptions allow viewers to establish a stronger emotional engagement with onscreen events (Bolens, 2012; Müller, 2014). In this sense, the motion may influence emotional responses (Simons et al., 1999).

1.3. Related studies in psychology

Mühlberger et al. (2008) assessed the emotional impacts of looming, receding, and static pictures on viewers in three different contexts: pleasant, unpleasant, and natural. The results demonstrated that the unpleasant looming pictures elicited strong responses both in valence and arousal ratings. The outcomes support the claim that alteration in physical distance affects emotional responses. Furthermore, the authors suggest that more sophisticated stimuli materials should be used for future experiments on the subject (Mühlberger et al., 2008). Several studies have evaluated the influence of motion stimuli on emotion (Detenber and Reeves, 1996; Detenber et al., 1998; Simons et al., 1999, 2000). Participants were shown still, and moving versions of the same clips were obtained from film and TV footage. Subsequently, their valence and arousal ratings were measured with self-reports and physiological measurement tools (e.g., skin conductance response and heart rate). Despite the different measurement tools and experiment designs (e.g., within-subjects vs. between-subjects) within the three experiments, the findings were consistent. Both in self-reports and physiological measurements, compared to still images, moving images appeared to be more arousing irrespective of whether the image was positive, negative, or neutral. The authors conclude that "the impact of image motion on the image-induced emotional response is inherent to motion itself" (Simons et al., 2000, p. 708).

In another study, Visch and Tan (2009) presented participants with different animated films in which moving objects depict chase scenes varied as five parameters—velocity, efficiency, fluency, detail, and deformation. Subjects categorized these clips into four genres (non-fiction, comedy, drama, and action) and also rated their emotional responses. Findings revealed that merely by watching different object movement patterns, viewers were consistently able to categorize them into genres. Furthermore, the movement patterns identified as fiction genres and their corresponding emotional reaction ratings were in line (e.g., comedy = response "funny", drama = "sad", and action = "impressive") (Visch and Tan, 2009). Overall, the above-mentioned studies in psychology exhibit guiding results concerning the emotional repercussions of motion, which can be applied to the context of camera movements in cinema.

1.4. Different ways of moving the camera

When faced with choices concerning camera movements and storytelling, cinematographers intuitively turn to their tacit knowledge (Calhoun, 2003; Pavlus, 2003; Lotman, 2016, 2021). Due to the nature of their work, filmmakers are often concerned with practical and narrative questions, such as whether there is a motivation to move the camera or whether executing a certain camera movement at any particular point would help the story (Nielsen, 2007).

Our study focuses on three established moving camera techniques: dolly, Steadicam, and handheld, which are compared to static cameras. We are interested in whether their distinct qualities produce different experiences for the viewers. Dolly is a wheeled cart-like device with a mounted camera. It can either be put on track or simply used on its own wheels. Dolly creates a smooth and stable movement. Steadicam is a special camera stabilization system invented by camera operator Garrett Brown in 1975. The Steadicam operator wears a vest that has an artificial arm attached that absorbs any friction of camera movement. The camera is mounted on the Steadicam sled which is connected to the arm. An important part of the Steadicam stability is not only the skill of "flying" it but also the skill of balancing the rig depending on the shoot (against the gravity with the drop-down speed and against the centrifugal force). Unlike the bulky and heavy dolly, Steadicam is more flexible to use and allows its operator to roam around freely. Steadicam generates smooth and stable movement akin to the dolly. However, depending on the skill level of its operator, the sense of slight "human touch" could be perceivable. Finally, as the name suggests, a handheld camera is a technique in which the camera operators operate the camera placed in their hands, shoulders, or hips, depending on the size of the camera. In contrast to dolly and Steadicam movements, handheld cameras may generate unstable images. The choice of the type of camera movement is made depending on the style and genre of the film in production, and what type of emotional experience the images are designed to convey for the viewers.

The chosen camera movement and related technical devices each have their different implications on the nature of the movement and how the viewer perceives it. They cause altering effects on the emotional engagement and bodily involvement of the viewer in the given scene. Thus, understanding the affective functions of different camera movement techniques is crucial for filmmakers. For instance, director of photography Vittorio Storaro (ASC) argues that compared to the limitations of dolly, Steadicam allows him to convey the "rush of feelings between the main characters" and determine an emotional state (Ferrara, 2001, p. 147). Steadicam inventor Garret Brown suggests that "Steadicam shots most closely resemble what humans see through our remarkably-stabilized eyeballs as we navigate our own daily 'movies'" (Pennington, 2020). Another Steadicam operator Ted Churchill states that Steadicam has the capability of "scaring half of the country to death" due to the sense of involvement it affords (Churchill, 1983, p. 119). In line with Brown, Jeff Mart also claims that unlike the sense of glide that dolly fabricates, Steadicam produces slightly wobbly and "less-than-perfect motions", which is closer to the genuine human experience of moving (Comer, 1993, p. 78). Director John Carpenter characterizes handheld shots as "moving chaos" (Ferrara, 2001, p. 114). Such examples illustrate the filmmakers' perspective that there are fundamental differences between the ways the camera is moved.

1.5. Embodiment of the cinematographer

The theories and practitioners' accounts presented above stress the cognitive and narrative potential of camera movement. However, they often attribute such potencies to the camera device itself as an autonomous entity, while disregarding the creative and cognitive mind behind the moving camera, the cinematographer. In our view, the cinematographer is the main source of the embodiment simulation of the viewer, and the camera is a means at their disposal. In movies, "the body of the spectator, the body of the film, and the body of the filmmaker" are intertwined (Gallese and Guerra, 2012, p. 189). The body of the film is the reflection of the author's embodied knowledge which, in return, is simulated by the viewers. MacDougall (2006) explains "...any image we make carries the imprint of our bodies [...] They are, in a sense, mirrors of our bodies, replicating the whole of the body's activity, with its physical movements [...] Corporeal images are not just the images of other bodies; they are also images of the body behind the camera and its relations with the world" (MacDougall, 2006, p. 3). Hence, cinematographers and directors "create, layer by layer, a living object sharing perceptual and cognitive structures with its viewer" (Gallese and Guerra, 2012, p. 189). In this sense, this filmic world, constructed by what Gallese and Guerra (2012; 190) refers to as the "filmic cognition of filmmakers", presents an "interaction between author's embodied knowledge gained in the real-life-situatedness and the modified representations of author's embodied knowledge gained therein" (Tikka, 2006, p. 146).

Tikka (2022) recently proposed a model labeled "Enactive Authorship". According to this proposal, embodied experiences of the author are simulated by the viewer *via* the protagonist's situatedness, and the author is the "embodied cognizer" (Tikka, 2022). We consider the camera as an active participant in any given scene and as the eye of the spectator. We propose to extend this model and argue that in the case of camera movement, cinematographers take the role of embodied cognizers and utilize their embodied experience, which is deeply rooted in their tacit knowledge, to create the movement of the camera in their prefilmic mental space to manipulate the emotions and the perception of viewers in accordance with the needs of the story. In exchange, when executed in filmic physical space and appearing on the screen, the mobile frame triggers the embodied simulation of the viewers and links them with the deliberately manipulated version of the cinematographer's cognition. If "film style creates the condition for an embodied film cognition" as Guerra suggests (Guerra, 2015, p. 143), then we would argue that it is the cinematographer who sets up such conditions.

We do not suggest that cinematographers must be given all the credit for devising a camera movement. Naturally, directors are also involved in creative decisions. However, the creative input of cinematographers in films is often overlooked in academic texts. For instance, reflecting on Deleuze's critique of Alfred Hitchcock's *Notorious,* Gallese and Guerra (2012 p. 200) wrote "Hitchcock aims to contact the viewer at a pre-cognitive level exploiting the potentiality of camera movements, and promoting an embodied approach capable of enhancing the suspense effect" (Deleuze, 1986; Gallese and Guerra, 2012). This account is in line with our proposal, as it exemplifies the deliberate cognitive input of authors to simulate viewer embodiment. However, it disregards the contribution of cinematographer Ted Tetzlaff (ASC) and gives all credit to Hitchcock.

An insight motivating the experiment in this contribution is that cinematographers can be understood as "embodied cognizers," who extend a modified and manipulated version of their bodily perceptions on the screen to viewers via camera movements. Cinematographers are likely intuitively aware of the different sensational and perceptual implications of different camera movement techniques through tacit knowledge and utilize this embodied understanding effectively. As testing the intuitions of cinematographers would be a challenging task, we start by exploring the effect of camera movements (and lack thereof) on viewers' emotional states in a controlled experimental setting.

2. Methods and materials

2.1. Stimuli

The stimulus scenes were designed to resemble scenes from actual movies in order to provide an experience more similar to watching a movie for the participants. Three scenes of different moods, suggesting an erotic encounter, a horror/thriller, and an ambiguous mood were each shot with four camera techniques, Steadicam, dolly, handheld, and static, resulting in 12 clips of approximately 45 s each. The zoom technique was not included as, in our view, zooming merely magnifies the image mechanically, instead of enhancing the embodied illusion of getting physically closer to an object in space. In this sense, zoom is "more transitive than transformative" (Sobchack, 2016, p. 80). Stationary shots were included to enable comparison with the moving ones.

In order to ensure that the main contrast would be the camera movements, we reduced other sources of variation while still ensuring the presence of essential visual elements of moody cinematic imagery. Visual components of the scene, such as lighting and placement of objects, were kept identical in all 12 stimulus variations. Furthermore, the use of music, sound, and human actors was avoided to keep the focus on the differences in the camera. While we made an effort to do so, in naturalistic stimuli like this, it is not possible to wholly eliminate the effects of other cinematographic variables such as focal length, shot size, lighting scheme, and framing. This is a tradeoff between ecological validity (naturalness) and experimental control.

In all three mood scenes, the scene starts with the camera tilting up from the ground and panning to the right. Each scene contains specific visual clues, such as blood trails on the ground in the horror scene and red rose petals in the romantic encounter. After the up-tilting, the camera reaches eye level, it starts moving toward a door by means of either Steadicam, dolly, or handheld and stops right in front of a closed door. Figure 1 shows frame grabs from all three scenes and Figure 2 illustrates the movement path of the camera. The starting and the ending frame of the camera movement as well as the pace and the duration of the movement were kept similar. In the stationary or static camera condition, the camera stays stable after tilting up and panning right. In cinematic terms, the camera movement refers to the physical displacement of the camera. When

FIGURE 1
Top to bottom: frame grabs from pleasant, neutral, and unpleasant scenes.

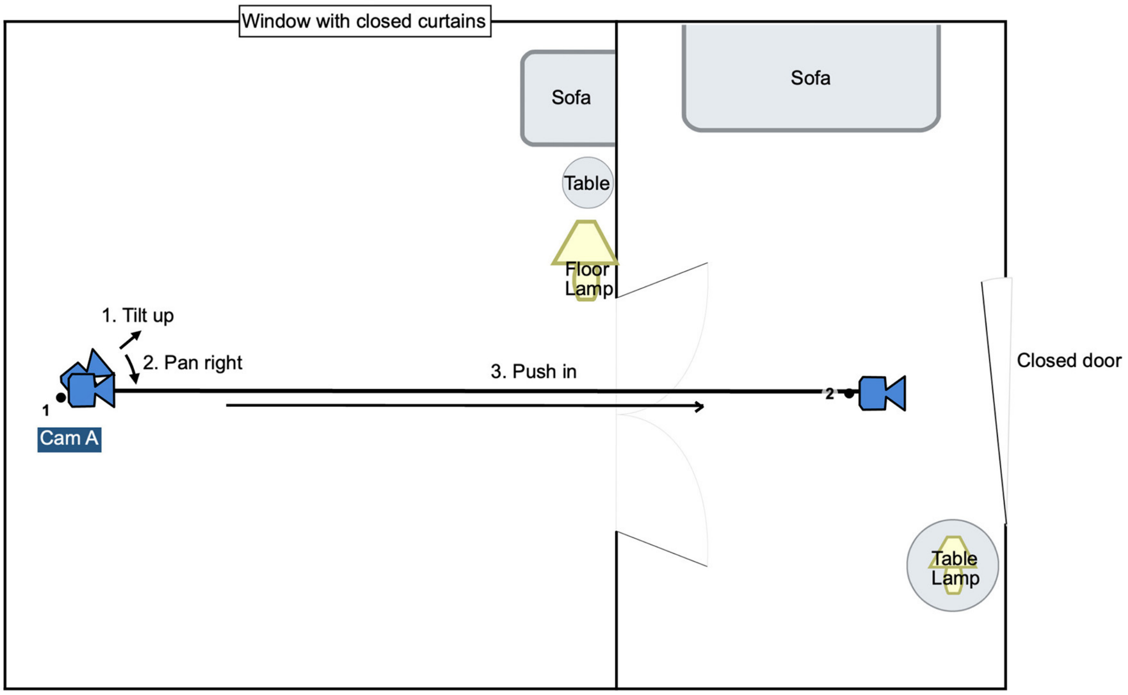


FIGURE 2
Floor plan and schematics of the stimuli scenes.

the camera is fixed with a tripod or any other means, the shot is often considered static, even if it pans or tilts.

All mood scenes were produced with a team of film professionals using professional cinema equipment. A cinematic lighting scheme was created in a controlled studio environment. The footage was captured by Arri Amira in 2K ProRes 444XQ LogC format and color graded in the DaVinci suite to further match contemporary professional film production standards.

All the stimuli scenes are open access. Access links to the materials can be found under the data availability statement section at the end of the article.

2.2. Participants

In total, 28 healthy women and 16 men (between the age of 19 and 68, with an average age of 28) participated in the experiment. The participants were recruited through university e-mail lists and Facebook groups. All subjects gave their written consent to participate in the study after orally being informed about the procedure and the duration of the experiment. No risk factors listed in the Ethics Committee of Tallinn University guidelines were present during the experiment and the ethical principles of informed consent were followed. To compensate for their time, each subject received a gift card to a local bookstore after the experiment had been completed.

2.3. Experimental design and procedure

Out of 12 clips, four stimulus sets were combined, each of which consisted of three different mood clips and three different camera movements. The underlying thought was to present each participant with only one camera movement variation from each mood condition in order to avoid familiarization by repetition. Hence, with that being the only condition, 12 clips were shuffled into four groups randomly (see Table 1 for stimuli sets). As a result, each participant saw three clips and each clip was seen a total of 11 times.

As participants were asked to write their ratings on separate paper assessment forms, each session started with oral instructions about how to fill out the assessment form after each film clip. Each session started with the showing of two training clips extracted from two feature films, in order to familiarize them with the alternating process of viewing a clip and immediately filling out the assessment form.

The first clip (60 s) was a Steadicam shot from "Goodfellas" (dir. Martin Scorsese, dop Michael Ballhaus), and the second clip (51 s) was the handheld camera shot from the opening scene of "Children of Men" (dir. Alfonso Cuarón, dop Emmanuel Lubezki). After this training session, participants were given a chance to ask questions. This was followed by the actual experiment, where the subjects watched the three clips from their assigned set. In the beginning of each clip, a fixation cross for 1,000 ms was displayed. The subjects were given 40 s to fill out the assessment form after each clip, after which, the next clip was automatically started. The experiment took place in a silent editing room located at the Baltic Film and Media School of Tallinn University. The subjects watched the clips alone, with lights off, on a 27-inch iMac computer.

TABLE 1 Stimuli sets.

	Erotic (Positive)	Ambiguity (Neutral)	Horror (Negative)
Static (STA)	1	2	3
Handheld (HH)	2	1	4
Dolly (DOL)	4	3	1
Steadicam (STE)	3	4	2

Set 1: Ambiguity (HH)—Erotic (STA)—Horror (DOL).
Set 2: Ambiguity (STA)—Erotic (HH)—Horror (STE).
Set 3: Ambiguity (DOL)—Erotic (STE)—Horror (STA).
Set 4: Erotic (DOL)—Ambiguity (STE)—Horror (HH).

2.4. Measures and statistical modeling

Participant experience assessment relied on self-reported measures in the form of a rating task. For emotional valence and arousal, we used a 5-scale version of the Self-Assessment Manikin (SAM) (Lang, 1980; Bradley and Lang, 1994), which consists of two scales that depict five manikins (stylized human figures). For valence, their expressions range from very pleasant to very unpleasant, and for arousal from very excited to very calm. These two scales were presented visually, without numerical values, arranged from left to right. For the statistical analyses, we coded them as ranging from 1 to 5, where 5 is the most pleasant and most exciting. As for the assessment of the involvement concerning the embodiment of the camera, following Heimann et al. (2014), we asked the following three questions:

1) On a scale of 1 to 5, how much did you feel involved in the scene?
2) On a scale of 1 to 5, how much did you feel as if the camera was your own eyes?
3) On a scale of 1 to 5, how much did you feel as if you were moving with the camera? (Was not asked for the static shots)

Here, the values were accompanied by the following descriptions: "Didn't feel at all", "Didn't feel", "Neutral/Unsure," "Felt", and "Felt strongly".

We employed a mixed-effects generalized linear regression framework to test the effect of movement on the five reported ratings (implemented using the lme4 and lmerTest packages in R; cf. Bates et al., 2015; Kuznetsova et al., 2017). While all outcome variables are technically ordinal, the linear model's assumptions are sufficiently met. In all cases, participants were modeled as a random effect, to account for repeated measurements (due to model convergence issues likely stemming from small data, we were only able to fit models with just random intercept). The movement was dichotomized as a binomial variable to simplify analysis (all moving cameras contrasted with the baseline of the static camera). We did not have any particular hypotheses concerning the relationship between mood and the outcome variables (emotions and involvement), so it was treated as a control variable (with ambiguous mood as a baseline). To assess whether movement affects the outcome variables in the respective models, we used a stepwise likelihood ratio test approach, comparing the mood model to a null (random effect only) model, and the full model (mood in interaction with movement) to the mood-only model. This, therefore, also indicates whether mood (averaging across movements) had an effect on the response, which we also consider interesting. We furthermore allow for interaction between mood and movement, assuming that effect may vary between scene types (as it indeed does, as indicated in Figure 2). The modeling differed slightly for Question 3, which concerned movement and was not asked about static scenes. Here, we followed the same stepwise procedure and included movement as a full 3-level variable (with the dolly as the baseline). In some cases, we carried out exploratory modeling on within-condition differences of interest (such as differences of the camera between a single scene), using a fixed-effects only linear regression approach that makes it convenient to report beta coefficients along with conventional *p*-values (while those data subsets do not include repeated measurements, given the nature of the sampling as described above).

3. Results

As discussed in the Introduction section, this study was motivated by multiple goals: to carry out experimentation on the effects of camera movements on viewer responses, but also to evaluate the suitability of experimentation for testing cinematographic theories such as embodiment (as discussed above). We, therefore, provide both statistical modeling for the movement-related hypotheses as well as exploratory results, accompanied by interpretation based on tacit cinematographic experience. Table 2 lists the results from the modeling as described in the methods section above in the form of likelihood ratio test *p*-values. In summary, after controlling for repeated measures and possible mood variation, movement in contrast to the static camera did not appear to significantly affect either valence or arousal ratings. Movement and static differ in terms of the perceived degree of involvement (Q1) and the sense of seeing the scene through one's own eyes (Q2), and the direction of the effects depends on the mood. Ignoring movement effects, we observe that valence and arousal differ between mood scenes (as somewhat expected). Given the multiple testing scenarios (five distinct questions), we also applied the Bonferroni correction (adjusting alpha to 0.01), but this ended up not affecting the interpretation.

3.1. Valence and arousal ratings

Irrespective of different types of moving or static frames, the overall atmosphere generated by the set dressing appeared to be the main agent which led the viewer to perceive the scene as either pleasant or unpleasant in valence (the mood-only model in comparison to the null model: $\chi^2(2) = 33.41$, $p < 0.001$). The averages of all condition combinations are illustrated in Figure 3 for valence as well as other questions. In other words, whether the camera moves by any means or stays stationary did not result in the audience feeling more or less pleasant when the nature of the scene is already comprehensible (valence, controlling for mood, compared to mood-only model: $\chi^2(3) = 0.84$, $p = 0.84$). The only exception to this pattern seems to be the erotic encounter (positive) scene shot by the dolly movement (mean valence 2.82), which is slightly lower than other camera types (but only the difference to handheld is significant; fixed-effects only linear model with the dolly as a baseline: handheld $\beta = 0.82$, $p = 0.03$).

In general, arousal does not appear to be strongly affected by a moving camera compared to a static camera (full model compared to mood-only: $\chi^2(3) = 4.56$, $p = 0.21$). As can be expected, the suspenseful horror scene with blood on the floor readily elicits stronger arousal than its calmer counterparts though. In the case of an erotic encounter (pleasant) scene, dolly movement and the static frame (both averaged 2.91) would appear to lead to calmer reactions, but this difference is not statistically significant. If this difference would be replicated with a larger sample, then one explanation could be the smooth and stable frame generated by the dolly and stationary camera, which are qualities that match with pleasant onscreen events in this case.

While there was no significant difference within the horror scene, the ambiguous scene (neutral) where the onscreen clues were not evident enough for meaning-making, elicited a difference in the dolly condition ($\beta = 1.27$, $p = 0.02$, compared to static). This can be explained by the anticipation of something being revealed soon, thus adding the excitement of "diving into the unknown" (as expressed by one of the participants). In a situation where the audience is uncertain about what to expect, the mobile frame may embody active exploration of the space, whereas the static frame provides a static point of observation.

The same qualities attributed to the dolly movement and the static frame, namely smoothness and stability, can both initiate excitement and calmness simultaneously. The perception and influence of such features depend on the circumstances under which the viewers are exposed to them. For instance, the exact same dolly movement that implies a presence of an ominous entity in a horror scene can also be an invitation to join the intimacy of an erotic encounter. In cinematographic practice, different camera movements must be taken into consideration together with the settings they are executed in, including components such as music, acting, directing, and sound. As Sobchack simply puts it, "the moving becomes particularly meaningful in every film's specific context" (Sobchack, 2016, p. 86).

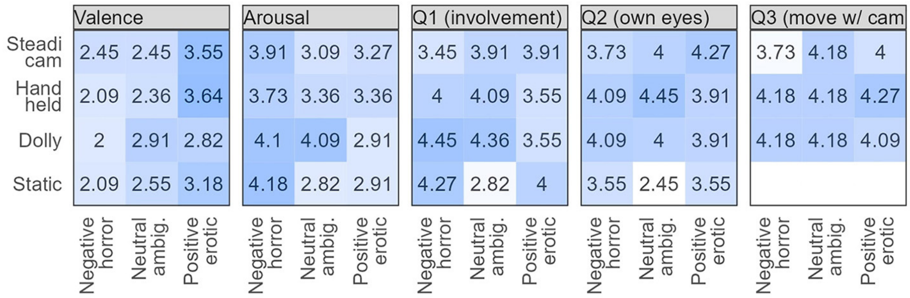
3.2. Degree of involvement

Question 1 of the involvement set of questions was stated as follows: on a scale of 1 to 5, how much did you feel involved in the scene? As indicated in Table 2, the model that includes movement describes the data better than mood only, i.e., a moving camera elicits different reactions from participants compared to static ($\chi^2(3) = 12.41$, $p = 0.006$). However, as illustrated in Figure 2, the effect depends on the scene. In the neutral scene, all moving cameras increase the feeling of involvement compared to static (all $p \leq 0.01$). In other scenes, moving cameras do not differ from static. In the full model (where moving cameras are aggregated and compared to static), the interaction between mood and camera

TABLE 2 Modeling outcomes as the likelihood ratio test *p*-values from stepwise comparisons of mixed-effects regression models (mood-only against the intercept-only null model, and subsequently the full model against that).

Model	Valence	Arousal	Q1, involved	Q2, own eyes	Q3, moving with the camera
Mood-only vs. null	**<0.001**	**<0.001**	0.28	0.66	0.72
Mood*movement, vs. mood-only	0.84	0.21	**0.006**	**<0.001**	0.99

Values below the convtentional alpha of 0.05 are highlighted in bold. * stands for interaction.



	Valence			Arousal			Q1 (involvement)			Q2 (own eyes)			Q3 (move w/ cam		
	Negative horror	Neutral ambig.	Positive erotic	Negative horror	Neutral ambig.	Positive erotic	Negative horror	Neutral ambig.	Positive erotic	Negative horror	Neutral ambig.	Positive erotic	Negative horror	Neutral ambig.	Positive erotic
Steadi cam	2.45	2.45	3.55	3.91	3.09	3.27	3.45	3.91	3.91	3.73	4	4.27	3.73	4.18	4
Hand held	2.09	2.36	3.64	3.73	3.36	3.36	4	4.09	3.55	4.09	4.45	3.91	4.18	4.18	4.27
Dolly	2	2.91	2.82	4.1	4.09	2.91	4.45	4.36	3.55	4.09	4	3.91	4.18	4.18	4.09
Static	2.09	2.55	3.18	4.18	2.82	2.91	4.27	2.82	4	3.55	2.45	3.55			

FIGURE 3
Mean rating values, ranging between 1 and 5, for combinations of camera movement and mood scene, averaged across participants. Values closer to 5 (darker background) correspond to a stronger response, e.g., the horror scene elicits stronger arousal than the erotic scene. In questions 1 and 2, moving cameras, in contrast to a static camera, lead to different immersion ratings.

movement is a significant effect. The beta coefficients for the two interaction terms of moving camera with the non-neutral moods are both about −1.4. Through a borderline result, it further illustrates the need to consider camera movements in the context of the scene.

Question 2 asked, "On a scale of 1 to 5, how much did you feel as if the camera was your own eyes?" The moving camera effect was significant compared to the mood control ($\chi^2(3) = 27.18$, $p < 0.001$). Similarly to Q1, this was mostly driven by the neutral condition, with stronger immersion from moving cameras. This may be explained as an apparent correspondence between the camera movement in the space and the subject's visual perspective. This suggests that movement enhances the embodiment of the camera as a kind of extension of the viewer's eye.

Finally, question 3 was intended as explorative, to probe the differences between moving cameras, without comparison to the static condition (where this question was omitted). It asked, "On a scale of 1 to 5, how much did you feel as if you were moving with the camera?" Responses to question 3 revealed little to no difference between camera movements (movement variable with three levels, dolly as baseline; full model vs. mood only: $\chi^2(6) = 3.02$, $p = 0.81$). The Steadicam average in the horror scene has the lowest value, but that difference is not statistically significant.

In summary, the experimental data paints a varied picture. Camera movement might not necessarily affect valence and arousal responses in audiences, although the scene and set dressing often do, as may be expected. Movement does seem to make a difference in terms of immersion, particularly in certain types of scenes, but participants did not pick up on the individual differences between the subtly different camera techniques when asked how much they felt like they were moving with the camera. The scores were similarly high across the board. This is an interesting finding in light of the theories and tacit experience accounts discussed above and invites future research on the topic.

4. Discussion

The empirical focus of our study has been on understanding how the viewers experience different camera movements in dramatized scenes with different moods. Any film viewing experience can be argued to build on the anticipation of future events based on previous events in the film as well as the accumulated knowledge due to lifelong experience of films and narratives (Kauttonen et al., 2014). The knowledge related to film genre conventions (i.e., media literacy), the way how films are structured, their audiovisual design, the character's appearance, and how the plot binds it all together dictates what type of experiences the viewers expect to be engaged with.

Our findings partially support the hypothesis that a moving camera enhances viewer engagement and the feeling of being involved more than a static camera. However, emotional responses of arousal and valence were similar between moving and static camera conditions, while the content of the three scenes appeared to make a difference. The findings are in line with the neuroscientific study by Tikka et al. (2018), where it was shown that the emotionally loaded narrative content of the film scene may override the formal differences between the scene representation (audiovisual vs. written representation). We hope that this contribution will inspire experimental designs in further studies. Some limitations and shortcomings (discussed below) were noted during the process of the experiment and the data analysis based on various feedback both from participants and colleagues.

4.1. Filmmakers' tacit knowledge

Cinematographers tend to draw upon their inherited or tacit knowledge, the pool of information built up by filmmakers throughout the history of moviemaking. It consists of "rules of thumb that have been passed around as the 'right' or best' way to construct shots and sequences" (O'Leary, 2003, p. 199). Such knowledge could be acquired over time through practice, or simply transmitted from master to apprentice. It is not a definite set of rules, but rather an ever-evolving collective endeavor. Individual experiences also add idiosyncratic layers to this knowledge. In the short term, it can be developed through "shorter, momentary windows of revelation", and in the longer term, it "is shaped by the experiences and collaborations a person has encountered" (Lotman, 2021, p. 34). Cinematographer Robert Richardson (ASC) explains that as his career progressed, his decisions in regard to moving the camera became "less bound to a sort of 'from the hip' sensibility" and more attentive (Pavlus, 2003, p. 41). Along the same line, his colleague in cinematography, Owen Roizman (ASC), reveals that he leaves the decision of whether to move the camera to his instincts and the camera moves only when it feels right (Calhoun, 2003).

When filmmakers pursue their "gut feeling", they may, in fact, be applying accumulated tacit knowledge (Lotman, 2021, p. 148) from lifelong experiences "transformed into learning through a cycle of learning involving experiencing, reflecting, thinking, and acting" (Kolb and Yeganeh, 2016, p. 101). "Tacit knowledge has a personal quality, which makes it hard to formalize and communicate" yet it is "deeply rooted in action, commitment, and involvement in a specific context" (Nonaka, 1994, p. 16). Attempting to understand movies through theories alone without considering filmmakers' tacit knowledge, particularly of what "camera movements can or ought to do", is unproductive since "filmmakers are rarely gripped by theories" (Morgan, 2016, p. 243). Or in Vorkapich's words, "what is there obtained perhaps accidentally should be sought, studied, and used consciously" (Vorkapich, 1930, p. 31). In this study, we attempted to measure and quantify what effect such intuitive decisions may have on the eventual audiences, by constructing a controlled yet relatively naturalistic experimental setting.

While a central aspect of filmmakers' tacit knowledge, camera moving techniques and their effects may be too subtle to be distinguished by non-professional viewers (Bordwell, 1977) and the relationship between camera movements and the human perceptual system went underexplored for a while (as noted by Sobchack, 1982), we see camera movements as a fruitful field of study for cognitive science and neurosciences, which in turn may also assist filmmakers in their art.

4.2. Limitations of the study

Our goal was to find a balance between creating natural, film-like stimuli, and retaining reasonable control over experimental variables to enable the quantification and interpretation of the results. We attempted to minimize the effect of potentially distracting variables and keep the stimuli scenes largely constant in terms of duration, pace, framing, lighting, focal length; the absence of a human agent, sound, and music). However, it can be assumed that viewers can still be affected by every component of shot design and engage with anything on the screen regardless of the intended contrast between the conditions. In addition, ordinary fiction movies do not usually lack (human) agents, sounds, or music. Therefore, while our stimuli are life-like, they might not look quite like typical movies. It is possible that in a more conventional cinema experience, the differences between camera movements could be less noticeable or vice versa.

Unlike some prior work (as discussed above), we set out to run the experiment in three scenes differentiated by mood and setting. This complicates the analysis, as the mood variable needs to be accounted for, but also allows for a richer interpretation, which we hope will inspire future research. Due to the temporal and financial limitations of a live experiment with human participants, we opted to have the same participant viewing multiple clips. While we control for these repeated measures in the statistical models, the small amount of data gathered here does not afford statistically reliable modeling of the sequence in which they watched the clips, which has the potential of affecting reactions. Future research should either seek to gather more data for more robust modeling or avoid showing participants multiple clips in succession. After the experiment procedure, some of the subjects orally expressed that irrespective of the content and the visual dynamics onscreen they felt excited simply because they did not know what they were about to see. Such excitement may have had an effect on arousal ratings, but we control for that in the statistical models via the random effects structure.

Proposing a cinematographer-oriented theory yet conducting an audience-oriented experiment might be considered contradictory. However, cinematographers devise the filmic world to manipulate viewers. In this sense, the cinematographer is the sender and the film viewer is the receiver. Whether a cinematographer succeeds in transmitting their perceptual bodily senses can only be understood by studying the receiver audience. The crux here was to design the experiment and stimulus with the cinematographer's insights. That being said, our experience of the process (including stimuli creation, experimental design, and data analysis) revealed that focusing on the creative process of the cinematographers might have been a better approach.

One of the shortcomings is related specifically to Steadicam. Steadicam Operator is a particular film profession that requires years of training to reach an adequate skill set and professional level of experience. The more skilled and experienced the operator, the more the Steadicam movement feels fluid and stable. In contrast, the floating of the horizontal line and excessive suggestion of human presence behind the device indicates an amateur operation of Steadicam. In our case, the Steadicam operator was somewhat novice; and in some Steadicam clips, the horizontal axis was not always straight, possibly resulting in a feeling of floating, which did not represent the usual fluidity of Steadicam shots.

5. Conclusion

In this study, we explored the views of embodied cognition on the interaction between the visual system and locomotion and sought to understand the effects of different camera movement techniques (which lead to differing camera movements on the screen) and self-reported viewer responses in terms of emotion and immersion. We explored the difference between a static and a moving camera, with three types in particular: Steadicam, dolly, and handheld, and their functions in narrative filmmaking. The insights from several professional cinematographers based on their tacit knowledge were reflected upon from the point of view of embodied simulation and embodied mind.

We highlight the utility of combining theory, professional experience-based accounts, and rigorous experimentation in controlled laboratory settings. The question remains, how exactly do the embodied experiences of the viewers and the embodiment of the tacit embodied knowledge of the cinematographers coincide in the filmmaking practice? The knowledge gained here will serve as a stepping stone for developing a more thorough understanding of embodiment, camera movement techniques, and also experimental methodologies to study them.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MY designed the experiment, conducted the data collection, and wrote the manuscript as the first author. PT contributed to the experiment design and writing of the manuscript. EL contributed to the experiment and stimuli designs. AK conducted the statistical analysis, wrote the relevant sections, and contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Tallinn University, Baltic Film, Media and Arts School (BFM) fund for Research and Creative Study and European Union Horizon 2020 funding programme (Grant No. 101035820) via FilmEU Research, Innovation, Transformation (FilmEU_RIT). The study by PT has been supported by the EU Mobilitas Pluss Top Researcher Grant (MOBTT90) and the Estonian Research Council. AK was supported by the CUDANERA Chair project, funded through the European Union's Horizon 2020 Research and Innovation Program (Grant No. 810961).

Acknowledgments

We thank Andero Uusberg from the University of Tartu, Maarten Coëgnarts from the University of Antwerp and Mati Mõttus from Tallinn University for their input during the experimental design process.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alefantis, P., Lakshminarasimhan, K., Avila, E., Noel, J.-P., Pitkow, X., and Angelaki, D. E. (2022). Sensory evidence accumulation using optic flow in a naturalistic navigation task. *J. Neurosci.* 42, 5451–5462. doi: 10.1523/JNEUROSCI.2203-21.2022

Anderson, J., and Anderson, B. F., editors. (2005). *Moving Image theory: Ecological Considerations*. Carbondale: Southern Illinois University Press.

Anderson, J. D. (1998). *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*. Paperback ed. Carbondale, IL: Southern Illinois University Press.

Barker, J. M. (2009). *The Tactile Eye: Touch and the Cinematic Experience*. Berkeley: University of California Press.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bolens, G. (2012). Kinesthetic empathy in charlie chaplin's silent films. In: Reynolds, D., and Reason, M. *Kinesthetic Empathy in Creative and Cultural Practices*, eds. Bristol: Intellect. p. 145–156.

Bordwell, D. (1977). Camera movement and cinematic space. *Cine-Tracts* 1, 19–25.

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Brown, G. (2003). *The Moving Camera Part One*. Available online at: https://www.garrettcam.com/the-moving-camera-part-1 (accessed January 17, 2023).

Calhoun, J. (2003). Putting the "Move" in Movie. *Am. Cinematogr.* 84, 72–85.

Churchill, T. (1983). Steadicam: an operator's perspective-part I. *Am. Cinematogr. 64, 36–39* 113–120.

Coëgnarts, M. (2017). Cinema and the embodied mind: metaphor and simulation in understanding meaning in films. *Palgrave Commun.* 3, 17067. doi: 10.1057/palcomms.2017.67

Coëgnarts, M., and Slugan, M. (2022). Embodying meaning visually: from perceptual dynamics to motion kinematics. *Art Percept.* 10, 137–158. doi: 10.1163/22134913-bja10035

Comer, B. (1993). Steadicam hits its stride. *Am. Cinematogr.* 74, 73–78.

Cutting, J. (2005). "Perceiving scenes in film and in the world. In: Anderson, J. D., and Anderson, B. F. *Moving Image Theory: Ecological Considerations*. Carbondale, IL: Southern Illinois University Press. p. 9–27.

Deleuze, G. (1986). *Cinema 1: The Movement-Image*. Minneapolis: University of Minnesota.

Detenber, B. H., and Reeves, B. (1996). A bio-informational theory of emotion: motion and image size effects on viewers. *J. Commun.* 46, 66–84. doi: 10.1111/j.1460-2466.1996.tb01489.x

Detenber, B. H., Simons, R. F., and Bennett, G. G. (1998). Roll 'em!: the effects of picture motion on emotional responses. *J. Broadcast. Electron. Media* 42, 113–127. doi: 10.1080/08838159809364437

Ferrara, S. (2001). *Steadicam: Techniques and Aesthetics*. Oxford; Boston: Focal.

Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 659–669. doi: 10.1098/rstb.2006.2002

Gallese, V., and Guerra, M. (2014). The feeling of motion: camera movements and motor cognition. *Ciné. Cie* XIV, 103–112. doi: 10.1093/oso/9780198793533.003.0003

Gallese, V., and Guerra, M. (2012). Embodying movies: embodied simulation and film studies. *Cine. J. Philos. Mov. Image* 3, 183–201.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. London: George Allen and Unwin LTD.

Grodal, T. (2002). *Moving Pictures: A New Theory of Film Genres, Feelings, and Cognition*. Oxford New York: Clarendon press Oxford University Press.

Guerra, M. (2015). Modes of action at the movies, or re-thinking film style from the embodied perspective. In: Coëgnarts, M., and Kravanja, P., editors. *Embodied Cognition and Cinema*. Leuven: Leuven University Press.

Hall, H. (1932). Cinematographers and directors meet discuss camera trucking problems. *Am. Cinematogr. XIII* 10, 47.

Heimann, K., Uithol, S., Calbi, M., Umilt,à, M. A., Guerra, M., Fingerhut, J., et al. (2019). Embodying the camera: an EEG study on the effect of camera movements on film spectators' sensorimotor cortex activation. *PLoS ONE* 14, e0211026. doi: 10.1371/journal.pone.0211026

Heimann, K., Umilt,à, M. A., Guerra, M., and Gallese, V. (2014). Moving mirrors: a high-density eeg study investigating the effect of camera movements on motor cortex activation during action observation. *J. Cogn. Neurosci.* 26, 2087–2101. doi: 10.1162/jocn_a_00602

Henderson, B. (1970). Toward a non-bourgeois camera style. *Film Q.* 24, 2–14. doi: 10.2307/1211215

Jääskeläinen, I. P., Klucharev, V., Panidi, K., and Shestakova, A. N. (2020). Neural processing of narratives: from individual processing to viral propagation. *Front. Hum. Neurosci.* 14, 253. doi: 10.3389/fnhum.2020.00253

Jääskeläinen, I. P., Sams, M., Glerean, E., and Ahveninen, J. (2021). Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage* 224, 117445. doi: 10.1016/j.neuroimage.2020.117445

Kauttonen, J., Hlushchuk, Y., Jääskeläinen, I. P., and Tikka, P. (2018). Brain mechanisms underlying cue-based memorizing during free viewing of movie Memento. *NeuroImage* 172, 313–325. doi: 10.1016/j.neuroimage.2018.01.068

Kauttonen, J., Kaipainen, M., and Tikka, P. (2014). *Model of Narrative Nowness for Neurocinematic Experiments* Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik. 11 pages.

Keating, P. (2019). *The Dynamic Frame: Camera Movement in Classical Hollywood*. New York, NY: Columbia University Press.

Kolb, D. A., and Yeganeh, B. (2016). Deliberate experiential learning: mastering the art of learning from experience. In: *Contemporary Organizational Behavior: From Ideas to Action*, eds E. D. Kimberly, A. B. Kayes, D.K. Christopher, (Boston: Pearson). p. 100–108.

Kovács, A. B., Raz, G., Valente, G., Svanera, M., and Benini, S. (2019). A robust neural fingerprint of cinematic shot-scale. *Projections* 13, 23–52. doi: 10.3167/proj.2019.130303

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82. doi: 10.18637/jss.v082.i13

Lakoff, G., and Johnson, M. L. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York, NY: Basic books.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In: *Technology in Mental Health Care Delivery Systems.*, eds S. B. Joseph, J. H. James, W. A. Thomas (Norwood, NJ: Ablex). p. 119–137.

Lotman, E. (2016). Exploring the ways cinematography affects viewers' perceived empathy towards onscreen characters. *Balt. Screen Media Rev.* 4, 88–105. doi: 10.1515/bsmr-2017-0005

Lotman, E. (2021). *Experiential heuristics in fiction film cinematography (dissertation/doctoral thesis)*. Tallinn University, Tallinn, Estonia.

MacDougall, D. (2006). *The Corporeal Image: Film, Ethnography, and the Senses*. Princeton, N.J: Princeton University Press.

Marzola, L. (2020). *German Cinema Comes to Hollywood: AC in the 1930s*. Am. Cinematogr. Available online at: https://theasc.com/articles/german-cinema-comes-to-hollywood (accessed January 31, 2022).

Metz, C. (1991). *Film Language: A Semiotics of the Cinema*. University of Chicago Press ed. Chicago: University of Chicago Press.

Morgan, D. (2016). Where are we?: camera movements and the problem of point of view. *New Rev. Film Telev. Stud.* 14, 222–248. doi: 10.1080/17400309.2015.1125702

Mühlberger, A., Neumann, R., Wieser, M. J., and Pauli, P. (2008). The impact of changes in spatial distance on emotional responses. *Emotion* 8, 192–198. doi: 10.1037/1528-3542.8.2.192

Müller, S. (2014). Embodied cognition and camera mobility in F. W. Murnau's the last laugh and Fritz Lang's M. *Paragraph* 37, 32–46. doi: 10.3366/para.2014.0108

Neisser, U. (1980). On "Social Knowing." *Pers. Soc. Psychol. Bull.* 6, 601–605. doi: 10.1177/014616728064012

Nielsen, J. I. (2007). *Camera movement in narrative cinema — Towards a taxonomy of functions (dissertation/doctoral thesis)*. Aarhus University, Aarhus, Denmark.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5, 14–37. doi: 10.1287/orsc.5.1.14

O'Leary, B. (2003). Camera movements in hollywood's westering genre: a functional semiotic approach. *Criticism* 45, 197–222. doi: 10.1353/crt.2004.0002

Pavlus, J. (2003). Robert Richardson, ASC on camera movement. *Am. Cinematogr.* 84, 41.

Pennington, A. (2020). *Industry Innovators: Garrett Brown*. Available online at: https://www.ibc.org/interviews/industry-innovators-garrett-brown/5338.article (accessed January 17, 2023).

Raz, G., and Hendler, T. (2014). Forking cinematic paths to the self: neurocinematically informed model of empathy in motion pictures. *Projections* 8, 89–114. doi: 10.3167/proj.2014.080206

Salt, B. (2009). *Film Style and Technology: History and Analysis*. 3rd ed. London: Starword.

Schonig, J. (2017). Seeing aspects of the moving camera: on the twofoldness of the mobile frame. *Synoptique* 5, 57–78.

Simons, R. F., Detenber, B. H., Reiss, J. E., and Shults, C. W. (2000). Image motion and context: a between- and within-subjects comparison. *Psychophysiology* 37, 706–710. doi: 10.1111/1469-8986.3750706

Simons, R. F., Detenber, B. H., Roedema, T. M., and Reiss, J. E. (1999). Emotion processing in three systems: the medium and the message. *Psychophysiology* 36, 619–627. doi: 10.1111/1469-8986.3650619

Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections* 6, 1–27. doi: 10.3167/proj.2012.060102

Sobchack, V. (1982). Toward inhabited space: the semiotic structure of camera movement in the cinema. *Semiotica* 41. doi: 10.1515/semi.1982.41.1-4.317

Sobchack, V. (1992). *The Address of the Eye: A Phenomenology of Film Experience*. Princeton, NJ: Princeton University Press.

Sobchack, V. (2016). 'The Active Eye' (Revisited): toward a phenomenology of cinematic movement. In: Ferencz-Flatz, C., and Hanich, J. *Studia Phaenomenologica, Volume 16/2016: Film and Phenomenology*. Bukarest: Zeta Books.

Tan, E. S. (2018). A psychology of the film. *Palgrave Commun.* 4, 82. doi: 10.1057/s41599-018-0111-y

Tikka, P. (2006). Cinema as externalization of consciousness. In: *Screen Consciousness: Cinema, Mind and World*. Amsterdam: Rodopi.

Tikka, P. (2008). *Enactive Cinema: Simultatorium Eisensteinense*. Helsinki: University of Art and Design.

Tikka, P. (2022). Enactive authorship: second-order simulation of the viewer experience—a neurocinematic approach. *Projections* 16, 47–66. doi: 10.3167/proj.2022.160104

Tikka, P., and Kaipainen, M. Y. (2014). From naturalistic neuroscience to modeling radical embodiment with narrative enactive systems. *Front. Hum. Neurosci.* 8, 794. doi: 10.3389/fnhum.2014.00794

Tikka, P., Kauttonen, J., and Hlushchuk, Y. (2018). Narrative comprehension beyond language: Common brain networks activated by a movie and its script. *PLoS ONE.* 13, e0200134. doi: 10.1371/journal.pone.0200134

Tivadar, R. I., Knight, R. T., and Tzovara, A. (2021). Automatic sensory predictions: a review of predictive mechanisms in the brain and their link to conscious processing. *Front. Hum. Neurosci.* 15, 702520. doi: 10.3389/fnhum.2021.702520

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge (Mass.) London: MIT Press.

Vesey, G. N. A. (1965). *The Embodied Mind*. Abingdon-on-Thames: Routledge.

Visch, V. T., and Tan, E. S. (2009). Categorizing moving objects into film genres: the effect of animacy attribution, emotional response, and the deviation from non-fiction. *Cognition* 110, 265–272. doi: 10.1016/j.cognition.2008.10.018

Vorkapich, S. (1930). Cinematics: some principles underlying effective cinematography. In: *Cinematographic Annual, 1930.*, ed H. Hal (Hollywood, CA: The American Society of Cinematographers. Inc). p. 29–33.

Zacks, J. M. (2010). The brain's cutting-room floor: segmentation of narrative cinema. *Front. Hum. Neurosci.* 4, 168. doi: 10.3389/fnhum.2010.00168

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 04 July 2023
DOI 10.3389/fpsyg.2023.1183660

OPEN ACCESS

EDITED BY
Celia Andreu-Sánchez,
Autonomous University of Barcelona, Spain

REVIEWED BY
Tina Kubrak,
Institute of Psychology (RAS), Russia
Tess Lankhuizen,
University of Amsterdam, Netherlands

*CORRESPONDENCE
Cynthia Cabañas
✉ ccaban01@mail.bbk.ac.uk

RECEIVED 10 March 2023
ACCEPTED 17 May 2023
PUBLISHED 04 July 2023

CITATION
Cabañas C, Senju A and Smith TJ (2023) The audience who knew too much: investigating the role of spontaneous theory of mind on the processing of dramatic irony scenes in film.
Front. Psychol. 14:1183660.
doi: 10.3389/fpsyg.2023.1183660

The audience who knew too much: investigating the role of spontaneous theory of mind on the processing of dramatic irony scenes in film

Cynthia Cabañas[1]*, Atsushi Senju[1,2] and Tim J. Smith[1]

[1]Cognition in Naturalistic Environments (CINE) Lab, Department of Psychological Sciences, Birkbeck, University of London, London, United Kingdom, [2]Research Centre for Child Mental Development, Hamamatsu University School of Medicine, Hamamatsu, Japan

As in real life, cinema viewers rely on spontaneous theory of mind (SToM) to interpret characters' mental states. Thus, analyzing cinematic structures offers a unique opportunity to examine ecologically valid sociocognitive processes. We conducted a proof-of-concept study ($N = 42$) to explore how SToM inferences impact film event comprehension in dramatic irony scenes, where knowledge divergence exists between the audience and characters. We hypothesized that spectators would focus more on characters' mental states in such false-belief inducing scenarios compared to scenarios without such disparity. We used six Harold Lloyd silent comedy clips in a narrative comprehension and spontaneous mental state attribution study with a between-subject (Knowledge Manipulation: Installation vs. Control) and within-subject (Phase: Context vs. Exploitation) comparisons. We provided critical information unknown to the characters only to the Installation group and withheld it from the Control group. By comparing differences in participants' descriptions of the clips during the Context phase (varying across groups) and Exploitation phase (same across groups), we evaluated viewers' processing of the same scenes based on their false- or true-belief representations. Our findings indicate that the Installation group used more cognitive mental state words during the Exploitation phase relative to the Context phase, suggesting that exposure to undisclosed critical information enhances the frequency of spontaneous epistemic state inferences and integration into event models of the exploitation. This research advances neurocinematics by highlighting spontaneous sociocognitive processes in event perception and comprehension and provides a novel dramatic irony film corpus and measures for future moment-to-moment SToM processing studies across cognitive-behavioral, physiological, and neural levels.

KEYWORDS

spontaneous theory of mind, dramatic irony, false belief attribution, event comprehension, neurocinematics, film comprehension

1. Introduction

Have you ever watched a suspense film and found yourself yelling at the screen, warning the protagonist of impending danger? In a heart-pumping scene from Hitchcock's (1960) "Psycho", as the unsuspecting protagonist, Marion Crane, steps into the shower, the audience is acutely aware of the danger lurking just outside the bathroom door and approaching her.

Hitchcock, a master of suspense, frequently used *dramatic irony*—a narrative device where the audience knows something that the characters do not—to heighten the tension and draw the audience deeper into the story. This moment, like many others in film, relies on the viewer's ability to understand the mental states of the characters on screen.

In everyday social situations, we constantly monitor what others know and do not know. For instance, parents often infer their child's knowledge gaps and beliefs to adapt guidance accordingly. This continuous adjustment of our understanding of others' mental states is essential for navigating complex social interactions. As social creatures, we rely on our theory of mind (ToM) to attribute mental states to ourselves and others, allowing us to make sense of differing thoughts and feelings in daily life (Premack and Woodruff, 1978). Importantly, cinema, literature, or theater also makes use of these ToM skills, also known as mentalizing abilities, implicitly motivating us to make sense of characters' actions by attributing and tracking their mental states to understand the stream of events from the narrative (Zunshine, 2006; Levin et al., 2013; Tan, 2013).

Since the emergence of ToM research, a wide range of stimuli has been utilized to study this sociocognitive process, including cartoons, animations, and photographs (e.g., Wimmer and Perner, 1983; Baron-Cohen et al., 1985; Zaitchik, 1990; Abell et al., 2000; Castelli et al., 2000; Gallagher et al., 2000). While these studies have been informative, they have been criticized for both their lack of ecological validity and for the excessive signposting and instruction to produce explicit mental state inferences (Bloom and German, 2000; Dziobek, 2012; Achim et al., 2013). For instance, the Sally–Anne task (Baron-Cohen et al., 1985) is a classic experiment in the field of developmental psychology that tests an individual's ability to understand false beliefs. The task involves presenting the participant with a story in which two characters, Sally and Anne, are present. Sally puts her toy in a basket and then exits the room. While Sally is absent, Anne moves the toy to a box. The participant is then asked to predict where Sally will look for the toy when she returns. The correct answer to pass this false-belief test is that Sally will look for the toy where she last left it, not where Anne moved it.

Despite the Sally–Anne task's foundational role in ToM research, the task has several limitations. One notable concern is its ecological validity, as the task presents a simplified scenario that does not adequately capture the complexity of real-life situations where we often need to integrate contextual information and spontaneously infer others' epistemic states in a more nuanced manner (Wellman et al., 2001; Ruffman and Perner, 2005). Additionally, the Sally–Anne task is primarily designed to assess ToM in young children (Gopnik and Astington, 1988; Astington and Gopnik, 1991), which limits its applicability in studying more advanced ToM abilities in older children and adults. The task may not be sufficiently challenging for older participants including individuals with autism spectrum conditions (ASCs), potentially resulting in ceiling effects or underestimating their ToM capabilities (Apperly, 2011; Senju, 2012; Livingston et al., 2019).

While the Sally–Anne task has its shortcomings, it has served as a critical starting point for research into ToM, particularly in highlighting the importance of false-belief understanding. Due to the simplicity and clarity of false-belief tasks that have allowed for more controlled experimentation, ToM research in adults has tried to adapt false-belief tasks for adults to understand the underlying cognitive mechanisms of this complex sociocognitive process. For instance, researchers have attempted to examine the curse of knowledge bias (Birch and Bloom, 2007; Bernstein et al., 2011; Sommerville et al., 2013) and higher-order ToM understanding (Kinderman et al., 1998; Stiller and Dunbar, 2007; Oesch and Dunbar, 2017). However, it remains unclear whether these tasks are adequate for exploring the intricacies of adult ToM as they may require more advanced conceptual knowledge or be influenced by working memory and executive function capacity (Brown-Schmidt, 2009; Ryskin and Brown-Schmidt, 2014). Addressing these concerns and identifying more suitable tasks are crucial for advancing adult ToM research and understanding individual differences in everyday ToM abilities.

The development of tasks that incorporate such false-belief structures together within naturalistic stimuli, such as films, could be a step in this direction as they require the integration of contextual information and the understanding of multiple mental states simultaneously (Levin et al., 2013; Tan, 2013). This approach may provide a more ecologically valid assessment of mentalizing abilities while still maintaining experimental control. In this study, we propose a novel approach to studying ToM by harnessing the engaging power of films featuring dramatic irony structures, which could potentially serve as a naturalistic, filmed adaptations of false-belief tasks.

The use of dramatic irony often follows a three-act structure (Lavandier, 2005 modified and extended to include Cohn, 2016): (1) An *establisher* scene sets up the situation and introduces the characters' goals. (2) An *installation* scene provides crucial information that one or more characters are unaware of, which sets the stage for the dramatic conflict to come. These oblivious characters are known as the victims of dramatic irony. Finally, (3) the *exploitation* scenes depict the victims' reactions and actions in response to their ignorance, which can lead to misunderstandings, decoys, or deception.

The structure of this scenario closely resembles that of the classical Sally–Anne task, but it is integrated with additional contextual information. For instance, in the mentioned iconic shower scene from "Psycho" first exhibits Marion's vulnerability as she is preparing to take a shower (*establisher*). While Marion, is under the water in the shower, the audience is able to catch a glimpse of shadow behind the curtain in the shower (*installation*). The audience quickly understands that Marion, the victim of dramatic irony, is unaware of the presence of this figure. As the shadow slowly approaches, it starts taking form into what seems an old lady prepared to violently attack Marion with a knife (*exploitation*), while the audience helplessly anticipates the consequences of Marion's false belief that she is alone in the bathroom.

While the field that studies theory of mind has historically centered around the investigation of (false) belief attributions, scholars such as Phillips et al. (2021) have recently highlighted the importance of examining the role of knowledge attributions in social interactions. Critically, dramatic irony creates a unique opportunity to distinguish between when viewers categorize a character as being ignorant and when they label them as holding false beliefs (Scott and Baillargeon, 2009; Baillargeon et al., 2010). In particular, when labeling characters as ignorant, we have to attribute lack of knowledge to them, whereas when attributing false beliefs, we assume they hold (incorrect) information not supported by reality. The temporal and contextual factors that influence individuals' tendency to make knowledge attributions are currently unknown.

On the contrary, over the past decade, many social neuroscience researchers have shifted toward using films as a rich source of naturalistic stimuli, enabling the examination of ToM processing in more realistic scenarios (Dziobek, 2012; Achim et al., 2013; Devine and Hughes, 2013). However, even studies that use audio-visual stimuli often examine explicit theory of mind, where participants are manifestly asked to infer mental states of characters in the film (Heavey et al., 2000; Dziobek et al., 2006; Golan et al., 2006; Devine and Hughes, 2013). In these tasks, it is assumed that individuals spontaneously engage in theory of mind reasoning during such scenarios and in everyday life [see Heyes (2014) for an argument about submentalizing]. However, there is a challenge to test such assumption, given that current tasks typically instruct participants to mentalize.

The need to overcome this critical limitation led to the development of implicit or spontaneous theory of mind (SToM) tasks that aimed at measuring the ability to infer mental states in naturalistic scenarios without explicit prompts or instructions. Some of these strategies include free-viewing paradigms combined with talk aloud tasks or *post-hoc* free recall comprehension questions which researchers can code the mentalizing skills and tendencies of participants (Klin, 2000; Barnes et al., 2009; Rice and Redcay, 2015; Altschuler et al., 2018; Bálint et al., 2018; Rooney and Bálint, 2018). Importantly, Apperly (2012) distinguishes between the ability to mentalize and the tendency to spontaneously pay attention to another person's mental states. This distinction is essential in SToM paradigms since there is increasing evidence that individuals with ASC may perform successfully in mentalizing tasks attending to socially relevant information when explicitly instructed to but might be less likely to mentalize spontaneously without explicit instruction or task demand (Senju, 2012; Dufour et al., 2013).

Several studies indicate that examining how and what we understand from a film narrative has the potential to reveal differences in mentalizing tendencies. For instance, Lahnakoski et al. (2014) observed differences in eye movements when viewers shifted their focus between characters and objects, while Yeshurun et al. (2017) found that neural representations of movie clips were more similar within groups who shared the same beliefs about a situation.

In this study, we present a proof-of-concept demonstration of the value of utilizing dramatic irony sequences in film as a naturalistic test of viewers' complex SToM processing. Our main goal was to examine whether dramatic irony structures naturally prompt audiences to make more inferences about characters' epistemic states and beliefs compared to control scenes without dramatic irony. We propose that by investigating the processing of these structures in films through a free recall task, we can gain valuable insights into individuals' spontaneous mental state inferences. This approach offers a point of reference, illustrating typical responses to the task and stimulus, and lays the methodological foundation for future investigations into the neural basis and individual differences in these processes.

In dramatic irony scenes, relevant information about the characters' ignorance and/or false beliefs is presumably extracted from the *installation* scene and integrated into a situational event model or "person schema" to understand characters in films (Smith, 1995; Zwaan and Radvansky, 1998; Loschky et al., 2020), drawing on their knowledge of real people. Moreover, Bálint et al. (2018) argued that by increasing the attentional resources allocated to characters and their facial expressions, close-ups could potentially boost the likelihood that a viewer's mental model of a narrative includes the mental states of the characters. Both cognitive ToM (recognizing others' beliefs, thoughts, and motivations) and affective ToM (inferring their emotions and feelings) are essential for understanding the divergence between our own beliefs and emotions and those of the victim of dramatic irony. For instance, in the "Psycho" example only by considering and incorporating Marion's beliefs into an event model of the scene, we can make sense of her calm, untroubled emotional expression in the shower while we see a threatening figure behind her.

Thus, we hypothesized that, similarly, by increasing attentional resources to the salient disparity of knowledge between character and audience, the structure of dramatic irony scenes (vs. control scenes) would prompt spectators to infer more often the mental state of characters, both cognitive and affective, and incorporate them into their event models. We examined this hypothesis by manipulating the audience's access to knowledge from the *installation* scenes in a narrative comprehension and spontaneous mental state attribution study, allowing us to compare how viewers process the same scenes depending on their ToM representations.

Participants in the Installation group watch the installation scene which contains crucial information to understand the dramatic irony conflict, while those in the Control group do not. We measured comprehension of the dramatic irony conflict and the frequency of mental state references to examine how each condition determined how participants reasoned about the events and described them. Critically, including both complementary measures can provide a comprehensive and nuanced understanding of how theory of mind is involved in the comprehension of dramatic irony.

Previous studies have used coding schemes that often identify both affective mental states, which refers to others' emotions or feelings (e.g., "*Marion looks relaxed and undisturbed in the shower*"), and cognitive mental states, which refers to others' thoughts, beliefs, or intentions (e.g., "*Marion thinks she is alone in the bathroom*") (Klin, 2000; Rice and Redcay, 2015; Altschuler et al., 2018; Rooney and Bálint, 2018). However, these types of mental states were previously collapsed together for later analysis. Crucially, there is a large body of evidence that shows that emotional and cognitive components of sociocognitive processes such as ToM and empathy are interdependent but separate mechanisms in the brain (Dziobek et al., 2008; Abu-Akel and Shamay-Tsoory, 2011; Zaki and Ochsner, 2012). Although ToM and empathy are distinct sociocognitive processes, some overlap exists in their definitions among various authors, particularly when comparing cognitive aspects of ToM and empathy with affective aspects of ToM and empathy. Nevertheless, empathy involves an experience-sharing component that is not necessarily inherent in ToM. Preckel et al. (2018) highlight that cognitive and affective empathy, as well as theory of mind (ToM), are underpinned by distinct, independent brain networks, while also acknowledging the interplay between these processes. Specifically, Cuff et al. (2016) note that while some empathy definitions focus on either affective or cognitive aspects, many encompass both. The authors further support this distinction by citing empirical evidence from research in personality, developmental disorders, and neurological studies, supporting the notion that cognitive and affective empathy are separate constructs. Specifically during film watching, Shany et al. (2021) found different neural patterns for affect sharing, affective ToM, and cognitive ToM. To capture these dissociative components in the processing of dramatic irony scenes, we considered cognitive and affective mental states both separately and together in our analysis.

We predicted that participants in the Installation group would understand the dramatic irony conflict that arises from the victim's ignorance of critical information more often than participants in the Control group. This prediction served as a manipulation check that exposure to *installation* scene is required to understand dramatic irony and that the *exploitation* scene alone does not contain sufficient information.

We expected the Installation group to use a higher frequency of overall mental state references (H1) than the Control group, in line with previous literature which demonstrated that increased number of mental state references in free recall was associated with more accurate mental state attribution (e.g., Bálint et al., 2018; Rooney and Bálint, 2018). We predicted participants in the Installation group would show a higher frequency of cognitive mental state references (e.g., beliefs, thoughts, and intentions) in their free recall responses compared to the Control group (H1.1), suggesting that exposure to critical information unknown to a character, promotes a more thorough understanding and integration of that characters' thought processes and mental perspectives. We also anticipated participants in the Installation group would demonstrate a higher frequency of affective mental state references (e.g., emotions, feelings, desires) vs. the Control group (H1.2), indicating that experiencing the installation scene enhances one's sensitivity to the characters' emotional experiences and the subtleties of their affective states.

Finally, to account for the difference in clip length between the Installation and Control groups, we examined the frequency of mental states specifically in participants in two different phases of the descriptions: the Context phase and the Exploitation phase (see Section 2.5). We did not expect to find differences in mental state references in the Context phase but expected the Installation group to differ from the Control group in the number of overall (H2), cognitive (H2.1), and affective (H2.2) mental state references in participants' descriptions of the Exploitation phase, where the dramatic irony conflict occurs.

2. Materials and methods

2.1. Design

The present study was an online experiment conducted on Gorilla.sc (Anwyl-Irvine et al., 2020) with a mixed-design: a between-subject variable (*Knowledge Manipulation*) with two levels (Installation vs. Control); a within-subject variable (*Phase*, which here denotes the part of the description that participants referred to) with two levels (Context vs. Exploitation) and two dependent variables, dramatic irony conflict comprehension (DIcomp) score and mental state reference frequency (MSRF) as a proxy of SToM tendency. Participants were randomly assigned to one of the two Knowledge manipulation conditions. The order of the presentation of the six clips (blocks) was randomized.

2.2. Participants

A convenience sample of 50 participants (33 female participants, age: M = 30, SD = 9.24) was recruited from the university student participant pool (SONA). Given that there were no previous studies we could use to power this study, the target sample size ($N = 42$) was derived from an a priori power analysis carried out using the software G*Power (Faul et al., 2009) for an estimated effect size of Cohen's d = 0.8 with sufficient power (0.9; $\alpha = 0.05$). Given the novelty in the experimental online design, the dropout rate for this study was uncertain. The initial sample size of 50 participants was chosen to compensate for the anticipated dropout rate due to exclusion criteria or technical errors.

The mean age of our sample (30 years old) and the large standard deviation is reflective of the diverse student population at our university, which includes a high proportion of mature and international students. Therefore, inclusion criteria consisted of normal or corrected-to-normal vision, an advanced English level to answer the comprehension questions and no previous diagnosis of autism spectrum conditions (ASCs). Given the limited number of trials (six clips), participants were excluded if they did not answer the free recall comprehension tasks for each of the six clips. They were also excluded if their descriptions for each clip did not include at least 1 sentence per fragment (2 for the Control group since these participants see *establisher* and *exploitation* scenes; 3 for the Installation group since these participants see *establisher*, *installation,* and *exploitation* scenes) demonstrating low effort in performing the task. To avoid un-blinding of conditions, this exclusion phase was performed by author CC, before the coding of comprehension and mental state references was performed by two independent coders (BS and EE) unaware of the nature of the task and the groups these participants belonged to.

From the final sample, eight participants were discarded: five for not completing all the measures or due to previously set exclusion criteria for data quality, one for having previously seen one or more of the films, and two for reporting an intermediate or lower English level. Experimental procedures were approved by Birkbeck, University of London Ethics Board (181949). All subjects provided written informed consent.

2.3. Film clip corpus design

Stimuli were six self-contained film excerpts taken from different Harold Lloyd comedy silent movies (U-certified). The videos convey a short storyline, with no sound, white text is presented on a black background in between scenes (intertitles with verbal information). The content of each clip with an illustrative still of each phase can be found in the Supplementary material. All of the films the clips were taken from were silent-era Harold Lloyd films, including "Never Weaken" (Newmeyer, 1921), "Girl Shy" (Newmeyer and Taylor, 1924), "The Freshman" (Newmeyer and Taylor, 1925), "For Heaven's Sake" (Taylor, 1926), and "The Kid Brother" (Wilde et al., 1927). This selection was inspired on the Silent Film task developed by Devine and Hughes (2013), who used silent comedy clips from a Harold Lloyd single film. The Silent Films task is designed to measure participants' explicit understanding of beliefs and desires and engaging for a broader audience including adults and older children with different language groups and children who may have low verbal ability. We aimed to build on these stimulus design criteria by selecting similar Harold Lloyd film clips albeit in a systematic way based on the identification of false-belief inducing situations in dramatic irony structures, with the intention to create a film corpus which could be used for future studies examining implicit measures of SToM.

There are several other benefits to this selection: First, given that these films are approximately a century old, it is very unlikely that participants have watched them; second, the silent film format controlled for the influence of audio or verbal information, focusing viewer's on the images as the main source of information and making the stimuli suitable across future differences in participant verbal IQ (e.g., in studying Autism); and third, given the canonical structure of dramatic irony described in the introduction (Lavandier, 2005), the identification of these scenes (*establisher, installation, and exploitation*) could allow us to make testable hypothesis about temporal dynamics and sub-processes of SToM.

Two versions of the clips were created: The Installation complete dramatic irony film clips composed of three scenes (*establisher* scene, *installation* scene, and *exploitation* scene) creating a divergence of beliefs between participants and characters, thus, depicting instances of deception and misunderstanding. To manipulate the knowledge context, in the Control version of the clips, the *installation* scene is edited out; therefore, participants only see two scenes: *establisher* scene and *exploitation* scene. For this purpose, we made sure that the characters' behaviors were congruent or plausible in Installation and Control conditions. A short description and the duration for each version can be found in Table 1. Critically, viewers in both groups watch exactly the same *exploitation* scenes, allowing us to compare how viewers process the same scenes when they know more than the victim vs. when they do not (they are as ignorant as the victim). Figure 1 provides a visual representation of both Installation and Control versions of the film clips for illustration purposes. Detailed copyright information and permissions for the reuse of these clips can be found on our dedicated project webpage: https://www.cinelabresearch.com/haroldlloydproject. Researchers interested in using these stimuli for their studies are encouraged to contact us directly through the form provided on the website for more information.

2.4. Procedure

Participants run the experiment through their web browser through an online experimental task engine (Gorilla.sc) used to ease participants access to the experiment at home. Immediately after watching each clip, we asked them to perform a free recall task by answering to a prompt on the screen saying, "Please, take about five minutes and write a paragraph about what happened in the clip you just saw and why it happened." Participants typed their responses in a text box. There was no time limit to answer to this prompt. Participants were not asked for specific aspects of characters' mental states to avoid potential bias allowing us to obtain only SToM responses (Barnes et al., 2009; Rice and Redcay, 2015; Bálint et al., 2018) and measure potential differences between conditions. The approximate total duration of the task for each group was ∼21 min for the Installation group and 15 min for the Control group. These durations ensured that participants in both Knowledge conditions were exposed to an adequate amount of content while keeping the overall task duration manageable. To prevent fatigue and ensure participant engagement, breaks were provided after each block, allowing participants to rest before continuing with the subsequent clips.

At the end of the experiment, they were asked whether they had seen any of the films and a short debrief question about whether they noticed any pattern across the clips (i.e., the dramatic irony structures) to ensure that they were not aware of the dramatic irony structures, potentially having an influence on their SToM. None of the participants reported having noted a pattern in the structure of the clips. Finally, a debrief was shared with the participants explaining the background of the study and what had been measured as part of the study.

2.5. Coding scheme

Free recall responses were coded by an independent blind researcher for the comprehension of dramatic irony conflict and the frequency of mental state use (see next section for measure definitions). To ensure inter-rater reliability, a second researcher coded 25% of a sample of descriptions at random. Initial inter-rater reliability was evaluated for both dramatic irony comprehension (DIcomp) and Overall, Cognitive, and Affective mental state reference frequency (MSRF) coding separately using Krippendorff's alpha, which was calculated to be 0.938 for DIcomp; 0.749 for Overall MSRF; 0.713 for Cognitive MSRF; and 0.725 for Affective MSRF, which are considered acceptable or above levels of agreement (Krippendorff, 2004).

2.5.1. Dramatic irony conflict comprehension

Using a grading scheme based on Barnes and Baron-Cohen (2012) and Lavandier (2005), participants are scored on their understanding of dramatic irony conflict in a narrative. Full understanding (2 points) requires explaining the victim's ignorance of critical information and its impact on their goals. Partial understanding (1 point) involves recognizing the victim's ignorance but not its consequences. Failed understanding (0 points) lacks any reference to the victim's ignorance or its impact. In the "Girl Shy" example (see Table 1), mentioning Harold's unawareness of the check earned partial understanding, while discussing how this relates to his goal to earn money for publishing his book earned full understanding. No mention of his ignorance or its impact resulted

TABLE 1 Summary of film clips with respective control and installation versions, including their duration and a brief description of the scenes.

Clip title	Version	Duration	Short description
Never Weaken I	Control	2 m 21 s	Harold aims to showcase an osteopathic clinic's effectiveness by healing a man on the street, attracting new clients to the clinic where his love interest works
	Installation	3 m 26 s	Harold devises a plan with an acrobat to stake a fake injury recovery, drawing the attention of potential clients to the osteopathic clinic where his love interest works
The Freshman	Control	3 m 26 s	Harold attempts to join the football team, successfully secures a spot, and enthusiastically heads to the field to play
	Installation	3 m 54 s	Harold excitedly tells a girl he made the team and eagerly goes to the field, not knowing that his real role is the water boy
Never Weaken II	Control	1 m 09 s	Harold proposes to a girl and overhears a conversation where he finds out that she is being proposed to by another man
	Installation	1 m 25 s	Harold proposes to a girl who accepts, but later misunderstands her conversation with her brother, thinking she's being proposed to by another man
Girl Shy	Control	2 m 26 s	Harold tries to publish his book but is rejected by the publisher and receives a rejection letter in the mail
	Installation	3 m 06 s	Harold attempts to publish his book, and although initially rejected, the publisher reconsiders. Harold, believing the letter contains a rejection slip, tears up the unopened envelope containing a check
For Heaven's Sake	Control	4 m 11 s	A missionary and his daughter write to Harold for help raising money for their mission. Harold comes across their mission cart and offers them a significant contribution
	Installation	5 m 14 s	Harold accidentally burns a mission cart, writes a check to compensate, but is mistaken for a generous donor for the mission
The Kid Brother	Control	1 m 32 s	Two men trick the sheriff into signing a permit for their traveling show
	Installation	5 m 06 s	Harold, dressed as his sheriff father, is tricked into signing the permit for the two men

in failed understanding. Since participants watched six clips, the possible total scores for dramatic irony conflict comprehension ranged from 0 to 12.

2.5.2. Mental state reference frequency

Based on the ToM coding scheme by Bálint et al. (2018), informed by Meins and Fernyhough (2006), participant descriptions were divided into subject–verb–predicate coding units. Coders identified explicit mental state references. Here, mental state reference was defined as "any reference to an individual's mental life, relating to desire, wish, emotion, will, mind, imagination, interest, intellect, or metacognition" (Bálint et al., 2018). These references were also categorized as (a) affective (e.g., feelings and desires) or (b) cognitive (e.g., memory, knowledge, and intention). To account for individual differences, participants received scores for the proportion of mental state references to total coding units, indicating their theory-of-mind responding level.

At a second coding stage, to account for the difference in clip length between the Installation and Control groups, a third blind coder identified sentences referring to Exploitation scenes to separate scores for frequency of affective and cognitive mental states in Context and Exploitation phases, relative to coding unit count. The first phase included the description of the Context, which differs per group, including the establisher and installation scenes in the Installation group and only the establisher scene in the Control group, since the latter did not watch an installation scene. We did not expect to find differences in mental state references in this phase. The second phase was composed of the description of the exploitation scenes, which are the scenes that are viewed in both Installation and Control groups (see Figure 1), where we did expect to find differences in MSRF. The primary responsibility of this coder was thus to identify when participants' descriptions started referring to the exploitation scene (Exploitation phase), which was clearly defined in the coding manual. The parts of participant descriptions not belonging to the Exploitation scene were categorized as Context. This task did not involve the interpretation of ambiguous mental states or the assessment of participants' understanding of conflicts but rather focused on a more straightforward identification process based on well-defined criteria.

3. Results

This analysis plan for this study was preregistered on the Open Science Framework [doi: 10.17605/OSF.IO/PQRU6] Additional analyses examined the relationship between dependent variables (DIcomp and MSRF) for the Installation group. R in R-studio was used for data management and statistical analysis, ensuring assumptions of normality and homogeneity of variances were met for *t*-tests and ANOVAs.

Overall MSRF and Cognitive MSRF were normally distributed, allowing parametric testing. Affective MSRF was slightly non-normal and positively skewed. As a result and to further validate our findings, we carried out both parametric tests using

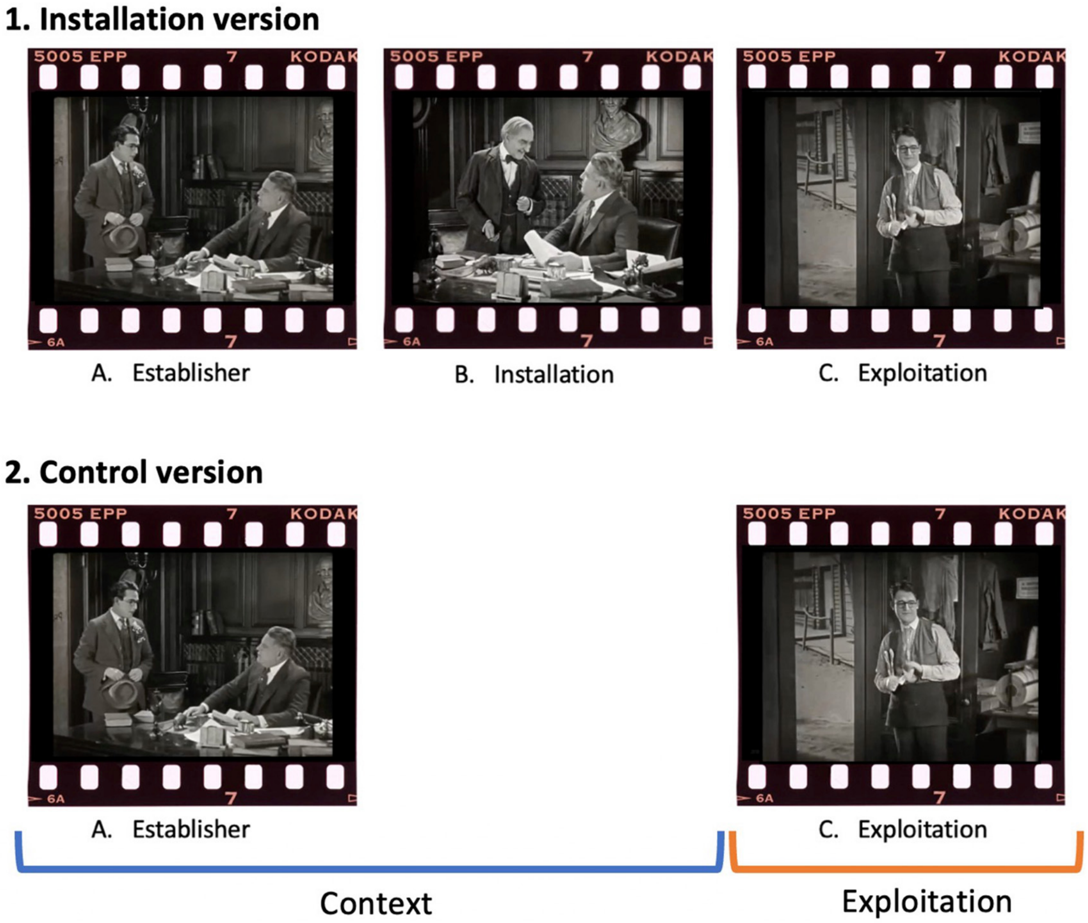


FIGURE 1
Example stills of both conditions from "Girl Shy" (Newmeyer and Taylor, 1924) included in the film corpus. In the Installation version, **(A)** Establisher: Harold visits a publishing house to inquire about the possibility of publishing his book. However, the publisher finds his book to be extremely comical, so they reject it and inform Harold that he will receive a rejection letter in the mail. **(B)** Installation: When Harold leaves, a senior employee convinces the editor to reconsider and publish the manuscript as a comedy. He then instructs the employee to send a check to Harold instead of the rejection letter. **(C)** Exploitation: Harold, downhearted and unaware of the content of the letter, tears it apart without opening it. 2. The control version only has **(A)** Establisher and **(C)** Exploitation; therefore, participants should interpret that Harold tore apart the rejection letter. The approximate duration of the clips was around 3 min. Stills taken with permission from Girl Shy (1924) © 2023 Harold Lloyd Entertainment, Inc. Reproduced with permission.

logarithmically transformed data and non-parametric tests, aiming to demonstrate the robustness and consistency of our results regardless of the specific statistical test employed (Field et al., 2012). Levene's test confirmed homogeneity of variances ($p > 0.05$). Table 2 displays summary statistics for measured variables per knowledge condition across all clips.

3.1. Manipulation check: DIcomp in installation vs. control groups

As a preliminary test, we assessed our manipulation check (higher DIcomp in the Installation group than Control) using a mixed-design ANOVA with Knowledge manipulation (Installation vs. Control) as a between-subject factor and Clip (1–6) as a within-subject factor and including participants' number in the error term. This accounted for individual variation and allowed us to similarity across clips in DIcomp. Significant main effects were found for Knowledge condition (F (1.234) = 344.44, $p < 0.001$, partial $\eta^2 = 0.60$) and Clip (F (5.1170) = 2.97, $p = 0.013$, partial $\eta^2 = 0.11$), with a significant interaction between them (F (5.1170) = 8.96, $p < 0.001$, partial $\eta^2 = 0.04$), indicating that the effect of Knowledge condition differed depending on the clip being shown.

We conducted six Welch's t-tests to compare DIcomp scores between groups for each clip. The results indicated significantly higher scores in the Installation group for five out of six clips (all $p < 0.05$), though the effect varied. As shown in Figure 2, Clip 1 showed no significant difference after Bonferroni correction ($p = 0.073$).

3.2. Hypothesis 1: MSRF in installation vs. control groups

To test H1 (higher MSRF in the Installation group than Control), a mixed-design ANOVA assessed the effects on MSRF of

TABLE 2 Descriptive and inferential statistics of all measures by Knowledge condition.

Measures	Condition	Mean	SD	95% CI lower	95% CI upper	*t*-test	*p*-value
Age	Control	29.00	8.16	25.18	32.82	−0.564	0.576
	Installation	30.52	9.11	26.38	34.67	-	-
Gender	Control	1.75	0.44	1.54	1.96	0.575	0.568
	Installation	1.67	0.48	1.45	1.89	-	-
English Level	Control	4.30	0.86	3.90	4.70	−0.309	0.758
	Installation	4.38	0.80	4.01	4.75	-	-
DIcomp	Control	1.75	1.59	1.01	2.49	−13.368	**0.000**
	Installation	9.71	2.19	8.72	10.71	-	-
Overall MSFR	Control	37.11	14.38	30.38	43.84	−1.476	0.149
	Installation	42.70	9.13	38.54	46.85	-	-
Cognitive MSFR	Control	21.65	8.84	17.51	25.78	−3.291	**0.002**
	Installation	31.32	9.97	26.78	35.86	-	-
Affective MSFR	Control	15.47	7.86	11.79	19.15	1.922	0.063
	Installation	11.48	5.06	9.17	13.78	-	-

DIcomp, Dramatic Irony Comprehension; MSRF, Mental State Reference Frequency; SD, standard deviation; CI, confidence interval.
Bold values denote significance at the level $p \leq 0.05$.

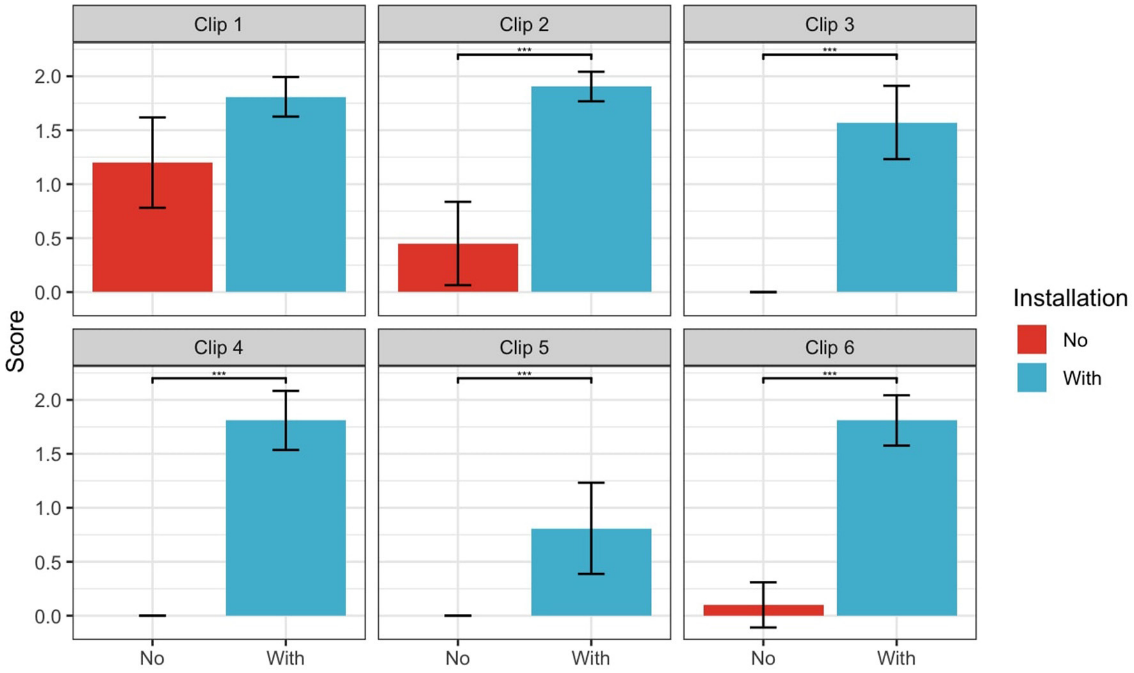


FIGURE 2
Dramatic Irony Comprehension (DIcomp) scores for each clip separately. Error bars represent 95% confidence intervals.

Knowledge manipulation (Installation vs. Control) as a between-subject factor, Clip (1–6) as a within-subject factor, and participants' number in the error term. By doing so, we accounted for individual variation and investigated the similarity across clips in MSRF. As shown in Figure 3, the results indicated a significant effect of condition on Overall MSRF (F (1, 228) = 5.197, p = 0.0236) and Cognitive MSRF (F (1.228) = 25.217, p < 0.0001). On the contrary, the effect of condition on Affective MSRF was not significant (F (1.228) = 3.457, p = 0.063). Logarithmically transforming and carrying out non-parametric, the data did not affect the results of the statistical tests so here we only report the parametric tests.

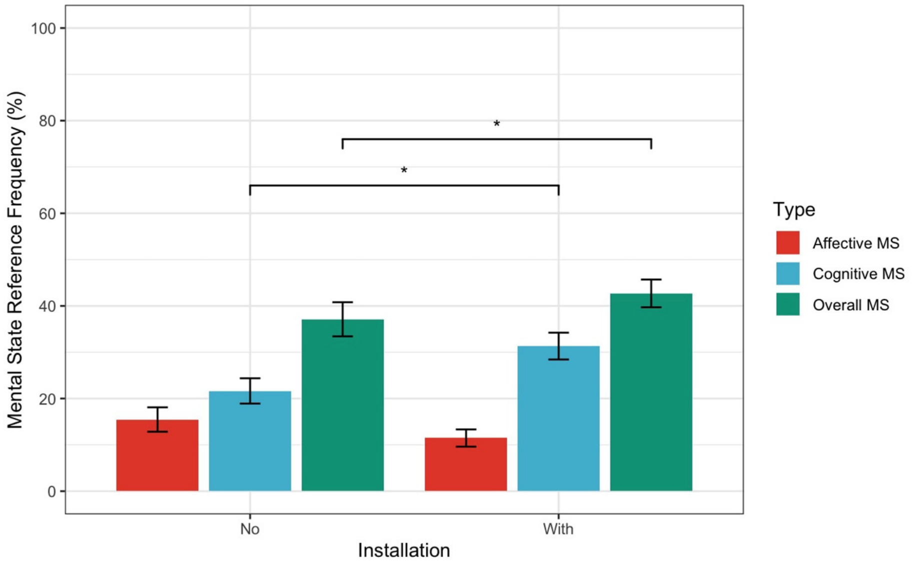


FIGURE 3
Comparison of Mental State Frequencies by condition across clips. Error bars represent 95% confidence intervals. *Indicates statistical significance with a p-value less than 0.05.

There was a significant effect of Clip on Cognitive MSRF (F (5.228) $= 2.555$, $p = 0.0284$) and Affective MSRF (F (5.228) $=$ 5.066, $p = 0.0002$) but not on Overall MSRF (F (5.228) $= 1.176$, $p = 0.3215$). The interaction between condition and Clip was not significant in any of the analyses, suggesting that the effects of Knowledge manipulation on Overall, Cognitive, and Affective MSRF did not vary across different clips. These preliminary tests were conducted to ensure that the MSRF was consistent across the six clips included in the study. Partial eta-squared (η_p^2) effect sizes for Overall, Cognitive, and Affective MSRF were 0.02, 0.10, and 0.04 for Knowledge condition and 0.03, 0.05, and 0.10 for Clip, respectively.

3.3. Hypothesis 2: DIcomp and MSRF in context vs. exploitation phase across groups

To test H2 (higher MSRF in Exploitation phase than in the Context only in the Installation group and not in the Control), we conducted a 2x2 mixed ANOVA assessing the effect of knowledge manipulation (Installation vs. Control) on mental state reference frequency (MSRF) in Exploitation vs. Establisher phases. Overall MSRF and Affective MSRF were normally distributed, while Cognitive MSRF was slightly skewed but reasonably symmetrical (see Supplementary Figure) to conduct ANOVA which is robust to non-normality with a large enough sample size. We conducted the same analysis transforming the data by squared rooting the cognitive MSRF values, to check for consistency as recommended by Field et al. (2012). All effects are reported significant at a p-value of < 0.05.

For H2.1 (Overall MSFR), there was a significant main effect of Knowledge condition (F (1.39) $= 4.67$, $p = 0.036$, $\eta_p^2 = 0.10$) and Phase (F (1.39) $= 18.67$, $p < 0.001$, $\eta_p^2 = 0.32$), with a significant interaction (F (1.39) $= 22.90$, $p < 0.001$, $\eta_p^2 = 0.37$). As shown in Figure 4, this significant interaction shows that participants in the Installation group used significantly more Overall MSRF when describing the Exploitation phase compared to Context phase; however, this was not the case for participants in the Control group.

For H2.2 (cognitive mental states), there was a significant main effect of Knowledge manipulation (F (1.39) $= 15.42$, $p < 0.001$, $\eta_p^2 = 0.28$) and Phase (F (1.39) $= 5.57$, $p = 0.023$, $\eta_p^2 = 0.21$), with a significant interaction (F (1.39) $= 34.12$, $p < 0.001$, $\eta_p^2 =$ 0.47), as shown in Figure 5. After transforming Cognitive MSRF, the main effect of Phase was no longer significant (F (1.39) $=$ 2.97, $p = 0.09$, $\eta_p^2 = 0.07$). The significant interaction between Knowledge manipulation X Phase reveals that participants in the Installation condition used significantly more cognitive MSRF when describing Exploitation phase compared to Context phase, whereas this difference was not observed in the Control condition.

For H2.3 (Affective MSRF), there was a main effect of Phase (F (1.39) $= 5.57$, $p < 0.05$, $\eta_p^2 = 0.30$) but no significant main effect of Knowledge manipulation (F (1.39) $= 3.14$, $p = 0.08$, $\eta_p^2 = 0.07$) or interaction (F (1.39) $= 0.53$, $p = 0.47$, $\eta_p^2 = 0.01$). Importantly, these results, depicted in Figure 6, suggest that participants in the Installation condition did not use more Affective MSRF when describing Exploitation phase compared to Context phase, and neither did the Control participants.

3.4. Supplementary analysis examining relationship between DIcomp and MSRF

In the final step to understand the complementary relationship between the two dependent variables measured and their individual differences, we focused on the Installation group. We analyzed

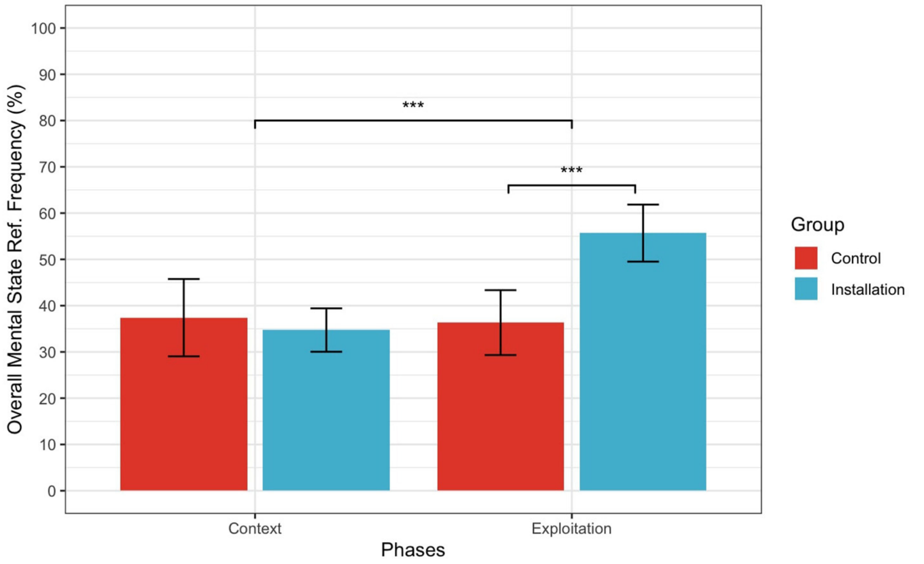


FIGURE 4
Interaction between Phase and Condition for Overall Mental State Frequency. Error bars represent 95% confidence intervals. *Indicates a *p*-value less than 0.05, **Indicates a *p*-value less than 0.01, ***Indicates a *p*-value less than 0.001.

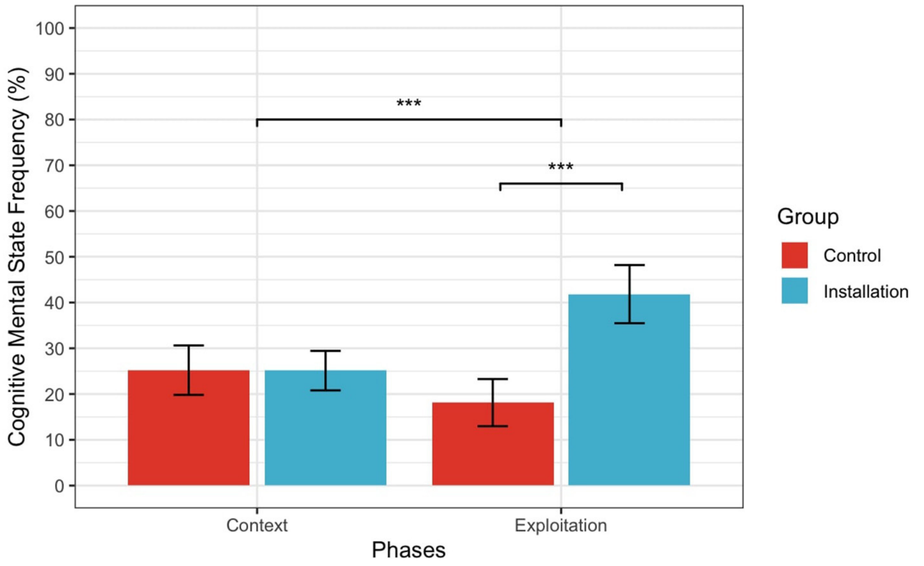


FIGURE 5
Interaction between Phase and Condition for Cognitive Mental State Frequency. Error bars represent 95% confidence intervals. *Indicates a *p*-value less than 0.05, **Indicates a *p*-value less than 0.01, ***Indicates a *p*-value less than 0.001.

whether viewers who watched the installation scene and either partially or fully understood the dramatic irony conflict used higher Overall, Cognitive, and Affective MSRF in their descriptions across clips compared to those who failed to understand it. We also examined whether this varied depending on the Context or Exploitation phase for each clip.

To investigate this, we classified participants from the Installation group based on their DIcomp levels and performed a series of linear mixed effects models using the lme4 package (Bates et al., 2009). The results of these LME models are shown in Table 3. We used a random intercept linear mixed effects model, nesting six measurement occasions (one per clip) of MSRF (Overall, Cognitive, and Affective) within each participant and including the DIcomp scores (failed, partial, or full understanding) as categorical predictors and their interaction. While we considered the possible total scores for dramatic irony conflict comprehension (ranging from 0 to 12) for the manipulation check, in the LME models, each clip was given a score of 0, 1, or 2 separately, allowing us to maintain the categorical nature of comprehension levels while still enabling us to analyze the relationship between DI Comprehension

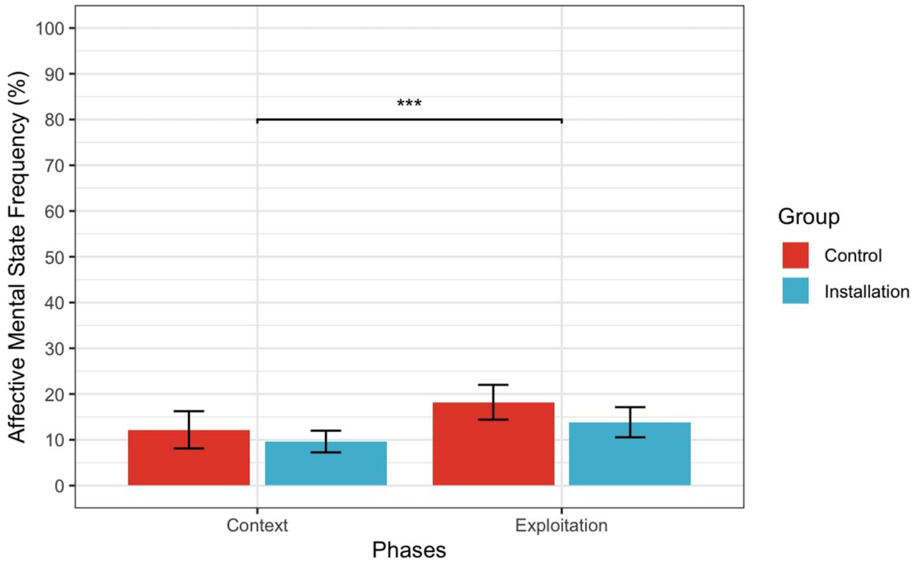


FIGURE 6
Interaction between Phase and Condition for Affective Mental State Frequency. Error bars represent 95% confidence intervals. *Indicates a p-value less than 0.05, **Indicates a p-value less than 0.01, ***Indicates a p-value less than 0.001.

and MSRF in a more nuanced manner. We built all models step by step to examine the effect of including the different terms in explaining mental state frequencies.

We included the variable "Clip" to check whether it was necessary to control for clip-level variables. The analysis for all DVs confirmed that adding "Clip" as a random intercept did not explain more variance and the fit of the model was worse for Overall MSRF (BIC: 2416.8 vs. 2411.3, $p = 1$); however, it was significant for Cognitive MSRF (BIC: 2364.8 vs. 2369.1, $p = 0.002$) and Affective MSRF (BIC: 2095.6 vs. 2118.4, $p < 0.001$). These results suggest that the measured DVs do vary across clips, indicating that some participants and clips tended to produce more MS references than others when considering the types of mental states independently, but this variation is balanced out when considering Overall MSFR.

The results of the LME for Overall MSRF revealed a significant intercept ($p < 0.001$), indicating that participants used mental state references to describe the clips, even when they did not understand the DI conflict. Critically, the effect of DIcomp on Overall MSRF was not significant for partial DI comprehension ($p = 0.752$), but it was for full DI comprehension ($p = 0.047$), suggesting that fully understanding DI conflict can predict a higher Overall MSRF across the whole description. The main effect of Phase on Overall MSRF was not significant ($p = 0.514$). However, as depicted Figure 7, the interaction between DIcomp and Phase [Exploitation] indicated a significant increase in the use of Overall MSRF during the Exploitation phase for participants with partial DI comprehension ($p = 0.019$) but not for full DI comprehension ($p = 0.086$). This suggests that partially understanding the conflict can predict a higher use of Overall MSRF when focusing only on the Exploitation phase.

Similarly, the LME for Cognitive MSRF found a significant intercept ($p = 0.005$), indicating that participants in the Installation group used more Cognitive MSRF, on average, when describing the clips even when they showed failed understanding (see Figure 8). However, there was no significant effect of DIcomp on Cognitive MSRF. The main effect of Phase was not significant for Cognitive MSRF, but the interaction between DIcomp and Phase [Exploitation] was significant for Partial DIcomp ($p = 0.004$), suggesting that the use of mental state references was more frequent when participants were describing the Exploitation compared to the Context of the clips for those with partial DI comprehension but not with full DI comprehension ($p = 0.061$).

Lastly, the LME for Affective MSRF did not find a significant intercept ($p = 0.111$), indicating that participants exposed to the installation scenes did not use significant Affective MSRF, on average, in describing the clips. As depicted in Figure 9, the effect of DIcomp on Affective MSRF was not significant, and the main effect of Phase was also not significant for Affective MSRF. The non-significant interaction term suggests that the effect of the Exploitation phase on Affective MSRF is not significantly different for participants with failed comprehension scores from those with partial ($p = 0.690$) or full DI comprehension scores ($p = 0.902$).

4. Discussion

In this proof-of-concept study, we presented a novel approach to explore spontaneous theory of mind (SToM) using film structures. In particular, we used dramatic irony structures that prompt viewers to attribute ignorance and false beliefs to film characters. Our design allowed us to compare true- vs. false-belief spontaneous attribution scenarios in naturalistic settings. Our findings show that when participants watched the full dramatic irony clip with the *installation* scene, they understand the dramatic irony conflict more often than when they did not watch the scene, demonstrating that the

TABLE 3 Linear mixed effects models of the effect of dramatic irony comprehension on mental state reference frequency for the Installation group.

	Overall MS references			Cognitive MS references			Affective MS references		
Predictors	Estimates	CI	*p*	Estimates	CI	*p*	Estimates	CI	*p*
(Intercept)	24.07	11.39 – 36.76	**<0.001**	18.99	5.64 – 32.34	**0.005**	7.31	−1.70 – 16.31	0.111
Partial DI comp	2.92	−15.26 – 21.10	0.752	3.66	−12.96 – 20.27	0.665	−4.10	−14.46 – 6.25	0.436
Full DI comp	13.70	0.20 – 27.21	**0.047**	7.58	−5.22 – 20.39	0.244	3.67	−4.29 – 11.62	0.365
Phase [Explo]	5.59	−11.23 – 22.41	0.514	1.53	−13.26 – 16.32	0.839	4.06	−5.44 – 13.56	0.401
Partial DI comp × Phase [Explo]	29.91	4.89 – 54.94	**0.019**	32.78	10.77 – 54.79	**0.004**	−2.87	−17.00 – 11.27	0.690
Full DI comp × Phase [Explo]	15.96	−2.30 – 34.23	0.086	15.32	−0.74 – 31.38	0.061	0.64	−9.67 – 10.96	0.902
Random effects									
σ^2	619.72			479.19			197.64		
τ_{00}	17.01 $_{\text{Participant\_num}}$			41.47 $_{\text{Participant\_num}}$			0.92 $_{\text{Participant\_num}}$		
	8.43 $_{\text{Clip}}$			57.05 $_{\text{Clip}}$			44.39 $_{\text{Clip}}$		
N	21 $_{\text{Participant\_num}}$			21 $_{\text{Participant\_num}}$			21 $_{\text{Participant\_num}}$		
	6 $_{\text{Clip}}$			6 $_{\text{Clip}}$			6 $_{\text{Clip}}$		
Observations	252			252			252		
Marginal R^2/Conditional R^2	0.216/0.247			0.170/0.312			0.056/0.232		

Bold values denote significance at the level $p \leq 0.05$.

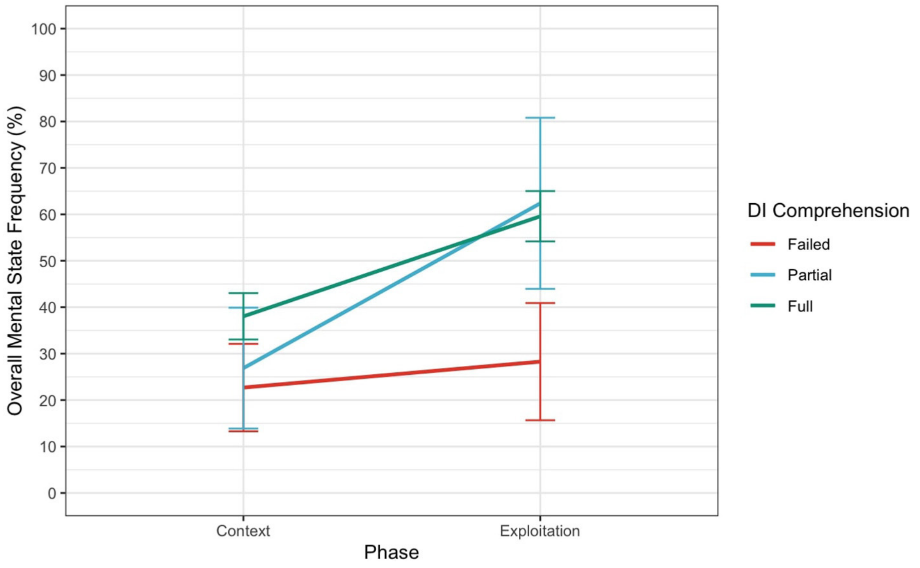


FIGURE 7
Interaction between Phase and DIcomp for Overall Mental State Frequency. Error bars represent 95% confidence intervals.

exposure to this particular scene was required to understand dramatic irony.

The results showed that overall, participants in the Installation group used a higher Cognitive MSRF than the Control group but not a higher Affective MSRF, which remained similar across conditions. These results imply that salient divergence of knowledge in dramatic irony prompted participants to spontaneously refer to characters' epistemic mental states such as belief and knowledge, rather than affective mental states, in their character models.

Moreover, when controlling for content delivered and focusing on the scenes that both groups had seen, the Installation group used a higher Overall and Cognitive MSRF in their descriptions of the Exploitation phase than the Control group but showed no difference in the Context phase. While it was reasonable to expect participants in the Installation condition to use a high

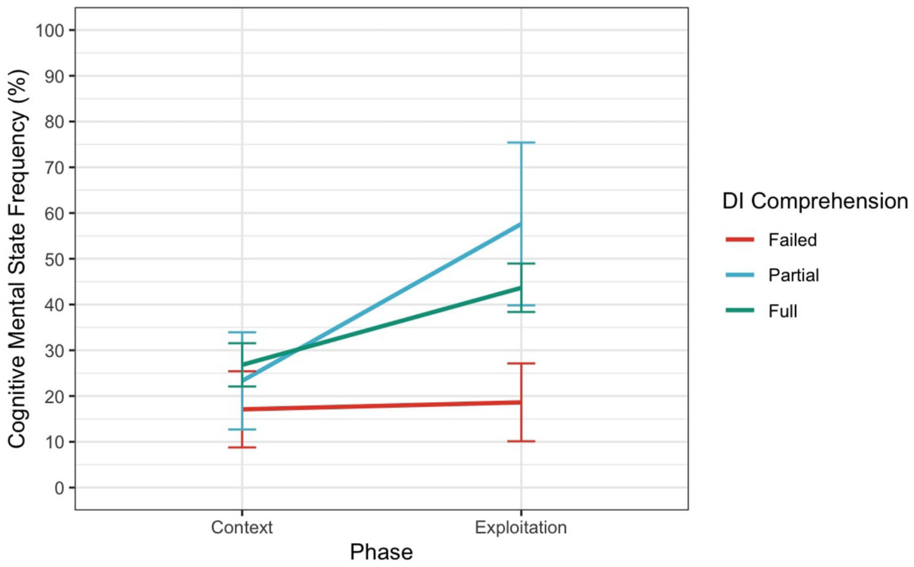


FIGURE 8
Interaction between Phase and DIcomp for Cognitive Mental State Frequency. Error bars represent 95% confidence intervals.

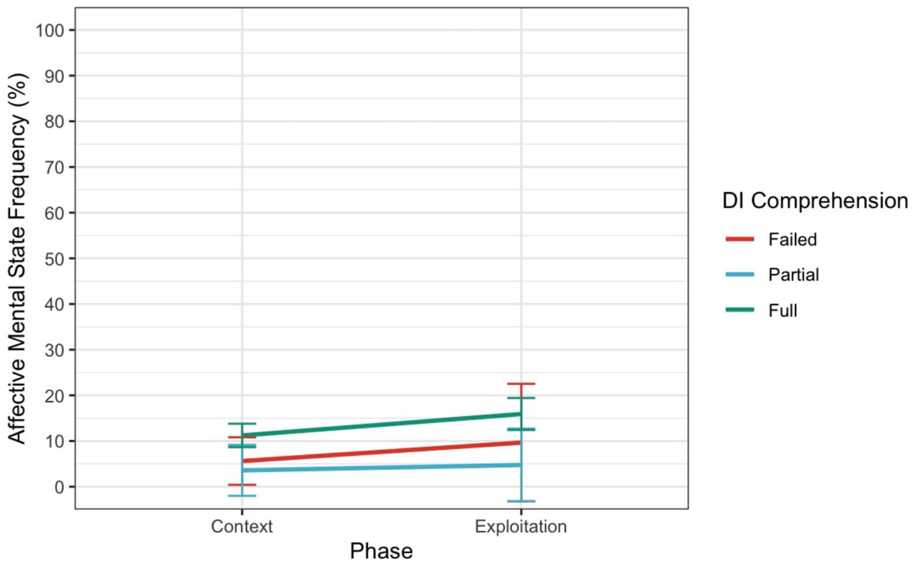


FIGURE 9
Interaction between Phase and DIcomp for Affective Mental State Frequency. Error bars represent 95% confidence intervals.

frequency of mental states to describe the Exploitation phase (given that this is where the dramatic irony conflict happens), it was also possible that these participants' event representation and retrospective recollection of the *establisher* scenes would be affected by the *installation* scene and dramatic irony conflict, prompting them to use a high frequency of mental states in the description of the Context phase. This was not the case. The present findings suggest that the salient disparity of knowledge in dramatic irony scenes prompts viewers to infer the cognitive mental states of characters and specifically when they access and retrieve event models of the *exploitation* scenes. In this regard, we also recognize that participants in the Installation condition had more events and characters to describe, which may have led to fatigue. However, if fatigue was to play a significant role, we would expect the Installation group to engage less with the mental states of characters, as the literature suggests that fatigue can negatively impact social problem-solving and empathic responding (Nelson et al., 2003; Nelson, 2018). This hypothesis would lead us to anticipate the opposite pattern of results to what we have found in our study. Moreover, given this fatigue bias, we might expect the Control group, who had less content to describe in the context scenes compared to the Installation group, to engage more with the mental states of characters due to the lack of fatigue. Yet, our findings do not support this prediction.

Finally, we investigated individual differences in DI comprehension in those participants that had watched the complete dramatic irony version and how this comprehension relates to mental state reference frequency. Our findings suggest that by default, participants used Overall and Cognitive but not Affective MSRF to describe the clips, even when they failed to understand the DI conflict. Moreover, we found that Overall MSRF increased with full understanding of DI conflict. Interestingly, partial understanding of DI conflict also predicted higher Overall and Cognitive MSRF (but not Affective MSFR) when participants described the Exploitation phase vs.s when they described the Context phase. Our findings suggest that participants tended to use a similar amount of cognitive and affective mental state terms when they watched scenes where they knew the same amount of information as the characters (as indicated by the significant intercepts in a second set of LMEs carried out for the Control group; see Supplementary material for this table). However, when participants were exposed to critical knowledge that a character did not possess, they describe the event in terms of their cognitive mental states, but they do not significantly refer to their affective mental states. Interestingly, this was the case even when they do not show understanding of dramatic irony. This suggests that the extent to which we focus on affective vs. cognitive mental states in describing a scene depends on our level of knowledge relative to the characters in the scene.

Taken together, these findings highlight the importance of differentiating between cognitive and affective mental states when measuring SToM through mental state reference frequencies. Participants use different frequencies of these two types of mental states depending on their comprehension, i.e., on the event model they built of the situation, thus implying that their differentiation is key for our understanding of the nature of SToM responses in dramatic irony scenes. This perspective aligns with the work of authors such as Shamay-Tsoory and Aharon-Peretz (2007), who found evidence for an anatomical distinction between affective and cognitive ToM processing, suggesting that they rely on partially separate anatomical substrates.

Similar to how verbal irony paradigms have been extensively used to probe the relationship between language processes and theory of mind and their individual differences (e.g., Hancock et al., 2000; Filippova and Astington, 2008; Nilsen et al., 2011; Pexman et al., 2011), dramatic irony in film can constitute a powerful tool for investigating social cognition processes in the domain of visual event perception and comprehension. Moreover, the present dramatic irony film corpus offers a unique opportunity to examine the underlying sociocognitive processes involved in event perception and comprehension in several ways.

First, it can allow us to examine whether and how individuals attribute knowledge or epistemic states in the *installation* scenes of dramatic irony. In particular, when labeling characters as ignorant, we attribute them a lack of knowledge due to either lack of perceptual access or due to other contextual factors such as repression, denial, or mental illnesses (Lavandier, 2005). Traditional ToM stimuli such as the Sally–Anne task and dramatic irony in film have in common that they both clearly cue lack of perceptual access; however, they achieve this in different ways. The Sally–Anne task manifestly shows a character leaving a room to signpost their lack of presence in the scene, while dramatic irony, which could be considered a cinematic version of the mentioned task, implies who has access to information through subtler means, embedded in a continuous, intricate narrative The complexity of ToM reasoning in these clips, as seen in the descriptions contained in Table 1, is much greater compared to the relatively straightforward Sally–Anne task. To understand who is ignorant in these scenes, viewers are inadvertently motivated to track knowledge as part of the broader event sequence. In silent films like the Harold Lloyd films chosen for our film corpus, long or medium-long shots were often used due to the composition trends of the time, where people walked in and out of the scenes similar to theater performances. In modern films, knowledge is indicated through cinematic techniques which are familiar to the viewer, such as flashbacks scenes, camera cutaways to different locations, point-of-view shots, or carefully choreographed staging in which character knowledge of foreground events may differ to background events. Other cinematic techniques include editing, which can shape the audience's experience of the narrative, and lighting, focusing, or camera movements, which can be used to direct the audience's attention to specific details within a shot.

Despite the absence of color and sound, as well as the presence of intertitles in these clips, the visual storytelling in these silent films relies heavily on facial expressions, body language, and context, which are essential components of real-life social interactions. These elements allow viewers to make inferences about characters' mental states, intentions, and emotions, thus providing a valuable stimulus for studying ToM processing. Additionally, while film editing techniques may not be naturalistic representations of reality, they do serve to guide and shape viewers' cognitive processes, allowing for the examination of how these cinematic choices influence ToM processing (Cabañas et al., 2022; Grall and Finn, 2022). In fact, the stylized nature of the stimulus may help to focus participants' attention on specific aspects of the narrative and ToM processing, allowing us to isolate these processes to study them with a degree of ecological validity.

It is essential to consider potential limitations in applying these findings to real-life social interactions due to unique cinematic elements and differing participant identification or empathy with characters. Nevertheless, film viewing offers a valuable opportunity to study cognitive processes, as it allows for spontaneous processing in a controlled environment that is challenging to achieve in real-world scenarios or virtual reality, where participants have a unique experience as they choose what is within their field of view at any moment. This controlled setting allows for a more uniform investigation of mental processes among participants. Nonetheless, future research should examine the relevance of our findings to real-world contexts and the potential limitations of generalizing results from film viewing to real-life social interactions.

Second, the knowledge manipulation in these clips seeks to address scenarios that induce either false beliefs (in the Installation condition) or true beliefs (in the Control condition), thereby creating differences in the divergence in the interpretations of the shared content (the Exploitation phase) based on prior knowledge, allowing us to compare how viewers process the same events depending on their false- or true-belief representations. In addition, this comparison speaks to the debate of automaticity and modularity in belief attribution [e.g., (Fodor, 1992; Leslie and Thaiss, 1992; Back and Apperly, 2010)]. Critically, the current findings suggest that false-belief inducing scenarios prompt a richer cognitive mental state representation than true-belief inducing scenarios accessed by participants in their free recall answers. In line with these results which suggest increased complexity of false-belief representations, there is evidence that adults are slower making judgements when a character had a false belief rather than a true belief (German and Hehman, 2006). Moreover, Phillips et al. (2011) found that while elderly adults perform similar to young adults on true-belief tasks, they perform worse in false-belief tasks. However, the mentioned studies often use explicit and repetitive ToM paradigms, which, as suggested previously, may not capture the full picture of everyday ToM processing. Employing measures that address cognitive effort to compare the sociocognitive processing of both dramatic irony clip versions could allow us to understand better whether and how false beliefs are more cognitively effortful than true beliefs in spontaneous mentalizing in a more ecologically valid context.

Note that the current results showed room for refinement for our research paradigm, in particular the choice of film stimuli. Notably, there was variation in the DI conflict comprehension scores, specifically for Clip 1, which resulted in no significant differences in DI comprehension between the groups, and Clip 5, which was similarly not well-understood in both groups. In Clip 1, formal cinematic cues may have helped participants understand dramatic irony even without the installation scene, while the absence of such cues in Clip 5 may have hindered its comprehension. Additionally, our coding scheme required participants to identify both ignorance and consequences for the victims of dramatic irony. However, we observed a "protagonist effect," where descriptions focused on the protagonist even when they were not the victim of dramatic irony, leading to insufficient criteria for DI comprehension. This is similar to Hutson et al. (2017), who found the "agent effect," where eye movements appear to be influenced by whether viewers perceive a character in the narrative as an agent or just a character who appears in the background without playing a significant role. Addressing these issues is crucial for appropriate comparison between Installation and Control conditions.

A potential limitation to address is that, precisely due to the expected intrinsic power of dramatic irony to motivate participants to attribute mental states to characters, we expected a certain overlap between comprehension of DI conflict and mental state scores. While both measures involve language-based descriptions of mental states, they are distinct constructs that do not necessarily depend on each other. Moreover, note that the coding scheme, based on Barnes and Baron-Cohen (2012) and the definition of dramatic irony conflict by professional scriptwriter theorist Lavandier (2005), did not require participants to use mental states to be scored as partial and full understanding. It is possible for participants to have a good understanding of the DI conflict without necessarily mentioning the victim's mental states in their descriptions. An example of a description scoring partial understanding without using mental state words is "the publisher exchanges the rejection letter for a check at the last moment." A description demonstrating full without mental state words might be "*Harold rips up the envelope containing the cheque instead of a rejection letter.*" Conversely, a participant might use numerous mental state references without necessarily having a good understanding of the dramatic irony conflict. For instance, answers such as "*Harold was very disappointed at himself, he felt a failure as a writer*" would not be awarded any points. Moreover, in the present study, we found that participants produced higher Overall and Cognitive MSRF when they partially understood the conflict than when they fully understood it, remarking that DIcomp and MSRF do not co-vary linearly, speaking to the fact that these measures address different aspects of dramatic irony processing.

In short, while there may be certain overlap between the comprehension measures and mental state references frequency, the results of the present study suggest that these are distinct constructs that do not necessarily depend on each other. The DIcomp measure provides valuable insights into individual differences in the ability to comprehend scenes and can be particularly useful in investigating the cognitive and affective processes involved in this type of scene comprehension. On the contrary, the frequency of mental state references demonstrates the extent to which individuals integrate mentalistic attributions into their character models. Although this often results in inferences about the consequences of false beliefs, it is not a strict requirement for comprehending the DI conflict. By examining both the understanding of the DI conflict and the frequency of mental state references, we obtain a more nuanced multi-dimensional of how SToM is involved in the comprehension of dramatic irony.

We recognize the potential influence of individual differences, especially in verbal abilities, on mental state descriptions. Although we did not explicitly control for verbal abilities, all participants possessed university-level English proficiency, establishing a baseline for language skills and minimizing the impact of individual differences. To further account for individual differences in verbal production, we adjusted for the length of participants' descriptions by dividing the number of mental state references by the number of coding units. Notably, differences in mental state references were observed within subjects when comparing Context and Exploitation phases, which mitigates the potential influence of individual differences in verbal abilities on our results. In future research, we plan to address individual differences more specifically, such as atypical theory of mind skills as it is common in ASC, and verbal abilities that may impact mental state descriptions, to better understand their role in our findings and enhance the generalizability of our results.

In future studies, the present dramatic irony film corpus and measures could be used in conjunction with other measures of SToM, such as eye tracking, psychophysiological monitoring of affective states and arousal, or functional neuroimaging during the processing of dramatic irony. This would enable researchers to identify individual differences in mentalizing processes, affective states, and help isolate neural structures responsible in the moment-to-moment false- vs. true-belief spontaneous attribution, such as the much-debated involvement of the medial prefrontal cortex (mPFC) in SToM (Bardi et al., 2016; Moessnang et al., 2017; Boccadoro et al., 2019). As already mentioned, these narrative devices are pervasive in cinema; therefore, researchers can identify these structures in open datasets to take advantage of free-viewing SToM paradigms (Eickhoff et al., 2020).

In concurrent work with this film corpus, we are exploring the role of supporting cognitive processes of ToM, such as working memory and attention, in the comprehension of dramatic irony. By examining the interplay between SToM and these cognitive processes, we aim to provide a more comprehensive understanding of the factors contributing to the successful processing and appreciation of dramatic irony in narrative contexts. By examining changes in eye movements and neural activity during the processing of the present dramatic irony film corpus, researchers could gain a better understanding of how SToM operates in real time during social interactions. On this note, we believe it is important to highlight that eye tracking, physiological, and functional neuroimaging techniques proposed should be used in conjunction with comprehension measures as the ones presented. As stated in Cabañas et al. (2022), simply investigating eye movements or brain activity without using additional comprehension measures to correlate with the observed brain response should not be viewed as conclusive evidence of viewers' cognitive representations, thereby constituting a type of reverse inference (Poldrack, 2006).

Moreover, our dramatic irony film corpus could be employed in research investigating developmental differences in SToM processing, by comparing how children, adolescents, and adults process and understand dramatic irony. This line of research would provide valuable insights into the development of mentalizing skills and social understanding across different age groups. The stimuli, procedures, and coding handbook used in our study are available upon request, making it accessible for researchers interested in further exploring this topic.

Lastly, our film corpus and measures could be adapted to examine the efficacy of interventions aimed at improving atypical ToM processing, such as in individuals with ASC. By using the film corpus as a tool to measure the effectiveness of these interventions, researchers can assess the real-world applicability of the developed strategies and their potential for improving social understanding in everyday life.

5. Conclusion

The implications of the present proof-of-concept study are three-fold. First, it demonstrates that the degree to which individuals emphasize affective vs. cognitive mental states is influenced by their level of knowledge relative to that of the characters in the scene. Second, our study highlights the utility of a novel corpus of dramatic irony film scenes as a means of investigating social cognition in ecologically valid contexts, enabling us to address knowledge, true- and false-belief attributions. Third, we provide measures of comprehension and mental state attribution, which address complementary aspects of social processing in scene perception and event comprehension, essential to allow the exploration of links between different levels of cognitive processing and eye tracking or neural dynamics. The integration of these measures with other techniques could have important implications for our understanding of moment-to-moment SToM and the neural underpinnings of social processing.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Birkbeck, University of London Ethics Board (181949). The patients/participants provided their written informed consent to participate in this study.

Author contributions

CC, AS, and TS contributed to conception and design of the study. CC collected the data, performed the statistical analysis, and wrote the first draft of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version.

Funding

This study was supported by an Economic and Social Research Council (ESRC) PhD studentship awarded to CC. ESRC Grant Reference: ES/P000592/1.

Acknowledgments

We thank the Harold Lloyd Trust for granting permission to use the mentioned clips from Harold Lloyd films for research and educational purposes. © 2023 Harold Lloyd Entertainment, Inc. We thank Anna Mozhina, for her assistance with film clip selection and description coding. We also thank Ben Stinson and Ella Edwards for their assistance with coding.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1183660/full#supplementary-material

References

Abell, F., Happe, F., and Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cogn. Dev.* 15, 1–16. doi: 10.1016/S0885-2014(00)00014-9

Abu-Akel, A., and Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia*, 49, 2971–2984. doi: 10.1016/j.neuropsychologia.2011.07.012

Achim, A. M., Guitton, M., Jackson, P. L., Boutin, A., and Monetta, L. (2013). On what ground do we mentalize? Characteristics of current tasks and sources of information that contribute to mentalizing judgments. *Psychol. Assess.* 25, 117. doi: 10.1037/a0029137

Altschuler, M., Sideridis, G., Kala, S., Warshawsky, M., Gilbert, R., Carroll, D., et al. (2018). Measuring individual differences in cognitive, affective, and spontaneous theory of mind among school-aged children with autism spectrum disorder. *J. Autism Dev. Disord.* 48, 3945–3957. doi: 10.1007/s10803-018-3663-1

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. K. (2020). Gorilla in our midst: an online behavioral experiment builder. *Behav. Res. Methods* 52, 388–407. doi: 10.3758/s13428-019-01237-x

Apperly, I. (2011). Mindreaders: the cognitive basis of. *Theory Mind* 3, 1093–1108. doi: 10.4324/9780203833926

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *Quart. J. Exp. Psychol.* 65, 825–839. doi: 10.1080/17470218.2012.676055

Astington, J. W., and Gopnik, A. (1991). Theoretical explanations of children's understanding of the mind. *Br. J. Develop. Psychol.* 9, 7–31. doi: 10.1111/j.2044-835X.1991.tb00859.x

Back, E., and Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition* 115, 54–70. doi: 10.1016/j.cognition.2009.11.008

Baillargeon, R., Scott, R. M., and He, Z. (2010). False-belief understanding in infants. *Trends Cogn. Sci.* 14, 110–118. doi: 10.1016/j.tics.2009.12.006

Bálint, K., Klausch, T., and Pólya, T. (2018). Watching closely: Shot scale influences theory of mind response in visual narratives. *J. Media Psychol. Theor. Method. Appl.* 30, 150–159. doi: 10.1027/1864-1105/a000189

Bardi, L., Desmet, C., Nijhof, A., Wiersema, J. R., and Brass, M. (2016). Brain activation for spontaneous and explicit false belief tasks overlaps: new fMRI evidence on belief processing and violation of expectation. *Soc. Cogn. Affect. Neurosci.* 3, 143. doi: 10.1093/scan/nsw143

Barnes, J. L., and Baron-Cohen, S. (2012). The big picture: storytelling ability in adults with autism spectrum conditions. *J. Autism Develop. Disord.* 42, 1557–1565. doi: 10.1007/s10803-011-1388-5

Barnes, J. L., Lombardo, M. V., Wheelwright, S., and Baron-Cohen, S. (2009). Moral dilemmas film task: a study of spontaneous narratives by individuals with autism spectrum conditions. *Autism Res.* 2, 148–156. doi: 10.1002/aur.79

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., et al. (2009). *Package 'lme4'*. Available online at: http://lme4.r-forge.r-project.org

Bernstein, D. M., Thornton, W. L., and Sommerville, J. A. (2011). Theory of mind through the ages: older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Exp. Aging Res.* 37, 481–502. doi: 10.1080/0361073X.2011.619466

Birch, S. A. J., and Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychol. Sci.* 18, 382–386. doi: 10.1111/j.1467-9280.2007.01909.x

Bloom, P., and German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* 77, B25–B31. doi: 10.1016/S0010-0277(00)00096-2

Boccadoro, S., Cracco, E., Hudson, A. R., Bardi, L., Nijhof, A. D., Wiersema, J. R., et al. (2019). Defining the neural correlates of spontaneous theory of mind (ToM): an fMRI multi-study investigation. *Neuroimage* 203, 116193. doi: 10.1016/j.neuroimage.2019.116193

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychon. Bullet. Rev.* 16, 893–900. doi: 10.3758/PBR.16.5.893

Cabañas, C., Senju, A., and Smith, T. J. (2022). Dramatic irony: a case study in the mutual benefit of combining social neuroscience with film theory. *Projections* 16, 84–104. doi: 10.3167/proj.2022.160106

Castelli, F., Happé, F., Frith, U., and Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage* 12, 314–325. doi: 10.1006/nimg.2000.0612

Cohn, N. (2016). "From visual narrative grammar to filmic narrative grammar: the narrative structure of static and moving images," in *Film Text Analysis* (pp. 94–117). Routledge.

Cuff, B. M. P., Brown, S. J., Taylor, L., and Howat, D. J. (2016). Empathy: a review of the concept. *Emotion Rev.* 8, 144–153. doi: 10.1177/1754073914558466

Devine, R. T., and Hughes, C. (2013). Silent films and strange stories: Theory of mind, gender, and social experiences in middle childhood. *Child Dev.* 84, 989–1003. doi: 10.1111/cdev.12017

Dufour, N., Redcay, E., Young, L., Mavros, P. L., Moran, J. M., Triantafyllou, C., et al. (2013). Similar brain activation during false belief tasks in a large sample of adults with and without autism. *PLoS ONE* 8, e75468. doi: 10.1371/journal.pone.0075468

Dziobek, I. (2012). Comment: towards a more ecologically valid assessment of empathy. *Emot. Review* 4, 18–19. doi: 10.1177/1754073911421390

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., et al. (2006). Introducing MASC: A movie for the assessment of social cognition. *J. Autism Dev. Disord.* 36, 623–636. doi: 10.1007/s10803-006-0107-0

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., et al. (2008). Dissociation of cognitive and emotional empathy in adults with asperger syndrome using the multifaceted empathy test (MET). *J. Autism Dev. Disord.* 38, 464–473. doi: 10.1007/s10803-007-0486-x

Eickhoff, S. B., Milham, M., and Vanderwal, T. (2020). Towards clinical applications of movie fMRI. *Neuroimage* 217, 116860. doi: 10.1016/j.neuroimage.2020.116860

Faul, F., Erdfelder, E., Buchner, A., and Lang, A-. G. (2009). Statistical power analyses using G* Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149

Field, A., Miles, J., and Field, Z. (2012). *Discovering Statistics Using R*. London: SAGE Publications.

Filippova, E., and Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Dev.* 79, 126–138. doi: 10.1111/j.1467-8624.2007.01115.x

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition* 44, 283–296. doi: 10.1016/0010-0277(92)90004-2

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., Frith, C. D., et al. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind'in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21. doi: 10.1016/S0028-3932(99)00053-6

German, T. P., and Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': evidence from compromised belief-desire reasoning in old age. *Cognition*, 101, 129–152. doi: 10.1016/j.cognition.2005.05.007

Golan, O., Baron-Cohen, S., Hill, J. J., and Golan, Y. (2006). The "reading the mind in films" task: Complex emotion recognition in adults with and without autism spectrum conditions. *Soc. Neurosci.* 1, 111–123. doi: 10.1080/17470910600980986

Gopnik, A., and Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Dev.* 59, 26–37. doi: 10.2307/1130386

Grall, C., and Finn, E. S. (2022). Leveraging the power of media to drive cognition: a media-informed approach to naturalistic neuroscience. *Soc. Cogn. Affect. Neurosci.* 17, 598–608. doi: 10.1093/scan/nsac019

Hancock, J. T., Dunham, P. J., and Purdy, K. (2000). Children's comprehension of critical and complimentary forms of verbal irony. *J. Cogn. Develop.* 1, 227–248. doi: 10.1207/S15327647JCD010204

Heavey, L., Phillips, W., Baron-Cohen, S., and Rutter, M. (2000). The awkward moments test: a naturalistic measure of social understanding in autism. *J. Autism Dev. Disord.* 30, 225–236. doi: 10.1023/A:1005544518785

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspect. Psychol. Sci.* 9, 131–143. doi: 10.1177/1745691613518076

Hitchcock, A. (Director). (1960). *Psycho. Paramount Pictures.*

Hutson, J. P., Smith, T. J., Magliano, J. P., and Loschky, L. C. (2017). What is the role of the film viewer? The effects of narrative comprehension and viewing task on gaze control in film. *Cogn. Res. Princ. Implicat.* 2, 46. doi: 10.1186/s41235-017-0080-5

Kinderman, P., Dunbar, R., and Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *Br. J. Psychol.* 89, 191–204. doi: 10.1111/j.2044-8295.1998.tb02680.x

Klin, A. (2000). Attributing Social Meaning to Ambiguous Visual Stimuli in Higher-functioning Autism and Asperger Syndrome: The Social Attribution Task. *J. Child Psychol Psychiatry Allied Discip.* 41, 831–846. doi: 10.1111/1469-7610.00671

Krippendorff, K. (2004). Reliability in content analysis: some common misconceptions and recommendations. *Hum. Commun. Res.* 30, 411–433. doi: 10.1111/j.1468-2958.2004.tb00738.x

Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., et al. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage.* 100, 316–324.

Lavandier, Y. (2005). *Writing Drama: A Comprehensive Guide for Playwrights and Scriptwriters.* Le Clown and l'Enfant.

Leslie, A. M., and Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition* 43, 225–251. doi: 10.1016/0010-0277(92)90013-8

Levin, D. T., Hymel, A. M., and Baker, L. (2013). "Belief, desire, action, and other stuff: theory of mind in movies," in *Psychocinematics: Exploring cognition at the movies* (pp. 244–266). Oxford University Press. doi: 10.1093/acprof:oso/9780199862139.003.0013

Livingston, L. A., Carr, B., and Shah, P. (2019). Recent advances and new directions in measuring theory of mind in autistic adults. *J. Autism Dev. Disord.* 49, 1738–1744. doi: 10.1007/s10803-018-3823-3

Loschky, L. C., Larson, A. M., Smith, T. J., and Magliano, J. P. (2020). The scene perception and event comprehension theory (SPECT) applied to visual narratives. *Top. Cogn. Sci.* 12, 311–351. doi: 10.1111/tops.12455

Meins, E., and Fernyhough, C. (2006). *Mind-Mindedness Coding Manual.* Durham: Durham University. [Unpublished].

Moessnang, C., Otto, K., Bilek, E., Schäfer, A., Baumeister, S., Hohmann, S., et al. (2017). Differential responses of the dorsomedial prefrontal cortex and right posterior superior temporal sulcus to spontaneous mentalizing. *Hum. Brain Mapp.* 38, 3791–3803. doi: 10.1002/hbm.23626

Nelson, D. W. (2018). Solving interpersonal problems when fatigued: the beneficial effect of positive mood. *Soc. Behav. Personal. Int. J.* 46, 1489–1497. doi: 10.2224/sbp.7497

Nelson, D. W., Klein, C. T. F., and Irvin, J. E. (2003). Motivational antecedents of empathy: inhibiting effects of fatigue. *Basic Appl. Soc. Psych.* 25, 37–50. doi: 10.1207/S15324834BASP2501_3

Newmeyer, F. C., and Taylor, S. (Directors). (1924). *Girl Shy. Pathé Exchange.*

Newmeyer, F. C., and Taylor, S. (Directors). (1925). *The Freshman. Pathé Exchange.*

Newmeyer, F. C. (Director). (1921). *Never Weaken. Hal Roach Studios.*

Nilsen, E. S., Glenwright, M., and Huyder, V. (2011). Children and adults understand that verbal irony interpretation depends on listener knowledge. *J. Cogn. Develop.* 12, 374–409. doi: 10.1080/15248372.2010.544693

Oesch, N., and Dunbar, R. I. M. (2017). The emergence of recursion in human language: mentalising predicts recursive syntax task performance. *J. Neurolinguistics* 43, 95–106. doi: 10.1016/j.jneuroling.2016.09.008

Pexman, P. M., Rostad, K. R., McMorris, C. A., Climie, E. A., Stowkowy, J., Glenwright, M. R., et al. (2011). Processing of ironic language in children with high-functioning autism spectrum disorder. *J. Autism Dev. Disord.* 41, 1097–1112. doi: 10.1007/s10803-010-1131-7

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., et al. (2021). Knowledge before belief. *Behav. Brain Sci.* 44, e140. doi: 10.1017/S0140525X20000618

Phillips, L. H., Bull, R., Allen, R., Insch, P., Burr, K., Ogg, W., et al. (2011). (2011). Lifespan aging and belief reasoning: influences of executive function and social cue decoding. *Cognition* 120, 236–247. doi: 10.1016/j.cognition.2011.05.003

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn. Sci.* 10, 59–63. doi: 10.1016/j.tics.2005.12.004

Preckel, K., Kanske, P., and Singer, T. (2018). On the interaction of social affect and cognition: Empathy, compassion and theory of mind. *Curr. Opin. Behav. Sci.* 19, 1–6. doi: 10.1016/j.cobeha.2017.07.010

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.*, 1, 515–526. doi: 10.1017/S0140525X00076512

Rice, K., and Redcay, E. (2015). Spontaneous mentalizing captures variability in the cortical thickness of social brain regions. *Soc. Cogn. Affect. Neurosci.* 10, 327–334. doi: 10.1093/scan/nsu081

Rooney, B., and Bálint, K. E. (2018). Watching more closely: shot scale affects film viewers' theory of mind tendency but not ability. *Front. Psychol.* 8, 2349. doi: 10.3389/fpsyg.2017.02349

Ruffman, T., and Perner, J. (2005). Do infants really understand false belief? *Trends Cogn. Sci.* 9, 462–463. doi: 10.1016/j.tics.2005.08.001

Ryskin, R. A., and Brown-Schmidt, S. (2014). Do adults show a curse of knowledge in false-belief reasoning? A robust estimate of the true effect size. *PLOS ONE*, 9, e92406. doi: 10.1371/journal.pone.0092406

Scott, R. M., and Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Develop.* 80, 1172–1196. doi: 10.1111/j.1467-8624.2009.01324.x

Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *Neuroscientist* 18, 108–113. doi: 10.1177/1073858410397208

Shamay-Tsoory, S. G., and Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia* 45, 3054–3067. doi: 10.1016/j.neuropsychologia.2007.05.021

Shany, O., Greental, A., Gilam, G., Perry, D., Bleich-Cohen, M., Ovadia, M., et al. (2021). Somatic engagement alters subsequent neurobehavioral correlates of affective mentalizing. *Hum. Brain Mapp.* 42, 5846–5861. doi: 10.1002/hbm.25640

Smith, M. (1995). *Engaging Characters: Fiction, Emotion, and the Cinema.* Oxford: Clarendon Press.

Sommerville, J. A., Bernstein, D. M., and Meltzoff, A. N. (2013). Measuring beliefs in centimeters: private knowledge biases preschoolers' and adults' representation of others' beliefs. *Child Dev.* 84, 1846–1854. doi: 10.1111/cdev.12110

Stiller, J., and Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. *Soc. Networks* 29, 93–104. doi: 10.1016/j.socnet.2006.04.001

Tan, E. S. (2013). *Emotion and the Structure of Narrative Film: Film as an Emotion Machine.* Routledge. doi: 10.4324/9780203812761

Taylor, S. (Director). (1926). *For Heaven's Sake. Paramount Pictures.*

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304

Wilde, T., Howe, J. A., and Lloyd, H. (Directors). (1927). *The Kid Brother. Paramount Pictures.*

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., et al. (2017). Same story, different story: the neural representation of interpretive frameworks. *Psychol. Sci.* 28, 307–319.

Zaitchik, D. (1990). When representations conflict with reality: the preschooler's problem with false beliefs and "false" photographs. *Cognition* 35, 41–68. doi: 10.1016/0010-0277(90)90036-J

Zaki, J., and Ochsner, K. N. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nat. Neurosci.* 15, 675–680. doi: 10.1038/nn.3085

Zunshine, L. (2006). *Why We Read Fiction: Theory of Mind and the Novel.* Ohio State University Press.

Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162. doi: 10.1037/0033-2909.123.2.162

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 14 July 2023
DOI 10.3389/fnins.2023.1173704

EDITED BY
Celia Andreu-Sánchez,
Autonomous University of Barcelona, Spain

REVIEWED BY
Emilia Zabielska-Mendyk,
The John Paul II Catholic University of Lublin, Poland
Vladimir Kosonogov,
HSE University, Russia

*CORRESPONDENCE
Javier Sanz-Aznar
✉ javier.sanz@ub.edu

RECEIVED 25 February 2023
ACCEPTED 27 June 2023
PUBLISHED 14 July 2023

CITATION
Sanz-Aznar J, Bruni LE and Soto-Faraco S (2023) Cinematographic continuity edits across shot scales and camera angles: an ERP analysis.
Front. Neurosci. 17:1173704.
doi: 10.3389/fnins.2023.1173704

Cinematographic continuity edits across shot scales and camera angles: an ERP analysis

Javier Sanz-Aznar[1]*, Luis Emilio Bruni[2] and Salvador Soto-Faraco[3,4]

[1]Section of Communication, Department of Hispanic Studies, Literary Theory and Communication, University of Barcelona, Barcelona, Spain, [2]Augmented Cognition Lab, Section for Media Technology, Department of Architecture, Design and Media Technology, The Technical Faculty of IT and Design, Aalborg University, Copenhagen, Denmark, [3]Multisensory Research Group, The Center for Brain and Cognition, Pompeu Fabra University, Barcelona, Spain, [4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Film editing has attracted great theoretical and practical interest since the beginnings of cinematography. In recent times, the neural correlates of visual transitions at edit cuts have been at the focus of attention in neurocinematics. Many Event Related Potential (ERP) studies studies have reported the consequences of cuts involving narrative discontinuities, and violations of standard montage rules. However, less is known about edits that are meant to induce continuity. Here, we addressed the neural correlates of continuity editing involving scale, and angle variations across the cut within the same scene, two of the most popular devices used for continuity editing. We recorded the electroencephalographic signal obtained from 20 viewers as they watched four different cinematographic excerpts to extract ERPs at edit points. First, we were able to reproduce the general time and scalp distribution of the typical ERPs to filmic cuts in prior studies. Second, we found significant ERP modulations triggered by scale changes (scale out, scale in, or maintaining the same scale). Edits involving an increase in scale (scale out) led to amplification of the ERP deflection, and scale reduction (scale in) led to decreases, compared to edits that kept scale across the cut. These modulations coincide with the time window of the N300 and N400 components and, according to previous findings, their amplitude has been associated with the likelihood of consciously detecting the edit. Third, we did not detect similar modulations as a function of angle variations across the cut. Based on these findings, we suggest that cuts involving reduction of scale are more likely to go unnoticed, than ones that scale out. This relationship between scale in/out and visibility is documented in film edition manuals. Specifically, in order to achieve fluidity in a scene, the edition is designed from the most opened shots to the most closed ones.

KEYWORDS

EEG, ERP, editing, cinema, neurocinematics, perception, film, continuity

1. Introduction

In his 1900 film *As Seen Through a Telescope*, G. A. Smith cut from a medium-wide shot (the standard at the time) of a man looking through a telescope, to a close-up of a woman's ankle. This had the effect of aligning the viewer's experience with that of the character in a seamless, continuous flow; a powerful narrative device was developed. Cinematography involves the creation of a sense of flow and continuity from a succession of shots with different viewing angles, times, spatial locations and characters. This sense of flow is created in the brain of the viewer, but it is strongly influenced by editing. It is not surprising that throughout the history of cinema, film editing has been extensively studied and remains one of the most essential technical aspects of the medium (Shklovsky, 1928/1971; Mitry, 1963/2002; Deleuze, 1984; Cutting et al., 2011).

In recent decades, with the interest of cognitive neuroscience in the study of different artistic disciplines (e.g., Ramachandran and Hirstein, 1999; Hasson et al., 2008; Van de Cruys and Wagemans, 2011), film editing has once again been at the forefront regarding cinematographic research. Addressing film editing is relevant for the advance of film studies as well as for understanding cognitive processes and their neural underpinnings (Matusz et al., 2019; Soto-Faraco et al., 2019). For example, the study carried out by Silva and her research team analyzed shot changes across editing cuts to expand neuroscientific knowledge about the memorization and mental organization of episodic events (Baldassano et al., 2018; Silva et al., 2019), while from a different perspective, Smith developed a theory applicable to continuity in film editing grounded on current knowledge about cognitive processes (Smith, 2012).

Edit cuts may serve a wide variety of purposes as narrative, aesthetic and emotional devices in the context of cinematography. These techniques may pursue the creation of smooth visual continuity flow (continuity edits) or else breaks that flag important narrative boundaries (such as those spanning different scenes). Here, we are especially interested in how different types of shots are combined in continuity editing. This is an essential aspect in cinematographic construction and typically represents one of the most relevant subjects of cinematography handbooks (Reisz and Millar, 1971; Marimón, 2015), as continuity editing is designed to help direct the viewer's attention toward the narrative of the film and away from the cinematographic technical artifact (Burch, 1969). Subverting these rules breaks continuity, and is sometimes used to expressively highlight shot changes, seeking to have an impact on the audience (Marimón, 2015). However, excessive use of forbidden editing [as it was called by Bazin, 1958/2004] can break the narrative virtuality of the film, diverting the viewer's attention from the narrative content.

Specifically, we are interested in how different types of cuts for continuity affect the viewer depending on changes in shot scale and the filming angle. Cuts involving scale changes (scale-out: from a close to an open shot, or scale-in: from an open to a close shot) are typically used in cinematic language for managing the emotional tension of the scene. A typical filmic structure would start with a wide shot showing the context of the staging, and make the scene unfold dramatically by scaling-in to a closer shot focusing on the main character(s) and their emotional expressions (Marimón, 2015). Cuts involving angle variation (the point of view of the camera moves with respect to the object being filmed) are another typical device used in cinematic language, for example when filming a conversation, alternating specific shot angles for each character. In film theory, the angle variation is bounded by the 30-degree rule, whereby 30-degree changes are considered to be the minimal variation needed to maintain a sense of continuity in the viewer (Shimamura, 2013; Marimón, 2015).

The study of cinematographic editing in cognitive neuroscience has employed different methods including measurement of response speed and accuracy, eye gaze, and neurophysiological measures with functional magnetic resonance imaging (fMRI) and electroencephalography (EEG). For example, with regard to fMRI, Magliano and Zacks (2011) analyzed cuts depending on their continuity or discontinuity in space, time, and action, and discovered that spatial–temporal changes and action changes produced different neural patterns, compared to purely continuity edits. Based on eye-tracking studies, amongst other analytical approaches, Smith and Henderson (2008) and Smith (2012) proposed the *Attentional Theory of Cinematic Continuity* (AToCC). The AToCC is based on the viewer's processing of the visual image in relation to gaze shifts and fixations, proposing how a sense of continuity is achieved across the shot change, and the editing techniques that favor it. For example, Smith proposes that attentional cues at the end of one shot may be used to produce a gaze shift in the viewer in order to make the cut to the next shot less noticeable. Regarding EEG approaches, the main lines of analysis cover the temporal and the frequency domains of neural responses. Heimann and her team (2017) combined both to investigate how the shot change affects neural activity when the 180-degree rule is broken. Their results showed that cuts, in general, elicit early event related potentials (ERPs) similar to those produced by syntactic violations in language and action sequences, and also suggested that the left–right reversal resulting when the 180-degree rule is broken caused an orienting deficit, reflected it in the event-related desynchronization/synchronization (ERD/ERS) pattern. We have recently reported a shot change study addressing ERD/ERS analysis, focusing on common patterns triggered by different types of continuity cuts (Sanz-Aznar et al., 2021). Our results showed a common pattern of ERD/ERS for continuity cuts related to theta and delta frequency bands, mainly in parietal electrodes. During the first 188 ms following an edit cut there is synchronization in theta rhythms and, between 250 and 750 ms after, a desynchronization in the delta frequency band.

In the present study we concentrate on the consequences of different types of continuity edit cuts on ERPs extracted from the EEG. ERP analysis provides a continuous measure of processing between a stimulus and a response with better temporal resolution than other physiological measures, making it possible to determine with precision the stage or stages of processing that are affected by a specific experimental manipulation (Luck, 2014). This possibility is important in the analysis of film cuts, given their fine temporal pattern within the timeline of the film. With ERPs, differences between cuts can be detected in a time resolved manner and potentially linked to specific stages of information processing (e.g., Reid and Striano, 2008; Sitnikova et al., 2008; Matran-Fernandez and Poli, 2015).

1.1. ERPs to shot change by cut: a brief review of findings

One of the first studies using ERPs to address edit cuts was conducted by Sitnikova and her research team (2008). In their study, ERPs following the moment of the cut were compared across three types of shot change: cuts in continuity with narrative coherence (take the bread loaf + cut a slice of bread), cuts that violated goal-related action requirements (take the bread loaf + place an iron on the loaf), and cuts that were unexpected but did not violate the goal-related requirements of the action (take the bread loaf + ironing a pair of the pants on an ironing board). Sitnikova et al. found significant differences in an anterior negativity N300/N400 component, which appeared in all three cases but was largest for unexpected cuts, and smallest for continuity cuts. In another study Reid and Striano (2008) focused on shot changes related to the completion of a previous activity in the scene. In their study, the shot after the cut could be the predictable completion of the previous action, or an unexpected ending. The ERPs after the unanticipated action endings showed a greater amplitude of the N400 component over frontal, central and parietal regions, with respect to anticipated cuts.

In a subsequent study, Francuz and Zabielska-Mendyk (2013) addressed the differences between related and unrelated cuts through ERP analysis. Unrelated cuts refer to those that involve a scene change (McKee, 1997), while related cuts refer to shot changes within the same scene. Related cuts maintain the visual unity of the filmic virtuality, at least within the immediate previous narrative context, whilst the unrelated cuts involve discontinuity (the concept of related and unrelated cuts was originally proposed in Carroll and Bever, 1976). Francuz and Zabielska-Mendyk found ERP differences in frontal electrode responses between 300 and 648 ms, unrelated cuts displaying more negative ERP compared to related cuts. This result is in line with the findings of Sitnikova et al. (2008) and Reid and Striano (2008). In addition, Francuz and Zabielska-Mendyk observed that the same difference extended over central electrodes (from 448 to 648 ms), and similar but opposed polarity differences in parietal electrodes, again with larger amplitude for unrelated cuts.

Heimann and her research team (2017) addressed the well-known 180-degree rule in film editing, typically used in dialogues between two characters. This is a conventional editing rule whereby all camera shots in the scene must be taken from one side of the virtual axis defined by an imaginary line linking the positions of the two interacting characters, and it is used to prevent confusion in the viewer (Murch, 1995/2001; Marimón, 2015). Heimann et al. found that cuts that violate the 180-degree rule triggered neuronal responses comparable to those that occur due to syntactic violations in language, such as an early left anterior negativity followed by a late positivity in the same area, and semantic violations such as a negative deflection around 400 ms in frontal scalp electrodes. This is in line with the studies discussed above addressing other breaches in filmic continuity.

In a different study, Matran-Fernandez and Poli (2015) compared ERPs elicited by shot changes (all consisting of related cuts) against a baseline condition without any cut. Compared to a no-cut baseline, shot changes produced a negative potential in the frontal electrodes, from around 100 ms up to 700 ms. Matran-Fernandez and Poli proposed to name the negative potential peaking between 380 and 420 ms, Post-Cut Negativity (PCN). This pattern, coherent with other studies (e.g., Reid and Striano, 2008; Sitnikova et al., 2008), defines the negative potential in frontal electrodes that is usually related with unexpected information, as is a typical neural reaction triggered by a cut. Analyzing the PCN they found a positive correlation with the duration of the shot preceding the cut, after controlling for luminance variations across cuts.

More recently, Andreu-Sánchez and her research team (2018) addressed the rhythmic aspect of editing techniques. To do this, they created two audio-visual clips with identical narrative content but different average shot length (ASL), a slow editing rhythm of 5.9 s and a faster one with an ASL of 2.4 s. The results showed that a faster edition triggered higher amplitude ERPs in occipital electrodes compared to slower editing, whilst slower editing triggered higher amplitude responses in the frontal and central scalp regions, compared to faster editing. Based on the results analyzed by ERP, frequency domain analysis and dipole estimation, they concluded that a faster editing rhythm increases attention, but at the same time decreases conscious processing.

Finally, Silva et al. (2019) analyzed editing cuts to study episodic memory encoding. They did not differentiate between types of cuts, but instead, they focused on differentiating those scenes that were remembered from those that had been forgotten, after viewing the film. They found significant differences in the ERPs of remembers vs. forgotten scenes between 600 and 1400 ms. Specifically, cuts that were recalled had elicited a more negative polarity in frontal, parietal and mid-temporal electrodes.

1.2. Inferences from the ERP findings so far

Based on the literature briefly reviewed above, the most characteristic response in the ERP signal triggered by shot changes is a large negative deflection in frontal electrodes between 300 and 700 ms, and a large positive deflection in the parietal scalp (Francuz and Zabielska-Mendyk, 2013). The negative deflection in frontal scalp electrodes could happen due to the overlap of evoked potentials with negative amplitudes, such as N300, N400 (Reid and Striano, 2008; Sitnikova et al., 2008) and SNW1 (Francuz and Zabielska-Mendyk, 2013), which can be associated to the neural processes triggered by unexpected events. This interpretation is supported by the fact that the amplitude of this frontal negativity is consistently larger when post-cut shots are least expected, such as for unrelated cuts (Francuz and Zabielska-Mendyk, 2013) and for violations of goal-related action expectations (Sitnikova et al., 2008). Sitnikova et al. (2008) noted that the latency of the N400 component for cuts is longer than in N400 arising from semantic inconsistencies when reading, due to the visual permanence of the incongruous information. Based on their investigations, they suggest that spatio-temporal information processing is reflected both in the N400 and in the late positivity.

In addition, the amplitude of the N400 could be related to the memory formation of the just-encoded event episode (Silva et al., 2019). The greater amplitude of N400 observed in the different comparisons between cuts reviewed above may reflect the mapping of visual input for semantic memory in a similar way that happens with language (Sitnikova et al., 2008). In particular, Matran-Fernandez and Poli (2015) suggest that this negative difference potential, located in the frontal and central scalp areas (which could be interpreted as N400 or more broadly as slow cortical potentials – SCP) may reflect the integration of new semantic information acquired after the shot change, built over the context of the previous shot.

On the other hand, the slow negative wave (SNW1) and slow positive wave (SPW) components, also present in some cases (Francuz and Zabielska-Mendyk, 2013) have been related to an orienting response (OR) or orienting reflex, reflecting an immediate response to a change in the environment (Sitnikova et al., 2008). The OR is a physiological and behavioral reaction that happens in response to new or unexpected stimuli (Öhman, 2021), preceding the orienting of sensory receptors toward salient events in the environment. Thus, the OR causes a non-conscious management of cognitive and attentional resources in order to process certain sensory information, called prominence or saliency (Evangelopoulos et al., 2013). Sitnikova et al. (2008) related the parietal positive deflection with the analysis of the observer content, even without the need for spatio-temporal inconsistencies to appear in the observed action. Regarding the positive deflection in parietal electrodes, Francuz and Zabielska-Mendyk (2013) ruled out the possibility that it reflects a P3 component, since the negative responses that co-occur in the frontal area make it incompatible with the P3a responses. Heimann and her research team (Heimann et al., 2017) detected a greater amplitude in the ERP between 400 and 600 ms after the cut (P4-6) in central right regions for shot changes that break the 180-degree rule, and lower in anterior left. This response, the authors interpreted, is associated with the adjustments of a detected violation without it reaching a level of visual awareness.

In summary, the consistent and large anterior negative and posterior positive deflections triggered in the ERPs have been generally related to the detection of incongruences due to the cut, and the ensuing adaptive responses in the brain. Here, the specifics vary from study to study, in part because of the differences in the concrete type of cut used, and the methodological and analytical approaches. For instance, one study showed that related cuts show late positive amplitudes in the central electrodes between 648 and 1800 ms, while unrelated cuts triggered negative amplitudes (Francuz and Zabielska-Mendyk, 2013), regardless of containing violations in goal-related action requirements (Sitnikova et al., 2008). In this sense, the responses triggered by unrelated cuts match with SNW2 (slow negative wave) responses, which usually appear after SNW1 responses (Francuz and Zabielska-Mendyk, 2013).

Despite most cut-related differences are seen in late components such as the ones discussed above, some studies have also reported earlier differences, though less consistently. Sitnikova and her research team (2008) reported the presence of components P1 (80 ms), N1 (180 ms) and P2 (220 ms), that reflect some variations due to the different types of shot changes, but not in all cases. They reported a greater negativity in the frontal and central areas, especially in the right hemisphere, between unrelated and related cuts as early as 150 to 250 ms. However, no differences were seen with expectation violations of goal-related action sequences. Heimann and her research team (2017) reported an early posterior negativity between 140 and 220 ms (N2), probably reflecting the recognition of a mismatch that implies bottom-up processes, followed by an early anterior positivity (P2). They interpret these early processes as the reflection that a reanalysis of the stimulus after a breach of expectation.

The differentiation between types of shot changes that has been studied with ERP (and other methods) has mainly focused on the coherence or incoherence of the narrative and visual content. Fewer studies, however, address variations between types of cuts used to create perceptual and narrative continuity, such as the present study, regarding continuity edits across shot scale and filming angle. The results of Andreu-Sánchez et al. (2018), larger frontal amplitude ERPs for cuts in slower compared to faster editing, and those by Matran-Fernandez and Poli (2015), who found larger Post-Cut Negativity after longer shots, fit well with this more fine-grained approach. Globally, these results would suggest that longer shots accumulate richer, more constraining pre-cut contexts, and therefore a more complex updating process would be required after the edit point (Matran-Fernandez and Poli, 2015).

1.3. Scope of the present research

Continuity editing is the basis for the construction of the narrative through the fragmentation of the staging, maintaining a sense of continuity in the viewer. Continuity is achieved via a set of editing techniques essential in film construction, conditioning how the spectator perceives the narrative (Reisz and Millar, 1971; Marimón, 2015), but their brain correlates have not been addressed before in detail. Instead, most studies have addressed the contrast between cut transitions preserving filmic (and narrative) continuity with cut transitions that break this continuity in different ways (scene transitions, semantic violations or breaches in the conventional rules). In the present study we address a comparison between different types of cut transitions meant to preserve continuity [that is, that they are related, in terms of time and space coherence, as described by Amiel, 2001/2005 and Burch, 1969], but change the scale or the filming angle across the cut. The possibility of locating differences in neural responses as a function of the shot scale and filming angle variations between the shots around the cut is consistent with the film theory developed in film editing. According to the scale law and the *30*-degree rule, the shot scale and angle variations larger than 30-degrees (45-degrees in North American film theory, e.g., Thompson, 1993) are relevant factors to produce a sense of continuity in the viewer, that is they are thought to go unnoticed (Marimón, 2015). One could therefore question whether variations of these, more subtle, types of edits will trigger differences in the typical neural responses to cuts, or any effects at all, in terms of ERPs, similar to the ones seen with more salient (and noticeable) cuts in the research reviewed above.

According to these past studies, the neuronal correlates registered from frontal and posterior electrodes are sensitive to different types of cuts between 300 and 700 ms (e.g., Sitnikova et al., 2008; Francuz and Zabielska-Mendyk, 2013). Therefore, if present, one would expect differences to be located mainly in the amplitudes of the signals between 300 and 700 ms, being negative in frontal electrodes and positive in posterior electrodes (Francuz and Zabielska-Mendyk, 2013). Specifically, according to previous research (Sitnikova et al., 2008), we expected to find differences in two time windows: from 250 to 350 ms, and from 350 to 600 ms.

2. Methods

These analyses have been performed on a dataset used in a previous research study (Sanz-Aznar et al., 2020, 2021), which addressed ERD/ERS but not ERPs. Below, we provide a brief description of the methods and design. The EEG recording was carried out on the participants while they passively viewed four film excerpts without any particular task. The analysis focused on the shot change (outgoing shot replaced by incoming shot) instant and subsequent time window. The specific pre-processing and data analyses are presented in full detail. Figure 1 visually represents the scheme of the experimental design.

2.1. Participants

Twenty subjects from the University of Aalborg, Denmark, chosen amongst undergraduate, master and, PhD students. The participants had an average age of 26 years (range 22–38 years old; 11 male and 9 female). Participants received 100 Danish Krone (approximately 13.42 Euro at the time of testing) worth of vouchers as compensation for their time, which they could redeem at a Danish cinema chain for tickets, popcorn and soft drinks. The experiment had the Aalborg University ethical approval signed letter with ID 2020-020-00504.

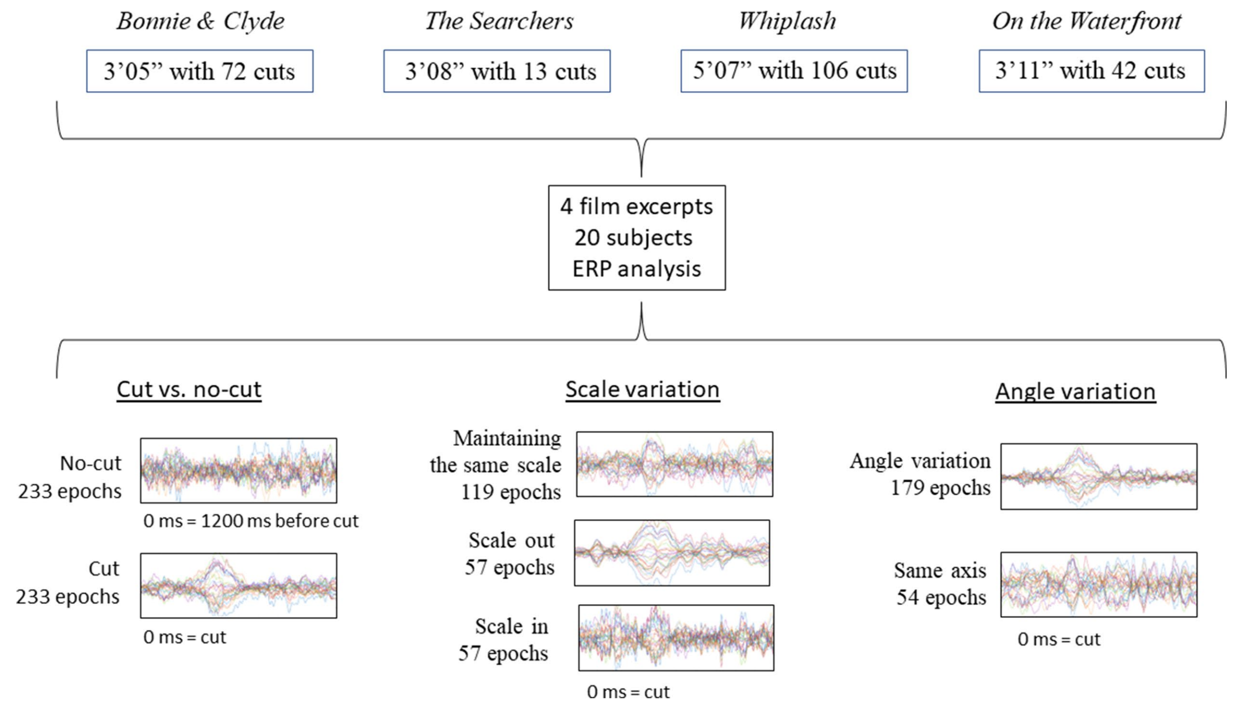


FIGURE 1
Visual representation of the study design.

2.2. Materials and procedure

To carry out the experiment, we extracted four fragments from four feature films: *Bonnie & Clyde* (3′05″, Penn, 1967), *The Searchers* (3′08″, Ford, 1956), *Whiplash* (5′07″, Chazelle, 2014) and *On the Waterfront* (3′11″, Kazan, 1954). The indicated films can be classified within the institutional mode of representation (IMR) defined by Burch (1987/2006). This mode of representation encompasses the usual style of cinema consumed by the typical film spectator in western countries. The selected excerpts contained shot changes edited in continuity without temporal or spatial breaks, that conform to the definitions of absolute connection or articulated montage (Amiel, 2001/2005), as well as the categories of continuity or proximity in terms of spatiality and rigorously continuous or hiatus in terms of temporality (Burch, 1969). These categories defined by Amiel and Burch are the ones that allow the spectator to keep the virtual sensation of a continuous space–time, allowing the possibility of a continuity cut.

We chose 4 films with clear differences in terms of rhythm, esthetics and cinematographic technique, trying to compensate as much as possible the influence of spurious aspects of one specific film in the results (Table 1)[1]. For each film excerpt, we selected the cuts that fitted the continuity criteria mentioned above and, in addition, had incoming shots longer than 1000 ms (to ensure sufficiently long analysis epoch): 72 usable cuts in *Bonnie & Clyde* (14 not usable), 13 in *The Searchers* (1 not usable), 106 in *Whiplash* (6 not usable) and 42 in *On the Waterfront* (4 not usable). Usable cuts refer those continuity cuts in which the incoming shot is longer than 1000 ms, that suppose the analysis epochs. The average length of the shots for *Bonnie & Clyde* is 2.3 s ($SD = 3.56$, $Min = 1.04$, $Max = 25.7$), for *The Searchers* is 11.83 s ($SD = 10.57$, $Min = 4.08$, $Max = 41.25$), for *Whiplash* is 2.95 s ($SD = 3.04$, $Min = 1.04$, $Max = 21.96$) and for *On the Waterfront* is 4.26 s ($SD = 2.37$, $Min = 1.33$, $Max = 13$). To control for sequential effects, the order of the 4 excerpts was randomized for each subject. Before the viewing begins a white image with a central fixation cross was shown for 1′30″, and for 15″ between excerpts.

The EEG recordings were carried out from 31 electrodes distributed according to the American Electroencephalographic Society 10–20 system, as the participants watched the movie excerpts without any particular task required (as movies are normally watched). The sampling frequency was 256 Hz. The devices used to amplify the EEG signal were two channel box g.Tec g.Gammabox connected to two biological amplifier g.Tec g.USB Amp. Both types of dispositive have 16 channels each, so connecting one device as a master and another as a slave allows

---

[1] To define the technical and esthetic differences between the film excerpts, we resorted to analysis references that provide general classifications regarding lighting esthetics (Revault D’Allonnes, 1991), film style (Bordwell, 1985; Langford, 2009; Thanouli, 2009), narrative-structural description (McKee, 1997), rhythmic ratio or cutting rates (Redfern, 2022) and the technical characteristic of whether the excerpt is black and white or color. Taking these characteristics as a reference, a first list of different fragments extracted from 20 films was elaborated. This first selection was reviewed by two experts in film analysis (Juan José Caballero Molina y Daniel Jariod Dato). Based on the report prepared by the experts, the list was definitely reduced to the four selected fragments.

TABLE 1 Film excerpts characteristics.

Film	Excerpt timecode[a]	Color or B&W	Rhythmic ratio[b]	Narrative value[c]	Aesthetics[d]	Filmic style[e]
Bonnie & Clyde	01:46:48:01 to 01:49:53:02	Color	27.24	Inflection point. Strong conflict.	Modern	Transition
The Searchers	00:01:35:00 to 00:04:42:20	Color	4.79	Character presentation. No conflict.	Classical	Classical
Whiplash	00:23:33:13 to 00:28:40:13	Color	21.89	Inflection point. Strong conflict.	Modern	Post-classical
On the Waterfront	00:21:24:08 to 00:24:35:16	Black and White	14.45	Scene before an inflection point. Low conflict.	Classical	Classical

[a]Timecode format: Hour:Minute:Second:Frame. [b]Average cuts per minute in the film excerpt. [c]Narrative value in relation to the cinematographic narration and the conflict shown in the excerpt (McKee, 1997). [d]Lighting style (Revault D'Allonnes, 1991). [e]Cinematographic style (Bordwell, 1985; Langford, 2009; Thanouli, 2009).

32 channels. The EEG signal was referenced to the right earlobe (online) and to the Fp1 electrode (offline). The recorded signal was analyzed using EEGLab (v2021.1) in MatLab (R2017a) environment, as well as the statistics. The recorded signal was re-referenced to the average of all scalp channels, Common Average Reference (Luck, 2014; Yao et al., 2019). The specific distribution of the electrodes on the scalp is represented in Figure 2A.

After recording the signal, a manual artifact rejection procedure was performed, and ICA was applied to separate the components caused by blinks and other muscle movements with a range of flagging from 90 to 100%. We applied a bandpass filter to the raw signal from 0.2 to 40 Hz. The average number of rejected epochs per subject was 18.22% (Min: 4.29%, Max: 42.06%). To carry out the ERP analysis, the signal was segmented in 1200 ms epochs, from −200 before to 1000 ms after the cut, with the 200 ms before to the cut used as baseline (following Matran-Fernandez and Poli, 2015).

2.3. Analyses

We selected two broad electrodes clusters of interest. The frontal cluster included Fp1, Fp2, Afz, F7, F3, Fz, F4, F8, Fc1 and Fc2, and the posterior cluster (parieto-occipital) included P7, P5, P3, Pz, P4, P6, P8, PO3, PO4, O1, Oz, O2, PO7 and PO8.

We first performed an analysis pooling all of the cuts together, simply comparing ERPs to cuts (stimulus) vs. no cut (absence of stimulus) segments (see below, for more details). The ERP for cut and no-cut condition are referenced to the same baseline, 200 ms before the cut [−200 ms, 0 ms], where 0 ms is the moment of the cut. Cut condition epochs included the first 1000 ms after the cut [0 ms, 1000 ms] and no-cut condition epoch included to the 1000 ms right before the reference [−1200 ms, −200 ms]. The analysis was run for comparison to the study by Matran-Fernandez and Poli (2015) comparing cuts vs. no-cuts between 380 and 420 ms using the Mann–Whitney test. As we re-referenced the electrodes to the average of all the scalp channels, we applied the analysis only to the frontal area and the posterior area independently, instead of grouping all the electrodes as done originally in Matran-Fernández and Poli.

Then we performed two comparative analyses contrasting the conditions (types of cuts) of interest (Figures 2B,C and Table 2). The first addressed scale variations, comparing the cuts as they scale in, scale out, or keep the same scale. To address cuts with scale variations we used a cluster-based permutation test, with the three conditions (keep, scale in, scale out) as independent variables. The second analysis focused on variations in filming angle, comparing cuts that keep the same filming axis (below 30°) with those that vary it (larger than 30°) using cluster-based permutation test. The statistical analysis Monte Carlo method, based on 500 randomizations (Heimann et al., 2017), with cluster correction, was applied in the two electrode clusters established (frontal and posterior) in both cases. We carried out time resolved point-by-point contrast across the first 0–1000 ms window after the cut.

Then, following Sitnikova et al. (2008), we focused on the average voltage amplitude in two time windows of interest; 250 to 350 ms, and 350 to 600 ms. These two time windows cover the typical times for the N300 and N400 epochs. Like Sitnikova, we used ANOVA to compare the different conditions (keep, scale in, scale out) as independent variables. We also applied t-tests for follow-up pairwise comparations. To reduce Type I error consequence of multiple comparison test p-values were adjusted for multiple comparison using Bonferroni correction and the accepted significance value for both cases was $p\text{-value} \leq 0.01$. To estimate the effect size, we applied partial eta squared for ANOVAs and Cohen's d for pairwise comparations.

3. Results

3.1. ERPs to cuts compared to no-cut periods

We first compared ERPs to cuts vs. a no cut baseline period of equal duration, as a reality check with respect to previous results using a similar method (e.g., Matran-Fernandez and Poli, 2015). We compared the average ERP in the interval [0–1000 ms] pooled over all the cuts (Figure 3, red line) with the average ERP of the interval [−1200 ms, −200 ms] prior to each cut (Figure 3, black line), which did not contain any cut. The ERPs were baseline

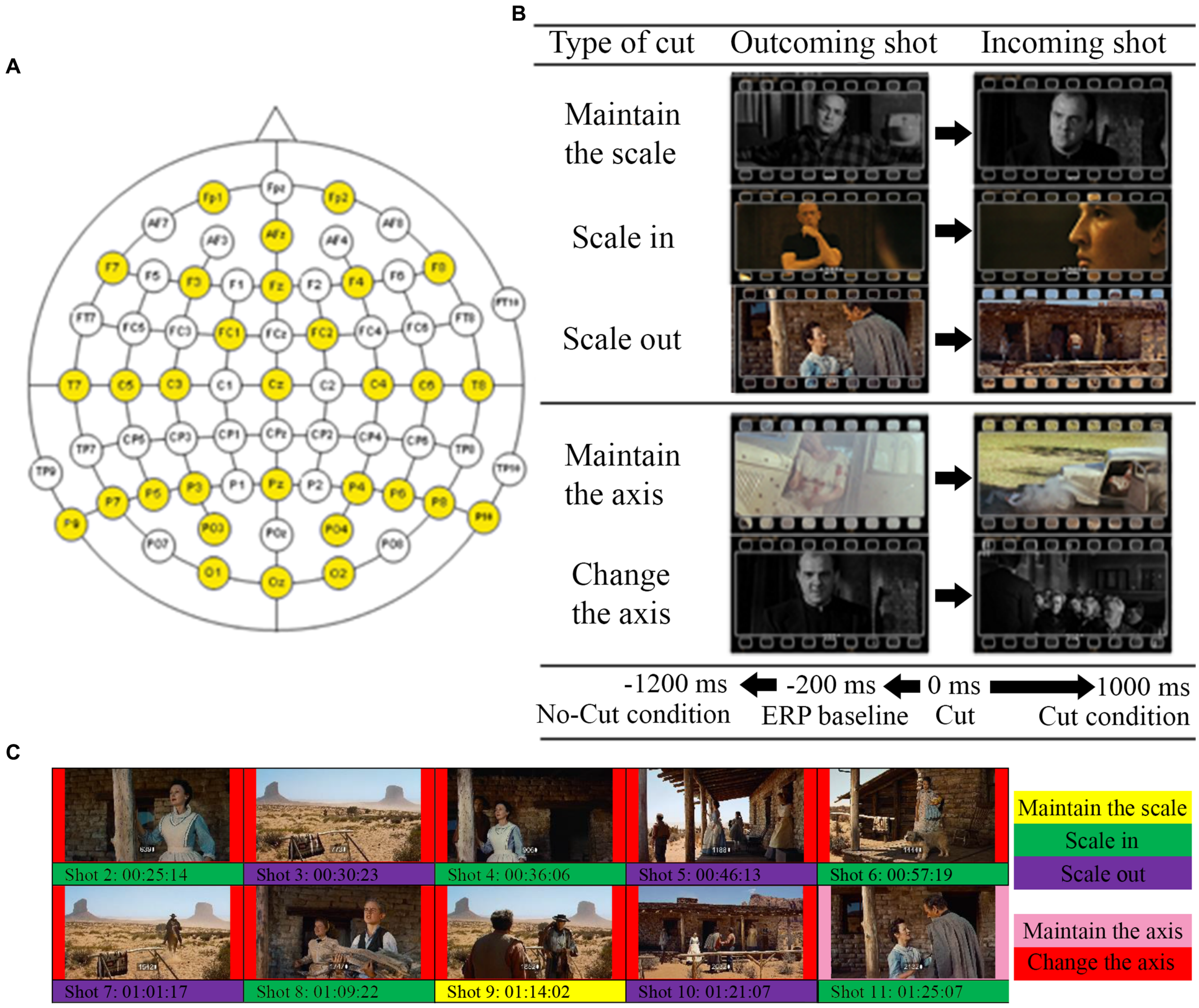


FIGURE 2
(A) Distribution of the electrodes used in the experiment (highlighted in yellow). **(B)** Examples for each type of cut used in the study. **(C)**. Example of type of cut segmentation for *The Searchers*.

TABLE 2 Number of usable cuts (epochs) per each condition.

Film excerpt	Cut	No-cut (before cut)	Keep scale	Scale in	Scale out	Same axis	Angle variation
Bonnie & Clyde	72	72	25	24	23	30	42
The Searchers	13	13	3	5	5	3	10
Whiplash	106	106	78	14	14	12	94
On the Waterfront	42	42	13	14	15	9	33
Total	233	233	119	57	57	54	179

corrected to the 200 ms interval before the cut [−200 ms, 0 ms]. As in previous experiments (e.g., Matran-Fernandez and Poli, 2015), the cut event triggered an ERP clearly differentiated from the signal where there is no shot change, as shown in Figure 3.

Cut-evoked ERPs display a negative potential from 200 ms onwards in frontal electrodes, and a positive potential in the posterior areas, compared to no cut events. Figure 4A shows the cut condition for frontal and posterior electrodes.

The frontal cluster showed a clear negative potential peaking between 300 and 400 ms, followed by a progressive return to baseline until the end of the epoch. This posterior positive deflection peaks between 300 and 400 ms, followed by return to baseline until the end of the epoch. The scalp distributions (Figure 4B) of the average of the signal recorded following the cuts confirm the positive deflection in the posterior scalp, and a negative one in the frontal scalp, between 200 and 800 ms.

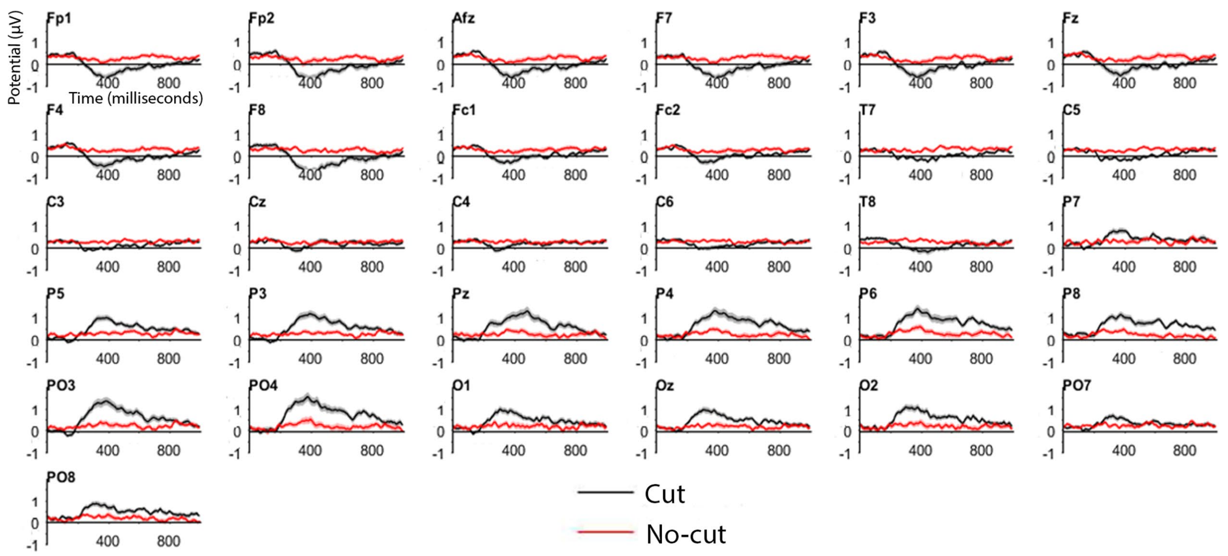


FIGURE 3
ERP for cut events (all cuts pooled, in black) and no-cut baseline (in red) for each electrode.

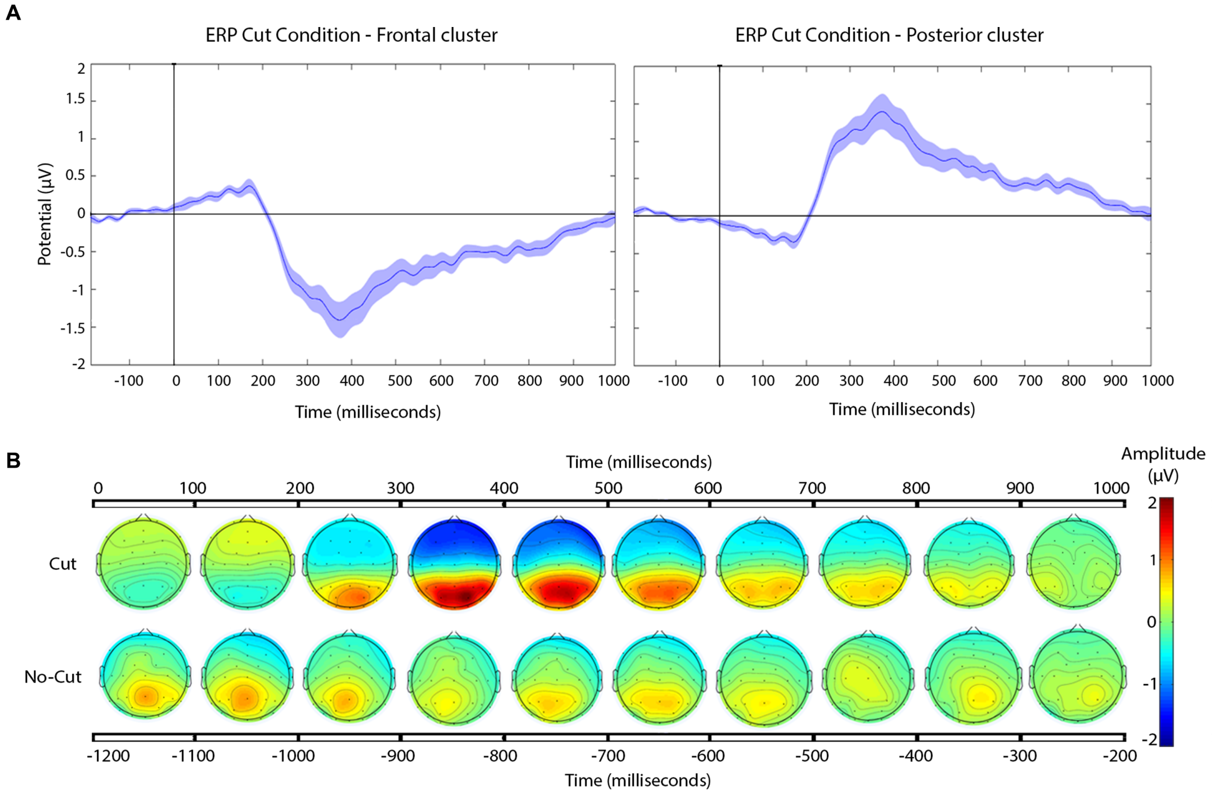


FIGURE 4
(A) Grand-average ERP for all cuts pooled for the frontal (left), and posterior (right) electrode clusters. Shaded areas represent the standard error of measurement (SEM) of ERP. **(B)** Scalp distribution of the voltage variation for all the cuts pooled together, compared to no cut, in 100ms time steps.

To further characterize this pattern, we applied a statistical approach similar to Matran-Fernandez and Poli (2015), comparing cut vs. no-cut between 380 and 420 ms using a Mann–Whitney test. We applied this comparison to the frontal and posterior electrode clusters separately, because unlike Matran-Fernandez and Poli, who used a whole scalp ERP (Global Field Power), our ERPs were already re-referenced the whole scalp.

We obtained a significant difference in both clusters, frontal (cut = −0.9261 μV vs. no cut = −0.043 μV $W = 66$, $p < 0.01$) as well as in posterior clusters (0.9229 μV vs. 0.0549 μV; $W = 187$,

$p < 0.01$). Based on the results obtained in this across-the-board ERP analysis, we can differentiate cut condition from no-cut condition with our set of materials. We then concentrate on specific cut types separately for the frontal and the posterior clusters.

3.2. ERPs following scales changes across the cut

To address scale variations between the shots across the cut, we carried out a point-by-point cluster-based permutation test (with one within-subject factor, scale change: scale in, scale out, or keep), separately for each electrode cluster (frontal, posterior), and established the significance level at *p*-value≤0.01. In frontal (Figure 5A) and posterior (Figure 5B) cluster analyses, the point-by-point cluster-based permutation test returned significant effects between 300 and 800 ms, according with the proposed hypothesis.

The magnitude of the negative frontal deflection (observed for all cuts pooled, in the previous analysis) varied as a function of the scale change across the edit (see Figures 5A,C). In particular, scale out produced a larger negative shift compared to those that keep the scale, whereas cuts with scale in produced a smaller negative shift, compared to keep-scale cuts. This was confirmed statistically, using the same time windows as in Sitnikova et al. The ANOVAs in the frontal cluster resulted significant in the 250–350 ms window [$F(2,75) = 50.37$, $p < 0.01$, $\eta^2 = 0.573$ (large effect)], and in the 350–600 ms window [$F(2,189) = 171.59$, $p < 0.01$, $\eta^2 = 0.645$ (large effect)]. The follow-up pair-wise *t*-tests confirmed the pattern described above in both time windows (ERP amplitude, scale out > maintain > scale in; see Table 3, for statistical values).

In the posterior cluster (Figure 5B) the ANOVA returned significant effects around the window 300–800 ms. The posterior electrode cluster displayed a positive shift between 300 and 800 ms, consistent with what had been observed in the overall analyses (all cuts pooled). Like for the negative shift in frontal electrodes, the amplitude of the posterior positive shift depended on the type of scale change across the cut, following a similar pattern: with respect to keeping scale, scale out produced stronger shift, and scale in produced a decrease in the positive shift. For consistency, we used the same time windows as before for the statistical confirmation of this pattern. The ANOVAs for the window 250–350 ms and the 350–600 ms window in the posterior were both significant {respectively, [$F(2,75) = 55.95$, $p < 0.01$, $\eta^2 = 0.599$ (large effect)], [$F(2,189) = 45.03$, $p < 0.01$, $\eta^2 = 0.605$ (large effect)]}. The follow up paired *t*-tests confirmed the pattern with the significance levels (see Table 3).

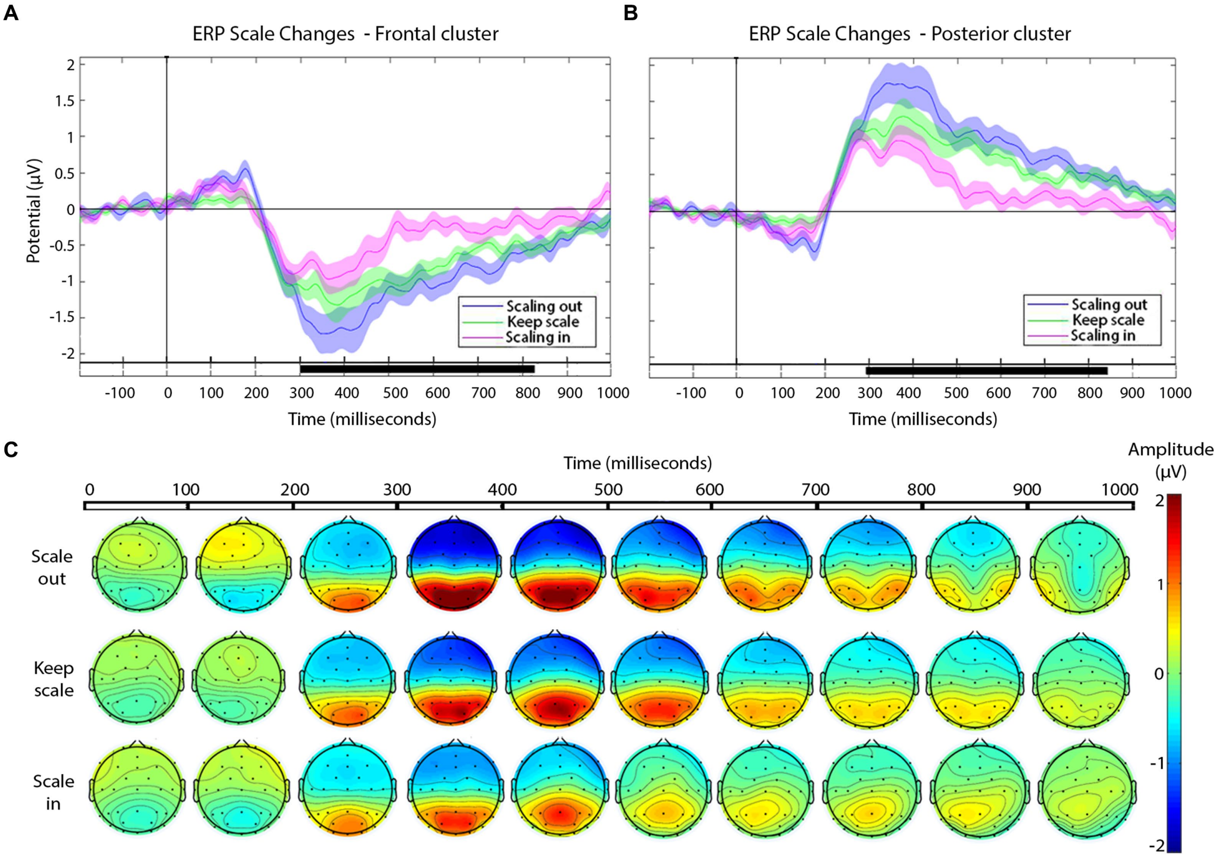


FIGURE 5
(A,B) Grand-average ERPs for each type of scale variation cuts (see legend) in the frontal **(A)** and posterior **(B)** electrode clusters. Shaded areas represent the SEM of ERP. In the timeline, black segments indicate significant effects in the cluster-based permutation test, *p*-value≤0.01. **(C)** Scalp distribution of the negative and positive deflections, as a function of scale change across the cut, presented in windows of 100ms after the cut.

TABLE 3 *t*-test results for the two time windows (250–350ms; 350–600ms) and clusters (frontal and posterior) for paired comparations between scale in, scale out and maintain scale.

	Time window	Comparation	*df*	*t*	*P*	*d*	Effect size interpretation
Frontal	250–350 ms	Out vs. keep	19	−2.11	0.05	−0.473	Medium
		Out vs. In	19	−3.54	<0.01	−0.794	Medium
		In vs. Keep	19	−2.10	0.05	−0.469	Small
	350–600 ms	Out vs. keep	19	−1.60	0.125	0.359	Small
		Out vs. In	19	−4.06	<0.01	0.909	Large
		In vs. Keep	19	−7.33	<0.01	1.64	Trivial
Posterior	250–350 ms	Out vs. keep	19	2.85	0.01	0.638	Medium
		Out vs. In	19	3.85	<0.01	0.337	Medium
		In vs. Keep	19	2.05	0.05	−0.008	Trivial
	350–600 ms	Out vs. keep	19	3.35	<0.01	0.750	Medium
		Out vs. In	19	6.61	<0.01	1.479	Large
		In vs. Keep	19	6.91	<0.01	1.545	Large

3.3. ERPs following angle variations across the cut

To address shot changes that vary the filming axis we started with a point-by-point approach as before, using cluster-based permutation test. The outcome is not as conclusive as for scale variations (see, Figure 6A).

As can be seen in Figure 6A, there were no significant differences in the ERP between 200 and 800 ms, the approximate segment of interest. For completeness, the scalp distributions are shown in Figure 6B. Therefore, the ERPs to this type of cut followed the general profile that was expected from previous research, as well as our own analysis of the pooled dataset, and the scale variation cuts: frontal negative deflection and posterior positive deflection, over the 200–800 ms window, with peaks happening 300–400 ms. Yet, there were no detectable ERP changes as a function of filming angle variation across the cut. The *t*-test for the window 250–350 ms and the 350–600 ms window in the frontal and posterior cluster were no significant. Results can be consulted in Table 4.

4. Discussion

At variance from most ERP studies addressing continuity violations or unrelated filmic cuts, we addressed potential differences between different types of continuity edits. Understanding how the human brain captures shot changes across cuts is important because they are amongst the most widely used devices for film composition in cinematography. In particular, we set out to analyze the ERPs following related film edits which involved differences in shot scale (scale in, scale out, or keep scale) and in shot angle (angle variation, vs. same axis). For example, scale and angle variation across shots help articulate different space and unity pieces and adding a hierarchical value of different shots within the scene whilst maintaining a feel of continuity in the viewer (Reisz and Millar, 1971; Marimón, 2015).

In order to align our results with those of previous studies we first analyzed the ERPs to all cuts pooled together, with respect to a no-cut baseline. Please note that such reality check does not only seek confirmation of previous findings, but also helps bring some coherence across the very different filmic materials and potential viewing conditions across different studies. Our results were overall consistent with the previous literature that analyzed the ERP triggered by shot changes, with very clear ERPs evoked by editing cuts (Francuz and Zabielska-Mendyk, 2013; Andreu-Sánchez et al., 2018; Andreu-Sánchez and Martín-Pascual, 2021). However, because of methodological difference we can only confirm the existence of the PCN (380–420 ms), described by Matran-Fernandez and Poli (2015) for all the electrodes, in frontal electrodes. Based on our across-the-board ERP analysis, we saw a negative potential appearing at a frontal cluster and a positive posterior deflection, peaking between 300 and 400 ms, both followed by return to baseline until the end of the epoch. This result coincides in frontal electrodes with that obtained by Matran-Fernandez and Poli, whose working hypothesis was that the cut would trigger an N400, due to abrupt change of visual information. However, our results differ in the posterior cluster, which in our case shows a clear positive deflection.

This divergence between results could be related to the different references used to calculate the ERPs across the two studies (averaged earlobes in Matran-Fernandez and Poli, 2015; whole scalp average in ours), making the direction and distribution of electrical fields not direcly comparable between studies. However, the temporal profile of the frontal negativity seen in our results fits well with the timing in Matran-Fernandez & Poli, as well as with that of several other ERP studies to filmic cuts (Reid and Striano, 2008; Sitnikova et al., 2008). About the positive deflection detected in posterior cluster, our results are generally coincident with most previous research (e.g., Francuz and Zabielska-Mendyk, 2013; Andreu-Sánchez et al., 2018; Andreu-Sánchez and Martín-Pascual, 2021).

In general, the ERP pattern can be described as a negative component in frontal electrodes and positive deflection in posterior electrodes, extending roughly from 200 to 800 ms, with peaks between 200 and 400 ms. These results, coincide with abundant previous literature (Reid and Striano, 2008; Sitnikova et al., 2008; Francuz and Zabielska-Mendyk, 2013), and provide

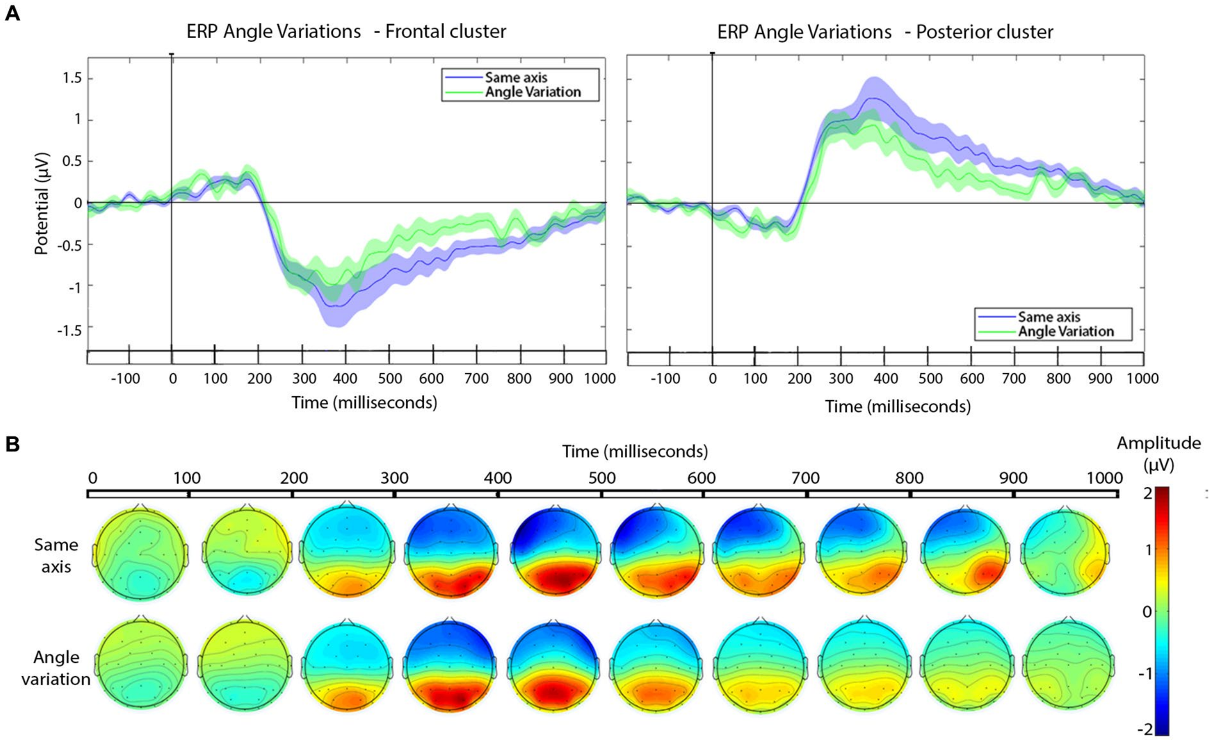


FIGURE 6
(A) ERPs for edits with and without angle variation across the cut for the frontal (left) and the posterior (right) electrode clusters. Shaded areas represent the SEM of ERP. In the timeline, significant effects in the cluster-based permutation test, *p*-value≤0.01 are indicated in black (no significant effects were found). **(B)** Scalp distribution of the voltage variations as a function of the angle variation (same or different axis) across the cut, presented in windows of 100ms after the cut.

TABLE 4 ***t*-test results for the two time windows (250–350ms; 350–600ms) and clusters (frontal and posterior) for paired comparations between same or different axis.**

	Time window	*df*	*t*	*P*	*d*	Effect size interpretation
Frontal	250–350 ms	19	−0.62	0.54	−0.139	Trivial
	350–600 ms	19	−2.40	0.03	−0.536	Medium
Posterior	250–350 ms	19	−1.08	0.29	0.241	Small
	350–600 ms	19	2.78	0.01	0.622	Medium

grounds for the investigation of variations as a function of cut type in the time windows of interest in the present study. These neuronal correlates have been related to the processing of cognitive inconsistencies triggered by the sudden change of visual input produced across the editing cut (Silva et al., 2019). This interpretation is coherent with our own results, but also with other studies addressing other ERP components, such as the study by Calbi et al. (2017). Calbi et al. analyzed the Kuleshov effect (Kuleshov, 1934/1994) by focusing on the ERP N170, which is characteristically elicited by faces. They concluded that the Kuleshov effect is the consequence of an attribution of expectations set by the shot preceding the cut as a function of the emotional coherence, or incoherence, across the edit.

As the main focus of our research, we registered effects in this frontal-negative and posterior-positive deflections related to different types of cuts depending on the scale variation. The negative-going deflection in the frontal cluster had a graded amplitude depending on the type of scale change across the cut, with similar scalp distributions in all cases: The largest shift corresponded to scale out, whilst scale in cuts led to the smallest (yet still significant) deflection, with ERPs to cuts keeping the scale falling in-between. Given the timing and scalp distribution of these negative shifts, one could relate them to the N300 and N400 components, following the interpretation of earlier ERP studies addressing film cuts (Hamm et al., 2002; Kumar et al., 2021). In particular, amplitude differences in the N300 and N400 are common in research that compares different types of cuts (Reid and Striano, 2008; Sitnikova et al., 2008), although they have so far been studied for edit transitions involving some degree of narrative or filmic continuity breach such as related vs. unrelated, predictable vs. unpredicted action continuations, or semantic and compositional incongruences. For instance, larger N400s are usually seen in cuts with action incoherence (Reid and Striano, 2008; Sitnikova et al., 2008). Based on these interpretations and the effects observed here, we might suggest the hypothesis that scale reductions from one shot to the next (scale in) may reduce the perception of incoherence, compared to keeping scale, and to a larger extent, scaling out. According to Matran-Fernandez and Poli (2015), the amplitude variation in N400 may reflect the integration of new semantic information built on the context of the previous shot. Interpreting our results under this light, a reduction in shot scale may imply a smaller amount of new information to integrate than scaling out.

On the other hand, the positive shift in the posterior cluster between 300 and 800 ms also varied in amplitude as a function of cut type in our results. Again, and in parallel to the anterior cluster, scaling out produced the largest amplitude in this component, followed by keep and then by scaling in. According to prior ERP studies on film editing, amplitude variations in the posterior responses between 350 and 450 ms have been related to a greater inconsistency between shots across an edit (Francuz and Zabielska-Mendyk, 2013). Based on these previous investigations, scale out would enhance neural responses related to visual incongruity, with these effects being smoother in the case of scale in.

Overall, the results from the frontal and parieto-occipital clusters are coherent, allowing us to propose the interpretation that scale out produce greater incoherence (at least, in terms of neural correlates) than scale in, with cuts that keep the scale constant being an intermediate case. According to this interpretation, one would therefore expect that scale out would be more noticeable on average to the viewers than scale in across shots in a related cut. This relationship between scale in/out and visibility is known in professional film editing and well documented in film editing handbooks. Specifically, to achieve smooth flow in a scene, the editing should be designed from the most open shots to closed ones (Reisz and Millar, 1971). Since it is not possible to sustain an incremental reduction of scale throughout each cut along a scene, the correspondence law (Marimón, 2015) is often applied. The correspondence law consists of keeping the same scale (among other aspects) to keep editing flow. The ERP results thus coincide with the praxis in cinema montage. However, it would be interesting to test these predictions in a behavioral experiment, using perhaps a similar approach as that of Smith and Henderson (2008) or Magliano and Zacks (2011), where they asked participants to explicitly detect edit points whilst watching a film.

Regarding the other main focus of our study, cuts based on the angle variation, the results did not show significant differences in the ERPs. In this case we cannot confirm the expectation, according to the film theory. According to the *30*-degree rule (Marimón, 2015) the cuts that imply less than 30^0 of variation (including those with no angle variation) should produce less continuity feeling in the viewer than the cuts that change the camera angle by more than 30^0. Cuts that break the *30*-degree rule are thought to be more 'aggressive' from a cognitive point of view and perceived by the viewer as a jump in the continuity flow (Königsberg, 1987; Marimón, 2015). Please note that although we have not found significant results in this case, there could be an actual difference, but the effect size might be small and our measurement not sensitive enough. Perhaps related to this, Smith et al. (2012) proposed that spatial memory is worse than identification memory, and therefore the spectator does not retain an allocentric representation of scenes. Maybe this poorer spatial reference explains the absence of significant results in N300 and N400 for angle variations. If this speculation where true, and in the light of our results, one would expect that, overall, angle variations should be less noticeable consciously than variations in scale.

In another investigation using the same dataset used here some of us have addressed (Sanz-Aznar et al., 2023) similar comparisons between cuts through ERD/ERS analysis, instead of ERPs. This analyses found differences between related cuts that vary shot angle vs. those that keep the same axis, as well as cuts that vary the scale of the shot vs. those that keep it. These differences were found in the first 125 ms after the cut, in central and posterior electrodes and between 300 and 1000 ms in frontal and parietal electrodes, in the frequency range Theta (3–7 Hz) and Delta (0.5–3 Hz). Theta frequency results relevant in the first 400 ms and delta from 500 ms.

5. Conclusion

We have investigated ERPs to film cuts using widely accepted editing techniques employed to confer a sense of continuity in the viewer, such as scale in/out, and angle variation. In all cases, these techniques have been developed with intuition and practice by filmmakers and montage professionals, to maintain a sense of continuity and flow across cuts. Our results seem to provide grounds for validation of this common editing practice from a neuroscience perspective and, importantly, may lay the basis for addressing other common editing techniques or even test new editing variations in a principled way. As one potential limitation of the current experimental approach, it should be noted that the subject's perceptive evaluation of the cut events are not taken into account. This decision was made consciously, because we sought to reproduce the cinematographic viewing experience as much as possible, hence favoring a passive viewing without specific task. Finally, the present results also raise some expectations regarding the viewer's awareness of visual edits in films. The results obtained allow us to hypothesize that the spectator could have a greater awareness of the visual transitions involving scaling out than when scaling in, with shot changes that keep constant scale as an intermediate typology. This means that in film editing, when a cut scales in, it should be more invisible for the spectator than a cut that scales out. No significant differences have been detected when comparing cuts with different camera angles between shots, and so this case remains less conclusive.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/m97xf/?view_only=a0427e4b8cee4c51bb09b0f978b95883.

Ethics statement

The studies involving human participants were reviewed and approved by the experiment had the Aalborg University ethical approval signed letter with ID 2020-020-00504. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JS-A and LB experimental design and carried out experiments. SS-F and JS-A performed data analysis and wrote the article. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Ministerio de Ciencia e Innovación (PID2019-108531GB-I00 AEI/FEDER), AGAUR Generalitat de Catalunya (2021 SGR 00911) grants to SS-F and NextGenerationEU (Margarita Salas) grant to JS-A.

Acknowledgments

We thank Mireia Torralba for her valuable help about Matlab coding and Cristina Baus for her advice about ERP analysis pipeline.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Amiel, V. (2001/2005). *Estética del montaje*. Madrid: Alba Editores.

Andreu-Sánchez, C., and Martín-Pascual, M. Á. (2021). Perception of cuts in different editing styles. *Profesional de la Información* 30:e300206. doi: 10.3145/epi.2021.mar.06

Andreu-Sánchez, C., Martín-Pascual, M., Gruart, A., and Delgado-García, J. (2018). Chaotic and fast audiovisuals increase attentional scope but decrease conscious processing. *Neuroscience* 394, 83–97. doi: 10.1016/j.neuroscience.2018.10.025

Baldassano, C., Hasson, U., and Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *J. Neurosci.* 38, 9689–9699. doi: 10.1523/JNEUROSCI.0251-18.2018

Bazin, A. (1958/2004). *¿Qué es el cine?* Madrid: Ediciones Rialp.

Bordwell, D. (1985). *Narration in the fiction film*. Madison: University of Wisconsin Press.

Burch, N. (1969). *Praxis du cinéma*. Paris: Gallimard.

Burch, N. (1987/2006). *El tragaluz del infinito*. Madrid: Ediciones Cátedra.

Calbi, M., Heimann, K., Barratt, D., Siri, F., Umilta, M., and Gallese, V. (2017). How context influences our perception of emotional faces: a behavioral study on the Kuleshov effect. *Front. Psychol.* 8:1684. doi: 10.3389/fpsyg.2017.01684

Carroll, J., and Bever, T. (1976). Segmentation in cinema perception. *Science* 191, 1053–1055. doi: 10.1126/science.1251216

Chazelle, D. (Director). (2014). *Whiplash* [Film]. Blumhouse Productions.

Cutting, J., Brunick, K., DeLong, J., Iricinschi, C., and Candan, A. (2011). Quicker, faster, darker: changes in Hollywood film over 75 years. *Iperception* 2, 569–576. doi: 10.1068/i0441aap

Deleuze, G. (1984). *La imagen-movimiento. Estudios sobre cine 1*. Barcelona: Ediciones Paidós.

Evangelopoulos, G., Zlatintsi, A., Potamianos, A., Maragos, P., Rapantzikos, K., Skoumas, G., et al. (2013). Multimodal saliency and fusion for movie summarization based on aural, visual, and textual attention. *IEEE Trans Multimedia* 15, 1553–1568. doi: 10.1109/TMM.2013.2267205

Ford, J. (Director). (1956). *The searchers* [Film]. Warner Bros.

Francuz, P., and Zabielska-Mendyk, E. (2013). Does the brain differentiate between related and unrelated cuts when processing audiovisual messages? An ERP study. *Media Psychol.* 16, 461–475. doi: 10.1080/15213269.2013.831394

Hamm, J. P., Johnson, B. W., and Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clin. Neurophysiol.* 113, 1339–1350. doi: 10.1016/S1388-2457(02)00161-X

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D. (2008). Neurocinematics: the neuroscience of film. *Projections* 2, 1–26. doi: 10.3167/proj.2008.020102

Heimann, K., Uithol, S., Calbi, M., Umiltà, M., Guerra, M., and Gallese, V. (2017). "Cuts in action": a high-density EEG study investigating the neural correlates of different editing techniques in film. *Cogn. Sci.* 41, 1555–1588. doi: 10.1111/cogs.12439

Kazan, E. (Director). (1954). *On the waterfront* [Film]. Columbia Pictures.

Königsberg, I. (1987). *The complete film dictionary*. London: Bloomsbury.

Kuleshov, L. (1934/1994). *L'art du cinéma et autres écrits*. Levier: L'Age d'Homme.

Kumar, M., Federmeier, K. D., and Beck, D. M. (2021). The N300: an index for predictive coding of complex visual objects and scenes. *Cerebral Cortex Commun.* 2:tgab030. doi: 10.1093/texcom/tgab030

Langford, B. (2009). *Post-classical hollywood: film industry, style and ideology since 1945: film industry, style and ideology since 1945*. Edinburgh: Edinburgh University Press.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge: MIT press.

Magliano, J. P., and Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cogn. Sci.* 35, 1489–1517. doi: 10.1111/j.1551-6709.2011.01202.x

Marimón, J. (2015). *El muntatge cinematogràfic. Del guió a la pantalla*. Barcelona: Edicions Universitat Barcelona.

Matran-Fernandez, A., and Poli, R. (2015). Event-related potentials induced by cuts in feature movies and their exploitation for understanding cut efficacy. Proceedings of the 2015 7th International IEEE/EMBS conference on neural engineering (NER) (pp. 74–77). Montpellier: IEEE.

Matusz, P. J., Dikker, S., Huth, A. G., and Perrodin, C. (2019). Are we ready for real-world neuroscience? *J. Cogn. Neurosci.* 31, 327–338. doi: 10.1162/jocn_e_01276

McKee, R. (1997). *Story: substance, structure, style and the principles of screenwriting*. New York: ReganBooks.

Mitry, J. (1963/2002). *Estética y psicología del cine*. Madrid: Siglo Veintiuno.

Murch, W. (1995/2001). *In the blink of an eye: A perspective on film editing*. Los Angeles: Silman-James Press.

Öhman, A. (2021). The orienting response, attention, and learning: an information-processing perspective. In H. D. Kimmel, OlstE. H. van and J. F. Orlebeke (Eds.) *The orienting reflex in humans* (pp. 443–471). London: Routledge.

Penn, A. (Director). (1967). *Bonnie & Clyde* [Film]. Warner Bros. Pictures.

Ramachandran, V., and Hirstein, W. (1999). The science of art: a neurological theory of aesthetic experience. *J. Conscious. Stud.* 6, 15–41.

Redfern, N. (2022). Analysing motion picture cutting rates. *Wide Screen* 9, 1–29.

Reid, V., and Striano, T. (2008). N400 involvement in the processing of action sequences. *Neurosci. Lett.* 433, 93–97. doi: 10.1016/j.neulet.2007.12.066

Reisz, K., and Millar, G. (1971). *The technique of film editing*. Oxford: Elsevier.

Revault D'Allonnes, F. (1991). *La Lumière au cinéma*. Paris: Seuil.

Sanz-Aznar, J., Aguilar-Paredes, C., Sánchez-Gómez, L., Bruni, L. Y., and Wulff-Abramsson, A. (2020). Methodology for detection of ERD/ERS EEG patterns produced by cut events in film fragments. *Lect. Notes Comput. Sci* 12196, 157–175. doi: 10.1007/978-3-030-50353-6_12

Sanz-Aznar, J., Sánchez-Gómez, L., Aguilar-Paredes, C., Wulff-Abramsson, A., and Luis Emilio Bruni, L. E. (2023). Differentiating types of cinematographic shot changes by cut through an EEG power spectral analysis. *Psychol. Aesthet. Creat. Arts* 17. doi: 10.1037/aca0000609

Sanz-Aznar, J., Sánchez-Gómez, L., Bruni, L. E., Aguilar-Paredes, C. Y., and Wulff-Abramsson, A. (2021). Neural responses to shot changes by cut in cinematographic editing: an EEG (ERD/ERS) study. *PLoS One* 16:e0258485. doi: 10.1371/journal.pone.0258485

Shimamura, A. (2013). "Psychocinematics: issues and directions" in *Psychocinematics: Exploring cognition at the movies*. ed. A. Shimamura (Oxford: Oxford University Press), 1–26.

Shklovsky, V. (1928/1971). *Cine y lenguaje*. Barcelona: Anagrama.

Silva, M., Baldassano, C., and Fuentemilla, L. (2019). Rapid memory reactivation at movie event boundaries promotes episodic encoding. *J. Neurosci.* 39, 8538–8548. doi: 10.1523/JNEUROSCI.0360-19.2019

Sitnikova, T., Holcomb, P., Kiyonaga, K., and Kuperberg, G. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *J. Cogn. Neurosci.* 20, 2037–2057. doi: 10.1162/jocn.2008.20143

Smith, T. (2012). An attentional theory of cinematic continuity. *PRO* 6, 1–27. doi: 10.3167/proj.2012.060102

Smith, T., and Henderson, J. (2008). Edit blindness: the relationship between attention and global change blindness in dynamic scenes. *J. Eye Mov. Res.* 2, 1–17. doi: 10.16910/jemr.2.2.6

Smith, T., Levin, D., and Cutting, J. (2012). A window on reality: perceiving edited moving images. *Curr. Dir. Psychol. Sci.* 21, 107–113. doi: 10.1177/0963721412436809

Soto-Faraco, S., Kvasova, D., Biau, E., Ikumi, N., Ruzzoli, M., Morís-Fernández, L., et al. (2019). *Multisensory interactions in the real world (elements in perception)*. Cambridge: Cambridge University Press.

Thanouli, E. (2009). *Post-classical cinema: an international poetics of film narration*. London: Wallflower Press.

Thompson, R. (1993). *Grammar of the edit*. Oxford: Butterworth–Heinemann.

Van de Cruys, S., and Wagemans, J. (2011). Putting reward in art: a tentative prediction error account of visual art. *Iperception* 2, 1035–1062. doi: 10.1068/i0466aap

Yao, D., Qin, Y., Hu, S., Dong, L., Bringas Vega, M. L., and Valdés Sosa, P. A. (2019). Which reference should we use for EEG and ERP practice? *Brain Topogr.* 32, 530–549. doi: 10.1007/s10548-019-00707-x

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 09 August 2023
DOI 10.3389/fpsyg.2023.1196209

OPEN ACCESS

EDITED BY
Miguel Ángel Martín-Pascual,
Spanish Public Television, Spain

REVIEWED BY
Kevin L. Blankenship,
Iowa State University, United States
Celia Andreu-Sánchez,
Autonomous University of Barcelona, Spain

*CORRESPONDENCE
Stuart S. Miller
✉ ssmiller@ksu.edu
John P. Hutson
✉ jphutson1@gmail.com

RECEIVED 29 March 2023
ACCEPTED 24 July 2023
PUBLISHED 09 August 2023

CITATION
Miller SS, Hutson JP, Strain ML, Smith TJ, Palavamäki M, Loschky LC and Saucier DA (2023) The role of individual differences in resistance to persuasion on memory for political advertisements. *Front. Psychol.* 14:1196209. doi: 10.3389/fpsyg.2023.1196209

The role of individual differences in resistance to persuasion on memory for political advertisements

Stuart S. Miller[1]*, John P. Hutson[1]*, Megan L. Strain[2], Tim J. Smith[3], Maria Palavamäki[4], Lester C. Loschky[1] and Donald A. Saucier[1]

[1]Department of Psychological Sciences, Kansas State University, Manhattan, KS, United States, [2]Psychology, University of Nebraska at Kearney, Kearney, NE, United States, [3]Department of Psychological Sciences, University of London Birkbeck College, London, United Kingdom, [4]Department of Media, Aalto University, Otakaari, Finland

When people see political advertisements on a polarized issue they take a stance on, what factors influence how they respond to and remember the adverts contents? Across three studies, we tested competing hypotheses about how individual differences in *social vigilantism* (i.e., attitude superiority) and *need for cognition* relate to intentions to resist attitude change and memory for political advertisements concerning abortion. In Experiments 1 and 2, we examined participants' intentions to use resistance strategies to preserve their pre-existing attitudes about abortion, by either engaging against opposing opinions or disengaging from them. In Experiment 3, we examined participants' memory for information about both sides of the controversy presented in political advertisements. Our results suggest higher levels of social vigilantism are related to greater intentions to counterargue and better memory for attitude-incongruent information. These findings extend our understanding of individual differences in how people process and respond to controversial social and political discourse.

KEYWORDS

social vigilantism, need for cognition, resistance to persuasion, attitudes, memory, political advertisements

1. Introduction

> Nothing limits intelligence more than ignorance; nothing fosters ignorance more than one's own opinions; nothing strengthens opinions more than refusing to look at reality.—Sheri S. Tepper

Opposing opinions about political and social issues often provoke sharp and contentious debate. It seems impossible to escape confrontations with opponents when people voice their positions on political or social issues. But when confrontation occurs, people may decide to engage the opponent by investing attention and effort to debate, or they may disengage from the opponent by diverting attention and conserving effort. We examined these decisions by investigating individual differences related to how people respond to attempted persuasion. Specifically, when faced with highly controversial social issues in political advertisements, do people engage in active resistance to attitude challenges, or engage in selective exposure away from them? Additionally, do these strategies for resisting persuasion affect memory for pro-attitudinal vs. counter-attitudinal arguments?

Social psychologists have examined individual difference factors that influence how receptive vs. resistant people are to attitude change attempts for decades (for reviews, see Petty and Wegener, 1998; Crano and Crislin, 2006). Stronger attitudes are harder to change (Krosnick et al., 1993; Eagly and Chaiken, 1995; Krosnick and Petty, 1995; Pomerantz et al., 1995; Zuwerink and Devine, 1996; Jacks and Devine, 2000; Jacks and Cameron, 2003; Visser et al., 2003; Skitka et al., 2005; Visser et al., 2006; Saucier et al., 2014), and challenges to strongly held attitudes provoke people to resist these attempts either by engaging with the persuasion attempt or by disengaging from it. Additionally, the strategies people use to resist such attempts vary with their individual differences in *need for cognition* (Cacioppo and Petty, 1982; Cacioppo et al., 1986) and *social vigilantism* (Saucier and Webster, 2010). In contrast to much previous work, here we do not focus on attitude change, but instead we investigate how individual differences in social vigilantism and need for cognition predict people's strategies to resist persuasion and their memory for the contents of the persuading arguments.

1.1. Strategies to resist persuasion

When people are motivated to resist persuasion, and preserve their pre-existing attitudes, they may use various strategies to actively or passively counter the specific persuasion attempt or attitude challenge. While people may use several strategies to resist persuasion (see Jacks and Cameron, 2003; Saucier et al., 2014), we focus on two: the active strategy of counterarguing by which people attack the opposing information, and the passive strategy of selective exposure by which people withdraw from and avoid the opposing information. Importantly, these two strategies involve fundamentally different choices to either engage with or avoid opposing information, which have important implications for later memory for the persuasive material.

1.2. Need for cognition

People higher in need for cognition are less likely to be persuaded by peripheral cues, such as the number of arguments (regardless of their quality) or the physical attractiveness of speakers (for a review, see Cacioppo et al., 1996), and prefer, are more receptive to, and have a better memory for cognitive vs. emotional appeals (Haddock et al., 2008). Importantly, higher levels of need for cognition are related to greater resistance to attitude change attempts (Haugtvedt and Petty, 1992) and therefore may also be related to greater counterarguing as a resistance strategy. Additionally, people higher in need for cognition may be more likely to engage with the information presented about an issue (and therefore less likely to engage in selective exposure), more likely to pay attention to it, and be more likely to recall it later (Peltier and Schibrowsky, 1994). Because need for cognition is related to the tendency to engage information (as opposed to ignoring it), selective exposure may be less likely to occur regardless of whether the information supports or opposes people's pre-existing attitudes about an issue. Existing findings are inconclusive on whether need for cognition is associated with greater tendencies to engage in selective exposure (e.g., Westerwick et al., 2017; Knobloch-Westerwick et al., 2020; Ryu and Vargas, 2021) or lesser tendencies to do so (e.g., Tsfati and Cappella, 2005).

1.3. Social vigilantism

Social vigilantism is the tendency for people to believe their own attitudes are superior to others' attitudes, to resist persuasion attempts and attitude challenges, and to attempt to impress their own attitudes on others (Saucier and Webster, 2010; Saucier et al., 2017). Higher levels of social vigilantism are related to greater resistance to persuasion. Specifically, people higher in social vigilantism have more extreme attitudes and show greater use of various resistance strategies—particularly counterarguing—in response to attitude challenges about a range of socially controversial topics (Saucier and Webster, 2010; Raimi and Leary, 2014; Saucier et al., 2014; Maki and Raimi, 2017; O'Dea et al., 2018). Counterarguing involves directly engaging and trying to defeat an opposing argument, and is a commonly used and effective technique for resisting persuasion (Cameron et al., 2002; Jacks and Cameron, 2003). Social vigilantism has consistently been related to counterarguing, but has been inconsistently related to using passive strategies to resist persuasion, with some studies finding that social vigilantism is positively correlated with selective exposure (O'Dea et al., 2018), but others not (Saucier et al., 2014).

Because it is still unclear whether social vigilantism and need for cognition are related to selective exposure to resist persuasion for strongly held attitudes, we investigated this question in the current study. We focused on these two individual difference constructs in particular because of their relevance to engaging with, and processing of, persuasive messages. Social vigilantism is important for understanding how belief superiority and motivations to impress self-professed superior attitudes upon others relate to counterarguing. Need for cognition is relevant for predicting whether people will engage in selective exposure because need for cognition represents individuals' tendencies to engage with information. To further understand how these individual differences relate to counterarguing and selective exposure, we additionally examined the consequence of these effects on memory for the provocative content. The objective of any political advertising is to first get you to attend to the content, then remember it and eventually incorporate the content into your own beliefs. As we were not examining attitude change for strongly held beliefs in our research, we stop at the earlier stage of analyzing whether memory for content is selective and congruent with participants' preexisting beliefs.

1.4. Linking resistance strategies to subsequent memory

Because of the likely differences in cognitive processing engendered by counterarguing vs. selective exposure strategies, there are strong reasons to hypothesize links between the strategies people report using in response to attitude-congruent vs. -incongruent information and their subsequent memory for it. People remember what they attend to in their environment (Loftus, 1972; Hollingworth and Henderson, 2002; Tatler et al., 2005; Zelinsky and Loschky, 2005; Pertzov et al., 2009). Counterarguing and selective exposure strategies likely differ in attention, thus also in memory. Specifically, in order to counterargue, one must attend to and process information, thus encoding memory for it.

1.4.1. Selective exposure and memory

Theories of selective exposure argue that people will avoid or tune out attitude-incongruent information. Memory experiments have found that people tend to have worse memory for attitude-incongruent information (Eagly and Chaiken, 1995). However, there are several important considerations concerning these selective exposure memory effects. Meta-analyses show that selective exposure effects on memory across studies are fairly weak (Eagly and Chaiken, 1995). However, these effects become stronger when individuals' attitude strength is included as a moderating variable, with selective exposure effects on memory being more common for people with weakly held, but highly partisan attitudes (Eagly and Chaiken, 1995). Conversely, people with strongly held beliefs tend not to show selective exposure effects, but rather use resistance strategies that increase their engagement with attitude-incongruent information. Critically, although people clearly engage in selective exposure, effects on memory are not entirely dependent on attitude congruence. Instead, moderating variables such as attitude strength, propensity to counterargue, and belief superiority are also important for understanding how people engage attitude-incongruent information (Albarracín and Mitchell, 2004; Brannon et al., 2007).

1.5. Overview of the current experiments

The current research extends that on selective exposure and memory in two important ways. First, we test whether social vigilantism and need for cognition are related to the strategies people use to resist persuasion for strongly held attitudes. Second, we test whether these individual differences relate to memory for attitude-relevant information as evidence of engaging with the contents of persuasion attempts (i.e., counterarguing) or ignoring them (i.e., selective exposure).

Across three experiments, participants viewed a series of political ads and completed attitude and persuasion questionnaires (Experiments 1 and 2), or memory measures (Experiment 3). We used both controversial and non-controversial ads to manipulate attitude congruence (congruent, incongruent, and neutral). We used the topic of abortion because: (1) it was a familiar topic to the participant pool used; (2) people typically have strong attitudes about abortion; and (3) pilot studies showed a bimodal distribution with a fairly even proportion of participants with highly pro-life vs. pro-choice attitudes.

In Experiments 1 and 2, participants viewed either a pro-choice or a pro-life ad and reported their intentions to engage in counterarguing and selective exposure. In Experiment 3, participants watched the videos and completed memory tests about the content of the videos. We examined how need for cognition and social vigilantism were related to (a) participants' intentions to respond with counterarguing or selective exposure (Experiments 1 and 2), and (b) participants' memory (Experiment 3).

2. Experiment 1

In Experiment 1, we tested two hypotheses: the *social vigilantism hypothesis* and the *need for cognition hypothesis*. Both hypotheses predict people higher in that individual difference variable, either social vigilantism or need for cognition, should report stronger intentions to engage in counterarguing. For selective exposure, the need for cognition hypothesis predicts that higher levels of need for cognition will be associated with weaker intentions to engage in selective exposure in response to attitude-*incongruent* political ads about the issue of abortion. However, because past research has found inconsistent relationships between social vigilantism and selective exposure, we made no predictions about social vigilantism and selective exposure in the current experiments. Social vigilantism could be negatively related to selective exposure because of a greater tendency to pay attention to (i.e., not ignore) a message that one wants to argue against. Alternatively, social vigilantism could be positively related to selective exposure because having less appreciation of opposing viewpoints and a greater belief in the superiority of one's own positions may lead those with higher levels of social vigilantism to ignore opposing information. We also included an attitude-*congruent* condition to test whether these patterns of relationships extended to situations where the message people receive is *congruent* with their attitudes about abortion, or whether these patterns of relationships are unique to situations where the message people receive is *in*congruent with their attitudes. Finally, although we were not interested in participants' possibility of attitude change after watching the ads, which we thought was highly unlikely, we nevertheless included attitude pre- and post-measures as a check.

2.1. Methods

2.1.1. Participants

We recruited college students from introductory psychology courses at Kansas State University in exchange for research credit. Our sample ($N = 232$) included 79 men and 153 women, ages 18 to 35 ($M = 19.60$, $SD = 2.30$), most of whom were White (79.9%). All data were collected prior to conducting our analyses.

2.1.2. Materials

To manipulate attitude congruence, we used 2 videos in Experiment 1 (a pro-life ad and a pro-choice ad). The abortion ads, including the arguments presented, were developed specifically for this research. We carried out pilot studies to select arguments based on pro-life and pro-choice participants' ratings, to ensure that both sets of arguments were rated by their respective supporters as approximately equal in their strength, persuasiveness, valence, and clarity. For more details, see Supplementary material.

2.1.2.1. Abortion ads

We created the pro-life (59 s long) and pro-choice abortion (1 min and 7 s) video ads using matching formats. The ads used intertitles to present the arguments, and had video imagery that by itself would be neutral, but when paired with the arguments would strengthen the arguments being presented. As shown in Table 1, we created intertitles brief enough to be read quickly in short videos, with parallel arguments for each ad.

The videos share a visual theme that focused on the hands of different people. The pro-Life video[1] mostly showed the hands of children, doing things like playing with Play-Doh or holding fruit. The pro-choice video[2] focused on the hands of adult women, doing things like searching on a computer or holding their face. We designed the pro-life ad to show the positives of being a child, and the pro-choice to show the difficulty of deciding to have an abortion.

---

[1] https://www.youtube.com/watch?v=eXKjmc4Ai9A

[2] https://www.youtube.com/watch?v=qTQXd2FCh_w

TABLE 1 Arguments for abortion ads.

	Pro-choice	Pro-life
1	Women today have the right to accomplish anything	Innocent lives should be protected
2	Women of all ages choose to have an abortion	Abortion is irresponsible and unsafe
3	There are many reasons for choosing an abortion	Life begins at conception
4	The rights of the fetus should not outweigh a woman's rights	Life should be given a chance

To immediately inform participants about the position each ad would take, each ad started with an intertitle stating the ad was paid for by either a pro-life or pro-choice group. At the end of each video, a final intertitle told viewers to either "Vote for Choice" (pro-choice video) or to "Choose Life" (pro-life video). The videos had instrumental background music. The pro-choice intertitle texts were slightly longer than the pro-life intertitle texts. The pro-choice video was 8 s longer to give participants time to read the ad's intertitles.

2.1.3. Procedure and measures

Participants completed all materials online through a Qualtrics survey in the following order: (1) informed consent, (2) demographic information, (3) abortion attitudes pretest, (4) social vigilantism scale, (5) need for cognition scale, (6) the pro-choice or pro-life video ad that was randomly assigned, in a between-groups design, (7) resistance strategies measure, and (8) abortion attitudes posttest. After completing the study, participants were thanked and debriefed.

2.1.3.1. Abortion attitude strength

We measured the strength of participants' attitudes about abortion using five items modified from Brannon et al. (2007). Participants rated "*The availability of abortion as a legal medical procedure is*" on five 9-point semantic differential scales: good–bad, foolish–wise, unnecessary–necessary, harmful–beneficial, oppose it–favor it. We averaged these items together to create a composite score where higher scores represented more pro-choice attitudes (pretest $M = 5.17$, $SD = 2.72$, $\alpha = 0.98$; posttest $M = 5.22$, $SD = 2.78$, $\alpha = 0.98$). Both the pretest and posttest distributions were multimodal, with distinct peaks at one and nine, and a smaller peak near the midpoint of the scale, demonstrating many participants had strong attitudes about abortion.

2.1.3.2. Social vigilantism

We used the social vigilantism scale (Saucier and Webster, 2010) to measure individual differences in the extent to which people generally think their beliefs are superior to others' beliefs and have a desire to impress their beliefs onto others. Participants responded to the items (e.g., "*I feel as if it is my duty to enlighten other people*") on 1 (*Strongly Agree*) to 9 (*Strongly Disagree*) scales. We averaged the 14 items to create composite scores where higher values represented higher levels of social vigilantism ($M = 5.02$, $SD = 1.10$, $\alpha = 0.85$).

2.1.3.3. Need for cognition

We measured people's preference for, and enjoyment of, deliberate thinking using the need for cognition scale (Cacioppo and Petty, 1982). Participants responded to the items (e.g., "*I would prefer complex to simple problems*") on 1 (*Strongly Agree*) to 9 (*Strongly Disagree*) scales. We averaged the 18 items (reverse-scoring appropriate items) to create composite scores where higher values represented higher levels of need for cognition ($M = 5.46$, $SD = 0.92$, $\alpha = 0.84$).

2.1.3.4. Resistance strategies

To measure participants' intentions to resist persuasion in response to viewing the pro-choice or pro-life video, we used items developed by Jacks and Cameron (2003) and Saucier et al. (2014) to measure eight different resistance strategies (e.g., counterarguing, attitude bolstering). Participants responded to these items with the instructions to "*rate how likely you are to respond in this way to the person who showed you the video.*" While we were only interested in counterarguing and selective exposure, we included the other items as filler material to distract participants from the nature of our study. To reduce the number of variables in our analyses, we averaged the two counterarguing items (e.g., *Respond by thinking about or verbalizing why the person's arguments are faulty*) to create a composite score ($M = 4.48$, $SD = 2.05$, $\alpha = 0.60$), and the two selective exposure items (e.g., *Respond by tuning-out the arguments that contradict my position*) to create a composite score ($M = 3.02$, $SD = 1.87$, $\alpha = 0.75$).

2.2. Results

The distribution of abortion attitudes was predominantly bimodal with the vast majority of participants scoring on one end or the other of the scale. Because we were most interested in the behavior of participants with stronger attitudes about abortion, we excluded participants who scored in the middle range of the scale (3.5 to 6.49) on the pretest of abortion attitudes.[3] For the remaining 153 participants, we coded whether the participants' abortion attitude was pro-life (scores 1 to 3.49, $n = 74$, 34 viewed the pro-life video and 40 viewed the pro-choice video) or pro-choice (scores 6.5 to 9, $n = 79$, 38 viewed the pro-life video and 41 viewed the pro-choice video). We created a variable to indicate whether the video was congruent (coded 1) or incongruent (coded 0) with participants' abortion attitudes. The final sample size we analyzed provided us with power > 80% to detect effect sizes > 0.20.

---

[3] On an ordinal scale of 1–9, we would have removed 4–6. However, because the scores were averaged, they were rational numbers rather than whole numbers. We therefore removed participants having the range of values between those that would round up to 4 and those that would round down to 6. Similar results were obtained in our analyses when less conservative selection criteria were used to exclude only participants who scored in the range 4 to 6.

2.2.1. Attitude change

We first tested our *a priori* assumption that participants with strongly held attitudes about abortion would not change their attitudes after viewing the videos. The results of a 2 (pretest/posttest abortion attitude) × 2 (pro-life/pro-choice video) mixed factorial ANOVA with repeated measures on the first factor showed no evidence of attitude change: abortion attitude $F(1, 151) = 0.65$, $p = 0.42$; abortion attitude × video condition interaction $F(1, 151) = 0.50$, $p = 0.48$. Given that we chose participants with the strongest abortion attitudes, the lack of attitude change after watching a single ad was not surprising but instead is consistent with the idea that people with strongly held attitudes would resist persuasion attempts.

2.2.2. Resistance strategies

Next, we examined the bivariate correlations between social vigilantism, need for cognition, counterarguing, and selective exposure. As predicted, social vigilantism was correlated with intentions to engage in counterarguing ($r = 0.31$, $p < 0.001$). Social vigilantism was not significantly correlated with selective exposure ($r = 0.11$, $p = 0.173$). Counter to our hypothesis, need for cognition was unrelated to counterarguing ($r = 0.01$, $p = 0.868$). However, consistent with our hypothesis, need for cognition was negatively related to selective exposure ($r = -0.23$, $p = 0.004$).

We next tested whether these relationships were moderated by viewing an attitude-congruent or attitude-incongruent persuasion attempt. We entered counterarguing and selective exposure as criterion variables in two separate regression models with attitude congruence (step 1), social vigilantism (step 2), and their interaction (step 3) as predictors. In two additional models, we entered attitude congruence, need for cognition, and their interaction as predictors of counterarguing and selective exposure. As expected, the video's congruence with participants' abortion attitudes affected their intentions to counterargue, such that participants intended to counterargue the counter-attitudinal message more than the pro-attitudinal message (Congruent: $M = 4.07$, $SD = 1.90$; Incongruent: $M = 5.35$, $SD = 2.08$; $b = -1.27$, 95% confidence interval lower $= -1.91$, upper $= -0.64$, $p < 0.001$). Consistent with the social vigilantism hypothesis, social vigilantism was positively related to counterarguing ($b = 0.58$, 95% confidence interval lower $= 0.30$, upper $= 0.87$, $p < 0.001$), and this relationship was not significantly moderated by attitude congruence as indicated by a non-significant social vigilantism X Congruence interaction ($b = -0.31$, 95% confidence interval lower $= -0.88$, upper $= 0.26$, $p = 0.282$). Need for cognition was unrelated to counterarguing intentions ($b = -0.04$, 95% confidence interval lower $= -0.39$, upper $= 0.30$, $p = 0.802$) and did not interact with the video condition ($b = 0.04$, 95% confidence interval lower $= -0.65$, upper $= 0.74$, $p = 0.899$).

Surprisingly, selective exposure intentions did not significantly differ by condition (Congruent: $M = 2.88$, $SD = 1.88$; Incongruent: $M = 3.05$, $SD = 1.73$; $b = -0.17$, 95% confidence interval lower $= -0.75$, upper $= 0.41$, $p = 0.559$). Consistent with the need for cognition hypothesis, need for cognition was negatively related to selective exposure ($b = -0.46$, 95% confidence interval lower $= -0.76$, upper $= -0.16$, $p = 0.003$) and did not interact with attitude congruence ($b = 0.23$, 95% confidence interval lower $= -0.38$, upper $= 0.84$, $p = 0.462$). Social vigilantism was not significantly related to selective exposure ($b = 0.18$, 95% confidence interval lower $= -0.09$, upper $= 0.45$, $p = 0.180$) and, interestingly, did not interact with the attitude congruence of the video ($b = 0.11$, 95% confidence interval lower $= -0.44$, upper $= 0.65$, $p = 0.698$).

2.3. Discussion

Our findings were consistent with previous research showing a persuasion attempt incongruent with a strongly held attitude is unlikely to change that attitude and would likely elicit stronger intentions to counterargue than a persuasion attempt congruent with that attitude. We found support for our social vigilantism hypothesis—higher levels of social vigilantism were related to stronger intentions to counterargue. However, social vigilantism did not interact with the effects of the attitude congruence of the persuasive message. Rather, our data suggest people's tendencies to argue and impress their beliefs on others may be an omnipresent goal, regardless of whether a message agrees or disagrees with their position on the issue. Whether social vigilantism is related to intentions to ignore attitude-relevant information was inconclusive.

Furthermore, we found support for the hypothesis that people with greater need for cognition would be less likely to ignore attitude-incongruent information. However, need for cognition was not correlated with counterarguing, suggesting that while need for cognition may be related to attending to, rather than ignoring, the information in a persuasive message, such attention may not be for the purpose of counterarguing.

Overall, these data suggest individual differences in social vigilantism and need for cognition are important for understanding how people process attitude-relevant information. The chronic motivation to influence others' attitudes by people high in social vigilantism appears to be related to their intentions to counterargue, regardless of the attitude-consistency of the topic. Furthermore, dispositional tendencies for careful thought are also related to being less likely to ignore information relevant to a strongly held attitude.

3. Experiment 2

In Experiment 1, social vigilantism was related to counterarguing regardless of whether the persuasive message was congruent or incongruent with participants' attitudes about abortion. Similarly the attitude congruence of the persuasive message did not moderate the negative relationship between need for cognition and selective exposure. In Experiment 2, we assessed how social vigilantism and need for cognition relate to resistance strategies in the context of an *un*controversial message to test whether social vigilantism and need for cognition predict counterarguing and selective exposure (respectively) more generally, regardless of the attitude-relevance of the information in a persuasive message. In a between-groups design, we used the same pro-choice and pro-life videos but added a third condition in which participants viewed a short video containing a relatively less politically controversial message about disabilities. This allowed us to test whether the associations between social vigilantism and counterarguing or need for cognition and selective exposure are specific to attitude-relevant persuasion attempts or whether they generalize to participants' responses to a message less relevant to attitudes about abortion. Additionally, in Experiment 1, we measured abortion attitudes at the start and end of the study to test our assumption that strongly held attitudes would not change. However, the pre-measure of abortion attitudes may have primed participants' attitudes about abortion, and thereby may have affected how they responded to the items measuring intentions to counterargue or ignore the persuasion attempt. Therefore, in Experiment 2 we waited to measure participants' attitudes about abortion until the end of the study and did not test for attitude change because the results of Study 1 strongly supported our *a priori* assumption that participants' attitudes about abortion were very *un*likely to change after exposure to either of our single political advertisements.

3.1. Methods

3.1.1. Participants

We recruited a new sample of college students from introductory psychology courses at Kansas State University in exchange for research credit. Our sample ($N = 234$) included 102 men and 132 women, ages 18 to 28 ($M = 19.18$, $SD = 1.72$), most of whom were White (81.4%). All data were collected prior to conducting our analyses.

3.1.2. Procedure and measures

We used the same procedure as Experiment 1, with the exception that participants only completed the abortion attitudes items at the end of the study. We also included a Non-controversial video[4] condition in the form of a public service announcement with the concluding message text: *Just because you do something differently, does not mean you are "disabled."* Visually, the ad is set on a series of steps, and people go up and down them in different and creative ways (e.g., dancing and crab walking). The pro-choice and pro-life videos were the same as Experiment 1. Participants were randomly assigned one of the three videos. Composite variables for each of the measures were calculated as described in Experiment 1 (abortion attitudes $M = 5.24$, $SD = 2.80$, $\alpha = 0.99$; social vigilantism $M = 5.18$, $SD = 1.09$, $\alpha = 0.85$; need for cognition $M = 5.83$, $SD = 1.05$, $\alpha = 0.84$; counterarguing $M = 4.40$, $SD = 1.99$, $\alpha = 0.63$; selective exposure $M = 3.12$, $SD = 1.91$, $\alpha = 0.81$). Again, the distribution of abortion attitudes was multimodal, with distinct peaks at one and nine, and a smaller peak near the midpoint of the scale.

3.2. Results

As in Experiment 1, we excluded participants who scored in the middle range of the scale (3.5 to 6.49) on abortion attitudes, resulting in a sample of 160 participants. We coded participants' abortion attitudes as described in Study 1 (pro-life $n = 72$, 24 viewed the non-controversial video, 28 viewed the pro-life video, and 20 viewed the pro-choice video; pro-choice $n = 88$, 29 viewed the non-controversial video, 28 viewed the pro-life video, and 31 viewed the pro-choice video). The final sample size we analyzed provided us with power > 80% to detect effect sizes > 0.20.

3.2.1. Resistance strategies

We first examined the bivariate correlations between social vigilantism, need for cognition, counterarguing, and selective exposure. As predicted, social vigilantism was once again correlated with intentions to engage in counterarguing ($r = 0.32$, $p < 0.001$). However, in contrast to Experiment 1, social vigilantism was significantly positively correlated with selective exposure ($r = 0.21$, $p = 0.008$). Need for cognition was again unrelated to counterarguing ($r = 0.07$, $p = 0.366$). As expected, need for cognition was again negatively related to selective exposure ($r = -0.20$, $p = 0.013$).

We next tested whether these relationships were moderated by attitude congruence. We entered counterarguing and selective exposure in separate regression models with the abortion attitude congruence of the video entered as dummy-coded aspects of the three video conditions (Congruent, Incongruent, or Neutral for the uncontroversial video in step 1), social vigilantism (step 2), and their interactions (step 3) as predictors. In additional models, we entered attitude congruence, need for cognition, and their interactions as predictors.

We replicated the finding from Experiment 1 that the attitude congruence of the persuasion attempt affected intentions to counterargue (Congruent: $M = 3.83$, $SD = 2.18$; Incongruent: $M = 5.16$, $SD = 1.96$; Congruent—Incongruent $b = -1.33$, 95% confidence interval lower $= -2.11$, upper $= -0.54$, $p < 0.001$). The Neutral condition ($M = 4.45$, $SD = 2.00$) did not differ from either the Congruent (Neutral—Congruent $b = 0.62$, 95% confidence interval lower $= -0.16$, upper $= 1.40$, $p = 0.116$) or Incongruent (Neutral—Incongruent $b = -0.71$, 95% confidence interval lower $= -1.50$, upper $= 0.09$, $p = 0.082$) conditions. Replicating the results from Experiment 1, we found social vigilantism was positively related to counterarguing ($b = 0.63$, 95% confidence interval lower $= 0.37$, upper $= 0.88$, $p < 0.001$), and social vigilantism did not interact with attitude congruence ($ps > 0.529$). Again, we found need for cognition was unrelated to counterarguing ($b = 0.19$, 95% confidence interval lower $= -0.11$, upper $= 0.49$, $p = 0.207$) and did not interact with attitude congruence ($ps > 0.120$).

For selective exposure, as expected, we found intentions to ignore the persuasion attempt were highest in the Incongruent condition ($M = 3.74$, $SD = 2.09$; Congruent—Incongruent $b = -1.37$, 95% confidence interval lower $= -2.09$, upper $= -0.65$, $p < 0.001$; Neutral—Incongruent $b = -0.80$, 95% confidence interval lower $= -1.53$, upper $= -0.07$, $p = 0.032$), followed by the Neutral condition ($M = 2.93$, $SD = 1.99$; Neutral—Congruent $b = 0.57$, 95% confidence interval lower $= -0.15$, upper $= 1.28$, $p = 0.119$), and lowest the Congruent condition ($M = 2.37$, $SD = 1.58$). In contrast to Experiment 1, social vigilantism was positively related to selective exposure ($b = 0.40$, 95% confidence interval lower $= 0.15$, upper $= 0.65$, $p = 0.002$). However, social vigilantism did not interact with attitude congruence ($ps > 0.307$). Replicating results from Experiment 1, we found need for cognition was negatively related to selective exposure ($b = -0.31$, 95% confidence interval lower $= -0.58$, upper -0.03, $p = 0.027$) and the interactions between need for cognition and attitude congruence were non-significant ($ps > 0.088$).

3.3. Discussion

In Experiment 2, we replicated our previous findings supporting our social vigilantism hypothesis: once again, higher levels of social

---

[4] https://www.youtube.com/watch?v=lxRFeBr-KK4

vigilantism were found to be related to stronger intentions to counterargue. We additionally replicated the finding that the relationship between social vigilantism and counterarguing was not moderated by the attitude congruence of the persuasive message. Replicating this finding gives us stronger confidence in concluding people who perceive their beliefs are superior and try to impress them onto others, may feel a greater need to argue in response to persuasion attempts—regardless of whether that attempt is congruent or incongruent with a strongly-held attitude.

Although social vigilantism did not correlate with intentions to ignore attitude-relevant messages in Experiment 1, we found a positive relationship between social vigilantism and selective exposure in the current experiment. The only difference between the two experiments was the inclusion of a less controversial message. It is possible that social vigilantism could be related to ignoring persuasive messaging about uncontroversial issues and including the neutral video increased the correlation between social vigilantism and selective exposure. However, prior research has shown that social vigilantism is related to more active resistance strategies (e.g., counterarguing, impressing one's own views on others) and unrelated to selective exposure regardless of the importance of the issue (Saucier et al., 2014, Study 2) suggesting that social vigilantism should be unrelated to selective exposure even for less controversial issues. We provided a further test of the relationship between social vigilantism and selective exposure in Experiment 3.

4. Experiment 3

Experiment 3 tested if the self-reported counterarguing and selective exposure intentions in Experiments 1 and 2 are related to participants' subsequent memory performance. Based on the resistance strategy results, people high in social vigilantism should be more engaged with all material, which would likely result in better memory for all video material. Because social vigilantism was inconsistently related to self-reported selective exposure across Experiments 1 and 2, memory for the contents of a persuasive message may provide a better test of selective exposure behavior. If social vigilantism is related to memory for attitude-relevant information, then it would suggest that social vigilantism is at best unrelated, if not negatively related, to selective exposure. Furthermore, in Experiments 1 and 2, people with higher levels of need for cognition showed less selective exposure for attitude-congruent and -incongruent information, so they may as a result have better memory for attitude-relevant information.

Importantly, the behaviors involved in selective exposure and counterarguing are inherently related to attention (i.e., selective exposure assumes fewer attentional resources are used to process counter-attitudinal information). As such, it is important to consider the effect the stimuli used may have on attention. Recent work on visual attention to videos shows that highly produced films and advertisements create a phenomenon known as attentional synchrony (Dorr et al., 2010; Smith and Mital, 2013), in which people show high convergence in where they look in videos on a moment-to-moment basis. Further, attentional synchrony persists despite large differences in top-down processes, which has been termed *the tyranny of film* (Loschky et al., 2015; Hutson et al., 2017). Given that memory is highly correlated with what a person attends to Loftus (1972), Hollingworth and Henderson (2002), Tatler et al. (2005), Zelinsky and Loschky (2005), and Pertzov et al. (2009), if the ads guide attention despite differences in attitude congruence, there could be a dissociation between participants reported resistance strategies from Experiments 1 and 2, and their memory. In other words, participants may believe they are engaging with the content differently due to their beliefs, but the video composition may be involuntarily guiding their attention and controlling what they recall.

4.1. Memory experiment hypotheses

4.1.1. Selective exposure

The selective exposure hypothesis predicts participants will have better memory for attitude-congruent vs. attitude-incongruent information.

4.1.2. Social vigilantism

Based on the results from Experiments 1 and 2, the social vigilantism hypothesis predicts that because social vigilantism was positively correlated with intentions to counterargue, participants higher in social vigilantism will be more likely to engage in processing attitude-congruent and attitude-incongruent information. This would result in better memory for both abortion ads but show no relationship for the less controversial video. Alternatively, if social vigilantism is positively related to selective exposure, participants higher in social vigilantism should show worse memory for the information in the attitude-incongruent videos because selective exposure is a resistance strategy that involves ignoring information that is inconsistent with one's attitude.

4.1.3. Need for cognition

Based on the results from Experiments 1 and 2, the need for cognition hypothesis predicts that because need for cognition was negatively related to selective exposure, participants higher in need for cognition will attend more to the information, and thus have better memory for the content of *all* of the videos.

4.1.4. Tyranny of film (null)

The tyranny of film hypothesis predicts that, due to the control filmmakers have over what information is presented, participants will have similar memory for the ads regardless of their attitudes, social vigilantism, or need for cognition.

4.2. Methods

4.2.1. Participants

A new sample of 118 participants were recruited from introductory psychology courses at Kansas State University to participate in the experiment (ages 18–41 [$M = 19.8$, $SD = 3.1$], 57% female, 86% were White). The obtained sample size provided >80% power to detect effect sizes > 0.10. All data were collected prior to conducting our analyses.

4.2.1.1. Individual difference scores

Participants completed the same individual difference measures for the memory experiment as in Experiments 1 and 2 (abortion attitudes $M = 4.89$, $SD = 2.82$; social vigilantism $M = 4.98$, $SD = 1.22$; need for cognition $M = 5.28$, $SD = 1.06$). Most participants identified as being either strongly pro-life or pro-choice, with a smaller group of participants indicating they had no strong attitude one way or the other. There were roughly an equal number of participants who identified as pro-life and pro-choice. In the current experiment, we elected to include the complete range of abortion attitude scores to use the full power of our sample size.

4.3. Stimuli

4.3.1. Videos

Participants viewed the same abortion ads as in Experiments 1 and 2 and the same non-controversial ad as in Experiment 2.

4.3.2. Memory test items

The memory test stimuli were developed to measure both recall and recognition memory, and visual and verbal memory. Free recall memory was of interest, because previous work has shown it may be more susceptible to top-down effects than recognition memory (Mandler, 2008). For free recall memory, participants were given prompts to recall as much verbal and then visual information as possible, as if they were explaining the video to a friend who had not seen it.

Participants completed 3 types of recognition memory items: argument recognition, visual multiple choice, and visual recognition. Argument recognition memory items presented participants with an argument, and they indicated whether it was worded exactly as in the ad they saw, or if it was reworded in some way (e.g., synonyms were used and/or verb tense was changed). Visual multiple-choice questions had a stem asking about a visual element of an ad (e.g., "What fruit was shown in the ad?"). Each question had four answer options. Finally, Visual Recognition memory items used video stills taken from the ad. For these items, participants indicated whether the image was presented as it originally was in the ad, or if it was mirror reversed (i.e., left/right reversed). All items were scored as correct (1) or incorrect (0).

Importantly, we used memory measures for different modalities and different levels of representation (Van Dijk and Kintsch, 1983), because memory effects could be different based on the how the viewer interacts with an ad. For example, if a viewer engages in counterarguing, they may be more likely to counterargue the text information presented, rather than the images. As such, the influence of counterarguing could be specific to memory for the text.

4.4. Procedure

The current experiment was conducted online via Qualtrics. Participants first watched all three ads (with order randomized for each participant) so that we could compare memory for each ad within-participants. After the videos, participants responded to the memory questions. The questions were organized into blocks based on their type in the following order for all participants: free recall, argument recognition, visual multiple choice, and visual recognition. Free recall questions were presented before argument recognition questions so that the presentation of the recognition items did not influence responses to the free recalls. Similarly, visual multiple-choice questions were presented before the visual recognition memory items so that seeing the visual recognition items could not influence responses to the visual multiple-choice questions.

4.5. Analyses

We ran multilevel logistic regressions separately for each video type (non-controversial & abortion ads), as well as for the different types of memory items. The random effects structure for all analyses included the participant and memory item. This random effects structure was determined to be best based on AIC values when compared to a random effects structure that only included the participant (Burnham and Anderson, 2004). To determine the best fixed effects (predictor) structure most likely to generalize at the population level, we used a model testing procedure, and selected the best model using AIC values to reduce the likelihood of Type I errors.

4.5.1. Signal detection analyses

The Argument and Visual Recognition memory items used an Old (i.e., seen in the experiment video)/New (i.e., not seen in the experiment video) format, which allowed us to use signal detection analyses. We conducted the signal detection analyses using logistic multilevel models with the probit link function (DeCarlo, 1998; Wright et al., 2009).

There are a few keys to interpreting signal detection analyses done with multilevel models. First, the intercept of the model is the overall bias (i.e., c = decision *criterion,* do participants disproportionately respond "Old" or "New"?). Second, the first predictor in the model is always whether the memory item was an "Old" or "New" item, and this is the overall sensitivity (i.e., d', the ability to discriminate between "Old" and "New" items). Effects of predictors on bias are indicated by adjustments to the intercept. Effects of the predictors on sensitivity are indicated by their interaction with "Old/New."

4.5.2. Free recall scoring

Unfortunately, participant free recall responses were very short ($M = 19.4$ words; $SD = 15.2$ words), and some participants simply wrote that they did not remember the video. This resulted in a floor effect, and there were no clear effects or trends with any of the predictor variables. As such the free recall data is not presented here.

4.6. Results

4.6.1. Results overview

Overall, Experiment 3 showed partial support for the tyranny of film and social vigilantism hypotheses, as well as some general top-down effects independent of attitude congruence that were not hypothesized *a priori*. Interestingly, many of these effects were found across measures and videos. Also, in contrast to Experiments 1 and 2, we found no effects of need for cognition on any of the memory measures, thus providing no support for the need for cognition hypothesis in terms of memory.

4.7. Non-controversial ad

4.7.1. Argument recognition memory

The best model only included attitude and social vigilantism as predictors (i.e., the simplest model). As expected, given the non-controversial nature of the ad, neither predictor influenced sensitivity or bias (all p's > 0.05). In other words, the model indicated individual differences did not significantly predict argument recognition memory for the non-controversial ad. Overall, for the non-controversial video argument recognition items, participants had very low sensitivity ($d' = 0.16$, $z = 0.90$, $p = 0.368$), but they showed a strong "Old" bias ($c = 0.62$, $z = 6.99$, $p < 0.001$). Thus, participants did well for unchanged "Old" items, but were well below chance for reworded "New" memory items.

4.7.2. Visual recognition memory

Surprisingly, given the non-controversial nature of the ad, for visual recognition memory the best model included the interaction of attitude and social vigilantism with recognition memory item type ("Old" vs. "Mirror reversed"). As shown in Table 2, participants were sensitive to the visual recognition memory items, and pro-choice participants showed higher overall sensitivity. Interestingly, however, the interaction of attitude and social vigilantism influenced sensitivity. As shown in Figure 1A, the interaction between attitude and social vigilantism on sensitivity created an "arch" shape. Specifically, for people higher in social vigilantism, pro-life participants were more sensitive than those who were pro-choice; conversely, for people lower in social vigilantism, this relationship reversed—pro-choice participants were more sensitive than pro-life participants (Figure 1A). Surprisingly, these results show, even when a video is on a *non*-controversial topic, a person's attitude toward a controversial topic and their level of social vigilantism, can interact to influence their visual recognition memory. None of the independent variables significantly influenced bias.

4.7.3. Visual multiple choice

The three individual difference variables in the accuracy model showed no significant effects on participants' memory for visual details (p's > 0.05). The performance predicted by the model was relatively low, 38%, but significantly above chance performance (25%).

When taken together, the recognition memory item results for the non-controversial ad were mostly consistent with our expectation that differences in abortion attitudes and social vigilantism would not have effects. However, there was one exception. For visual recognition memory, there were effects of attitude, social vigilantism, and their interaction.

4.8. Abortion ads

4.8.1. Argument recognition memory

As with the non-controversial ad, there was only a significant "Old" bias. Participants were more likely to indicate that memory items had appeared in the video ($c = -0.61$, $z = 5.39$, $p < 0.001$). Although none of the individual difference measures were significant, there was a non-significant trend toward an interaction of attitude congruence and social vigilantism on sensitivity ($b = -0.04$, $z = -1.76$, $p = 0.08$). Participants lower in social vigilantism tended to show higher sensitivity for the attitude-congruent video, but participants higher in social vigilantism tended to show better memory for the attitude-incongruent video. In contrast to Experiments 1 and 2 where social vigilantism was related to counterarguing regardless of the attitude congruence of the persuasive message, our memory results suggest that social vigilantism was positively related to engaging with attitude-incongruent messages and negatively related to engaging with attitude-congruent messages.

4.8.2. Visual recognition memory

For the abortion ads, we replicated some of the attitude and social vigilantism effects found for the non-controversial ad. The best model included "Old"/"New," Attitude, and social vigilantism, but there were no attitude congruence effects. Overall, as shown in Table 3, there was an "Old" bias, and participants were sensitive to the memory items. Interestingly, participants who were more pro-choice had higher sensitivity. This effect was not predicted.

4.8.3. Visual multiple choice

Consistent with the hypothesis that social vigilantism would be related to better memory, higher levels of social vigilantism related to better memory for attitude-*in*congruent content (Table 4; Figure 2). This U-shaped pattern is especially clear for the pro-life video. At lower levels of social vigilantism, pro-life

TABLE 2 Summary of multilevel logistic signal detection analysis for non-controversial ad visual recognition memory.

Variable	*B*	*SE(B)*	*z*	Sig. (*p*)
Intercept [Bias]	0.29	0.17	1.79	0.072
"Old"/"New" [sensitivity]	0.73	0.33	2.23	0.026
Attitude	−0.005	0.02	−0.25	0.805
Social vigilantism	−0.02	0.04	−0.43	0.664
"Old"/"New" × attitude	0.10	0.04	2.80	0.005
"Old"/"New" × SV	0.03	0.09	0.38	0.703
Attitude × SV	−0.005	0.01	−0.37	0.713
"Old"/"New" × attitude × SV	−0.07	0.03	−2.51	0.012

The intercept of the model is the overall bias. "Old"/"New" shows the overall sensitivity to the memory items. Attitude, Social Vigilantism (SV), and Attitude × SV show adjustments to bias. Interactions with "Old"/"New" show adjustments to sensitivity. The continuous variables were centered for the interaction. Statistically significant effects are shaded.

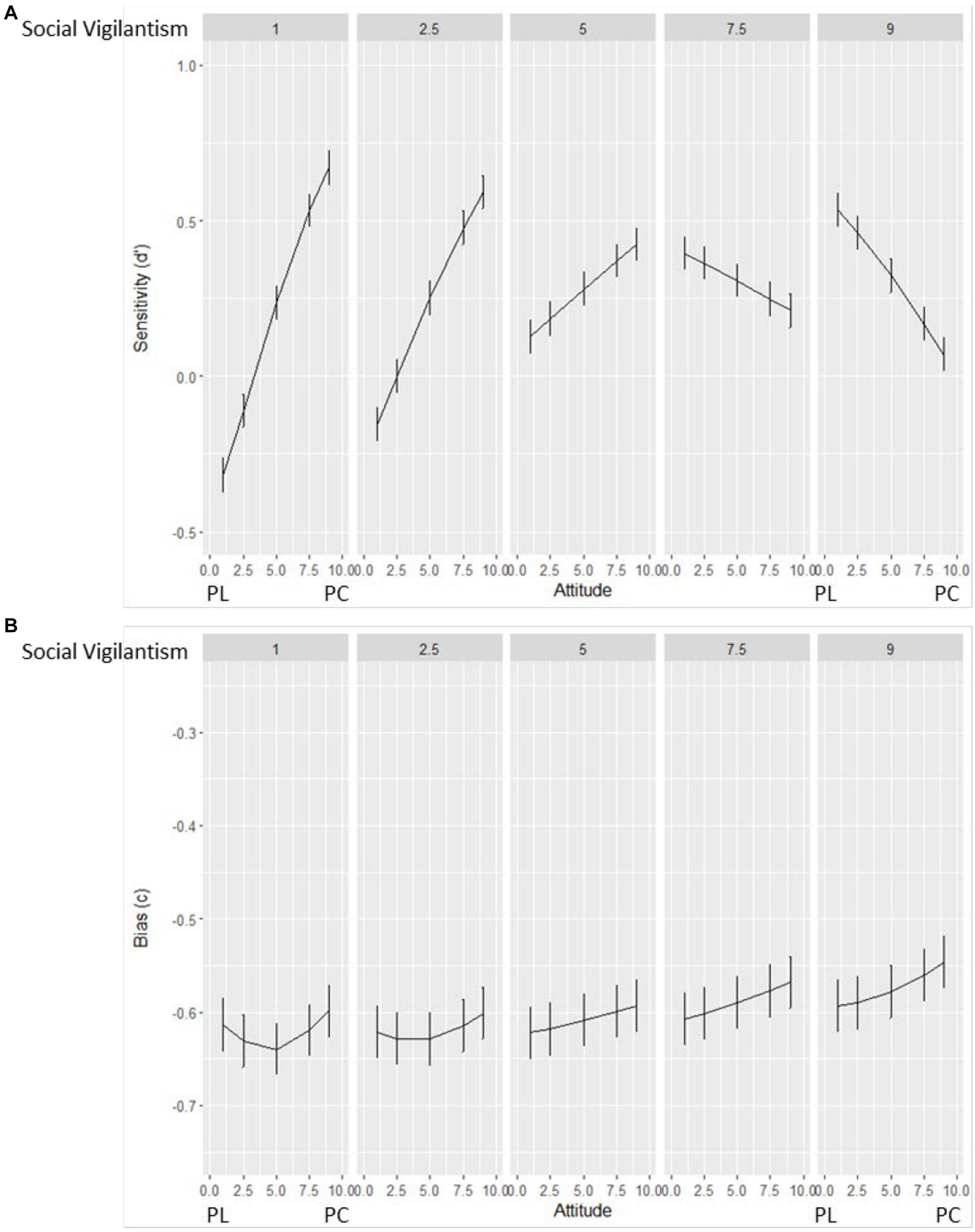


FIGURE 1
Signal detection analysis for Non-controversial visual recognition memory. **(A)** The Y-axis is the predicted sensitivity (d′) to the visual recognition memory items. The X-axis is attitude PL = pro-life, PC = pro-choice (1 = most pro-life; 9 = most pro-choice). The panels labeled at the top of the graph are cross sections of the social vigilantism (SV) measure (1 = Very low in SV; 9 = Very high in SV). **(B)** The Y-axis is the predicted bias (c). All other axes are the same as for **(A)** (Attitude on the X-axis and Social vigilantism for the panels). Error bars are 1 standard error.

participants showed better memory, and, at higher levels of social vigilantism, pro-choice participants showed better memory. For the pro-choice video, we found approximately the same general trend in reverse, a partial "arch" pattern, similar to visual recognition memory for the non-controversial ad (Figure 1A), and the non-significant trend we found for argument recognition memory. In this case, at lower levels of social vigilantism, pro-choice participants had better memory for the pro-choice video; however, the slope did not reverse direction at higher levels of social vigilantism.

A trend that emerged from the abortion ad memory results was that social vigilantism moderated the effect of attitude congruence on memory performance. People lower in social vigilantism had better memory for attitude-congruent items, while those higher in social vigilantism had better memory for attitude-*in*congruent items—which produces a "U" pattern, or conversely an arch pattern. Although this effect was only

TABLE 3 Summary of multilevel logistic signal detection for abortion ad visual recognition memory.

Variable	*B*	*SE(B)*	*t*	Sig. (*p*)
Intercept [Bias]	−0.46	0.08	5.60	<0.001
"Old"/"New" [sensitivity]	0.79	0.16	4.80	<0.001
Attitude	0.0004	0.009	−0.04	0.97
Social vigilantism	0.03	0.02	−1.45	0.15
"Old"/"New" × attitude	0.05	0.02	2.58	0.01
"Old"/"New" × SV	−0.06	0.05	−1.22	0.22
Attitude × SV	−0.01	0.007	1.41	0.16
"Old"/"New" × attitude × SV	−0.01	0.01	−1.07	0.28

The intercept of the model is the overall bias. "Old"/"New" shows the overall sensitivity to the memory items. Attitude, Social vigilantism (SV), and Attitude × SV show adjustments to bias. Interactions with "Old"/"New" show adjustments to sensitivity. The continuous variables were centered for the interaction. Statistically significant effects are shaded.

TABLE 4 Summary of multilevel logistic for abortion ad visual multiple-choice memory.

Variable	*B*	*SE(B)*	*z*	Sig. (*p*)
Intercept	−0.26	0.23	−0.94	0.349
Attitude	0.02	0.02	0.90	0.367
Social vigilantism	−0.01	0.06	−0.23	0.816
Video	−0.33	0.23	−1.47	0.141
Att. × SV	0.02	0.02	1.10	0.271
Att. × video	0.04	0.02	2.20	0.028
Video × SV	0.07	0.05	1.5	0.134
Att. × SV × video	−0.03	0.01	−2.33	0.020

Describes model for predicted accuracy for the visual multiple-choice memory questions. The continuous variables were centered for the interaction. Statistically significant effects are shaded. SV = social vigilantism.

significant for the visual multiple-choice questions, the argument recognition items trended in the same direction. It is interesting that social vigilantism moderated the effects of attitude congruence for items based on visual information and trended for argument items, but not for the visual recognition items. One possibility, based on Mandler's (2008) dual process model of recognition memory, is immediate visual recognition memory operates at a perceptual level (Langley et al., 2008) that may not be affected by top-down (here, attitude or personality-driven) processing.

4.9. Discussion

Both attitude congruence and social vigilantism influenced memory for political ads, but the presence of effects varied with memory type (argument recognition vs. visual multiple choice). First, for the non-controversial ad, which was included as a baseline condition, viewers' visual recognition memory surprisingly showed differences based on their attitudes and level of social vigilantism. Participants who were pro-choice were more sensitive for visual recognition memory items (which was also found for the abortion ads). Thus, the non-controversial ad showed top-down effects of attitude and social vigilantism on memory, even for a non-controversial topic. Additionally, the arch shaped interaction created in Figure 1A was also partially found in Figure 2 for the abortion ads visual multiple-choice memory measure.

Importantly, the abortion ad results showed effects of both attitude and social vigilantism. At higher levels of social vigilantism, participants showed better memory for attitude-*in*congruent information. This effect was strongest for the visual multiple-choice measure, and somewhat less so for argument recognition. These findings combine to indicate an effect of attitude congruence, social vigilantism, and their interaction on memory for politically controversial content. However, in contrast to Experiments 1 and 2, we found no support for the need for cognition hypothesis.

An interesting insight from the current work and an area for future inquiry is the level of memory representation at which the top-down effects occurred. Van Dijk and Kintsch (1983) proposed three levels of representation for text: surface, propositional, and situational. Subsequent work verified people do encode and retrieve all three levels of representation (Schmalhofer and Glavanov, 1986; Fletcher and Chrysler, 1990). In the current work, the recognition memory items manipulated the surface structure of the arguments presented in the videos, but the propositional representations of the text base remained the same (i.e., the wording of the arguments changed, but the meaning stayed the same). Thus, the recognition memory results showed individual difference effects at the surface level.

5. General discussion

We live in a highly politically divisive era, in which political discourse is increasingly insular. People are continuously exposed to political video content through various media, which is increasingly filtered to create

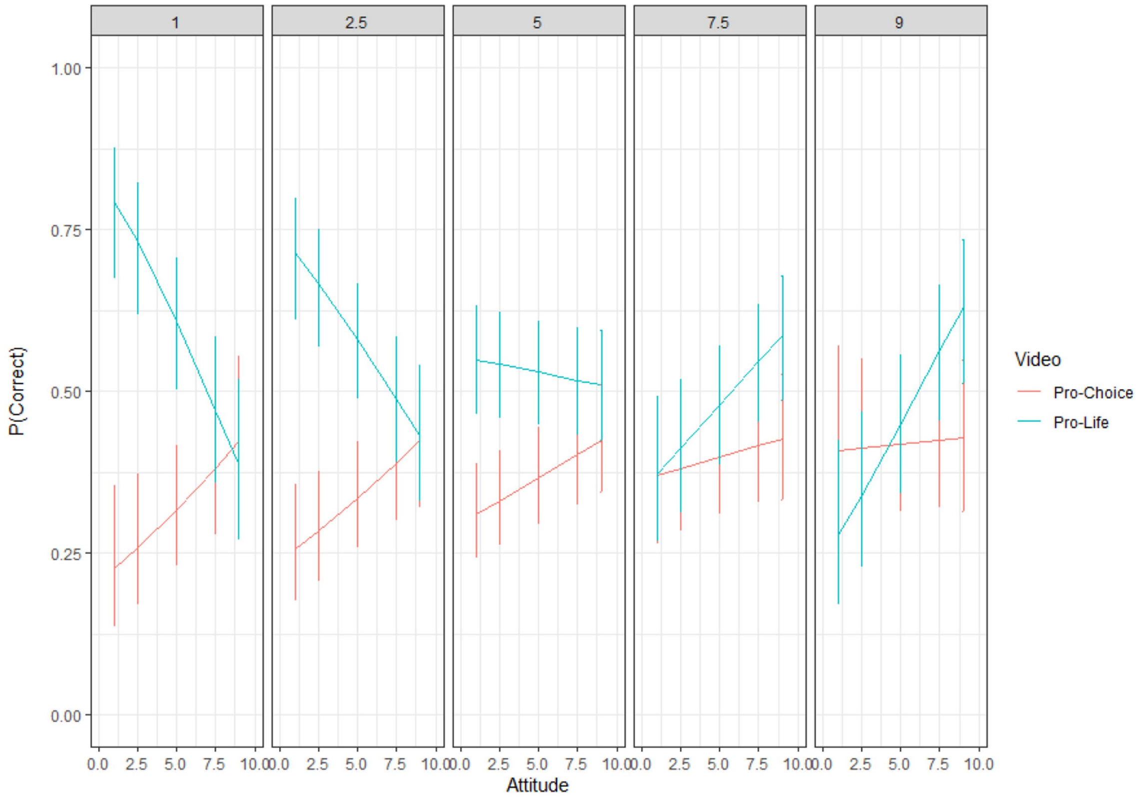


FIGURE 2
Abortion ads' visual multiple choice. Y-axis is predicted accuracy. The X-axis shows abortion attitude (low scores = more pro-life; high scores = more pro-choice). The panels labeled at the top of the graph are cross sections of the social vigilantism (SV) measure (1 = very low in SV; 9 = very high in SV). Error bars are 1 standard error.

political "echo chambers" consistent with people's pre-existing beliefs. Highly produced videos, such as political ads, are designed to create similar attentional patterns across viewers, which could produce similar memory across viewers as well. Some people, who are high in social vigilantism, choose to actively debate with those with whom they disagree, often trying to persuade them to change their minds, while others, who are high in need for cognition, may be less argumentative, but still greatly value thinking through difficult problems. Such individual differences could create large differences in people's memory for the contents of political videos. This inspires our key question: how do individual differences in social vigilantism and need for cognition interact with the attitude congruence of political videos in determining viewers' reported resistance strategies and what they remembered from such videos?

We tested several hypotheses about how social vigilantism and need for cognition would interact with responses to political content that was congruent vs. incongruent with viewers' attitudes. The Tyranny of Film hypothesis predicted individual differences in social vigilantism and need for cognition would not be predictive of viewers' memory, because video makers are experts at guiding viewers' attention and the information available to process. The three alternative competing hypotheses stated people may either (i) avoid or tune-out counter-attitudinal information (selective exposure), or (ii) engage with incongruent information more if they had higher belief superiority and desire to impress their beliefs on others (social vigilantism), or (iii) that viewers' who enjoy engaging in demanding cognitive tasks would engage more (need for cognition).

In Experiments 1 and 2, we found evidence of both need for cognition and social vigilantism affecting the resistance strategies adopted when viewing videos containing controversial political content congruent vs. incongruent with their attitudes. As expected, people reported greater intentions to resist persuasive video content that was attitude-incongruent in terms of counterarguing and selective exposure. However, consistent with the social vigilantism hypothesis, those higher in social vigilantism were more likely to report the intention to counterargue than those who were lower in social vigilantism. Interestingly, people higher in social vigilantism were more likely to report intentions to counterargue even when the video content was attitude-congruent. We also found evidence for the need for cognition hypothesis, with people higher in need for cognition reporting lesser intent to practice selective exposure to ads inconsistent with their attitudes.

In Experiment 3, we investigated whether these reported intentions were consistent with viewers' memory for the same political videos and found those higher in social vigilantism had better recognition memory for ad content that was attitude-incongruent. Notably, for people lower in social vigilantism these results showed evidence consistent with selective exposure, though in Experiment 1, participants did not report intending to ignore attitude-incongruent political video content, and in Experiment 2, attitude-congruence did not moderate the significant positive relationship between social vigilantism and intention to engage in selective exposure. Thus, the data across Experiments 1–3 show a degree of divergence between our self-report measures and memory for information one disagrees with. Findings from Experiment 3 are also consistent with the hypothesis that higher levels of social vigilantism would be associated with counterarguing resistance strategies, because the better memory for attitude-incongruent content suggests people higher in social vigilantism might be motivated to gather information about opposing viewpoints in order to directly challenge those arguments. Experiment 3 found no support for the need for cognition hypothesis. Nevertheless, Experiments 1 and 2 showed that participants higher in need for cognition were less likely to report intentions to engage in selective exposure for attitude-incongruent information. Future work should address this inconsistency.

Interestingly, we also found effects of attitude, but not attitude congruence, and social vigilantism, for memory for the non-controversial ad. This ad was included as a baseline measure, for which we expected to find no attitude effects. Nevertheless, viewers' (seemingly) irrelevant attitudes toward a different and highly controversial political topic (abortion) produced differences in their memory for the *non*-controversial ad. The same pattern of results (arch or "U" shape) was also partially found for the visual multiple choice measure for the abortion ads. Future work will test whether this is a reliable relationship, which may be similar in nature to the effects of political ideology on attention (Dodd et al., 2012).

5.1. Limitations and future directions

People do not always behave in ways consistent with how they say they will behave. Thus, one limitation of Experiments 1 and 2 is that participants self-reported their intentions to engage in counterarguing and selective exposure. It may be socially undesirable for people to claim they would ignore or counterargue against attitude-incongruent persuasion attempts, and this may have affected how participants self-reported their intentions to engage in resistance strategies. Additionally, in Experiments 1 and 2, higher levels of social vigilantism were related to stronger intentions to counterargue regardless of whether the video was congruent or incongruent with participants' attitudes. This finding might have been due to participants attempting to maintain consistent responses across the counterarguing measures and the social vigilantism items pertaining to argumentative tendencies. However, our memory measures in Experiment 3, which did not have these same limitations, provided further support for the social vigilantism hypothesis. To be more confident in concluding higher levels of social vigilantism are related to counterarguing against attitude-incongruent positions, future research could also assess whether higher levels of social vigilantism are associated with paying more attention to attitude-incongruent information. For example, eye movement or eye blink measures (e.g., Nakano et al., 2009; Andreu-Sánchez et al., 2021a,b) would be more direct measures of attention.

Additionally, because the social vigilantism scale seems to capture both belief superiority as well as motivations to impress those beliefs onto others, it would be interesting to examine which of these two constructs is more strongly associated with counterarguing behaviors. This could be studied by including both the social vigilantism scale and the general belief superiority scale (Raimi and Jongman-Sereno, 2020) as competing predictors of counterarguing behavior. It may be that belief superiority alone would be sufficient for eliciting several of the resistance to persuasion strategies identified in past research, but we suspect that the motivation to impress one's beliefs onto others would more strongly predict counterarguing specifically.

5.2. Conclusion

Our research contributes to our understanding of how people resist politically charged attempts at persuasion by showing how higher levels of social vigilantism and need for cognition are related to greater intentions to engage with, rather than tune out, information that opposes their strongly held attitudes. The results of our memory study further demonstrate how levels of social vigilantism relate to cognitive processes that may facilitate engagement through better memory for opposing arguments, perhaps because individuals higher in social vigilantism are tracking the information they are motivated to argue against. Together, our findings highlight the significant role individual differences play in how people process and respond to attempts to change their strongly held attitudes.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/k6j3z/?view_only=ed561b0a2a7b4f9ab5ece40c64489211.

Ethics statement

The studies involving human participants were reviewed and approved by Kansas State University Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

SM: manuscript writing and experimental design, materials, data collection, and analyses for Experiments 1 and 2. JH: manuscript writing and experimental design, materials, data collection, and analyses for Experiment 3. MS: manuscript writing and study design, materials, data collection, and analyses for the pilot study. TS, LL, and DS: manuscript writing and advised on the design for all three experiments. MP: creation of video stimuli used in all three experiments. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1196209/full#supplementary-material

References

Albarracín, D., and Mitchell, A. L. (2004). The role of defensive confidence in preference for proattitudinal information: how believing that one is strong can sometimes be a defensive weakness. *Pers. Soc. Psychol. Bull.* 30, 1565–1584. doi: 10.1177/0146167204271180

Andreu-Sánchez, C., Martín-Pascual, M. Á., Gruart, A., and Delgado-García, J. M. (2021a). Viewers change eye-blink rate by predicting narrative content. *Brain Sci.* 11:422. doi: 10.3390/brainsci11040422

Andreu-Sánchez, C., Martín-Pascual, M. A., Gruart, A., and Delgado-García, J. M. (2021b). The effect of media professionalization on cognitive neurodynamics during audiovisual cuts. *Front. Syst. Neurosci.* 15:598383. doi: 10.3389/fnsys.2021.598383

Brannon, L. A., Tagler, M. J., and Eagly, A. H. (2007). The moderating role of attitude strength in selective exposure to information. *J. Exp. Soc. Psychol.* 43, 611–617. doi: 10.1016/j.jesp.2006.05.001

Burnham, K. P., and Anderson, D. R. (2004). Multimodel Inference. *Sociol. Methods Res.* 33, 261–304. doi: 10.1177/0049124104268644

Cacioppo, J. T., and Petty, R. E. (1982). The need for cognition. *J. Pers. Soc. Psychol.* 42, 116–131. doi: 10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., and Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: the life and times of people varying in need for cognition. *Psychol. Bull.* 119, 197–253. doi: 10.1037/0033-2909.119.2.197

Cacioppo, J. T., Petty, R. E., Kao, C. F., and Rodriguez, R. (1986). Central and peripheral routes to persuasion: an individual difference perspective. *J. Pers. Soc. Psychol.* 51, 1032–1043. doi: 10.1037/0022-3514.51.5.1032

Cameron, K. A., Jacks, J. Z., and O'Brien, M. E. (2002). An experimental examination of strategies for resisting persuasion. *Curr. Res. Soc. Psychol.* 7, 205–224.

Crano, W. D., and Crislin, R. (2006). Attitudes and persuasion. *Annu. Rev. Psychol.* 57, 345–374. doi: 10.1146/annurev.psych.57.102904.190034

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychol. Methods* 3, 186–205. doi: 10.1037/1082-989X.3.2.186

Dodd, M. D., Balzer, A., Jacobs, C. M., Gruszczynski, M. W., Smith, K. B., and Hibbing, J. R. (2012). The political left rolls with the good and the political right confronts the bad: connecting physiology and cognition to preferences. *Philosoph Transact R Soc B Biol Sci* 367, 640–649. doi: 10.1098/rstb.2011.0268

Dorr, M., Martinetz, T., Gegenfurtner, K. R., and Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *J. Vis.* 10, 1–28. doi: 10.1167/10.10.28

Eagly, A. H., and Chaiken, S. (1995). "Attitude strength, attitude structure, and resistance to change" in *Attitude strength: Antecedents and consequences*. eds. R. E. Petty and J. A. Krosnick (Mahwah, NJ: Lawrence Erlbaum Associates Inc), 413–432.

Fletcher, C. R., and Chrysler, S. T. (1990). Surface forms, textbases, and situation models: recognition memory for three types of textual information. *Discourse Process.* 13, 175–190. doi: 10.1080/01638539009544752

Haddock, G., Maio, G. R., Arnold, K., and Huskinson, T. (2008). Should persuasion be affective or cognitive? The moderating effects of need for affect and need for cognition. *Pers. Soc. Psychol. Bull.* 34, 769–778. doi: 10.1177/0146167208314871

Haugtvedt, C. P., and Petty, R. E. (1992). Personality and persuasion: need for cognition moderates the persistence and resistance of attitude changes. *J. Pers. Soc. Psychol.* 63, 308–319. doi: 10.1037/0022-3514.63.2.308

Hollingworth, A., and Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *J. Exp. Psychol. Hum. Percept. Perform.* 28:113. doi: 10.1037/0096-1523.28.1.113

Hutson, J. P., Smith, T. J., Magliano, J. P., and Loschky, L. C. (2017). What is the role of the film viewer? The effects of narrative comprehension and viewing task on gaze control in film. *Cognit Res Principles Implic* 2:46. doi: 10.1186/s41235-017-0080-5

Jacks, J. Z., and Cameron, K. A. (2003). Strategies for resisting persuasion. *Basic Appl. Soc. Psychol.* 25, 145–161. doi: 10.1207/S15324834BASP2502_5

Jacks, J. Z., and Devine, P. G. (2000). Attitude importance, forewarning of message content, and resistance to persuasion. *Basic Appl. Soc. Psychol.* 22, 19–29. doi: 10.1207/S15324834BASP2201_3

Knobloch-Westerwick, S., Mothes, C., and Polavin, N. (2020). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Commun. Res.* 47, 104–124. doi: 10.1177/0093650217719596

Krosnick, J. A., Boninger, D. S., Chuang, Y. C., Berent, M. K., and Carnot, C. G. (1993). Attitude strength: one construct or many related constructs? *J. Pers. Soc. Psychol.* 65, 1132–1151. doi: 10.1037/0022-3514.65.6.1132

Krosnick, J. A., and Petty, R. E. (1995). "Attitude strength: an overview" in *Attitude strength: Antecedents and consequences*. eds. R. E. Petty and J. A. Krosnick (Mahwah, NJ: Lawrence Erlbaum Associates Inc), 1–24.

Langley, M. M., Cleary, A. M., Kostic, B. N., and Woods, J. A. (2008). Picture recognition without picture identification: a method for assessing the role of perceptual information in familiarity-based picture recognition. *Acta Psychol. (Amst)* 127, 103–113. doi: 10.1016/j.actpsy.2007.03.001

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cogn. Psychol.* 3, 525–551. doi: 10.1016/0010-0285(72)90021-7

Loschky, L. C., Larson, A. M., Magliano, J. P., and Smith, T. J. (2015). What would jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLoS One* 10, 1–23. doi: 10.1371/journal.pone.0142474

Maki, A., and Raimi, K. T. (2017). Environmental peer persuasion: how moral exporting and belief superiority relate to efforts to influence others. *J. Environ. Psychol.* 49, 18–29. doi: 10.1016/j.jenvp.2016.11.005

Mandler, G. (2008). Familiarity breeds attempts: a critical review of dual-process theories of recognition. *Perspect. Psychol. Sci.* 3, 390–399. doi: 10.1111/j.1745-6924.2008.00087.x

Nakano, T., Yamamoto, Y., Kitajo, K., Takahashi, T., and Kitazawa, S. (2009). Synchronization of spontaneous eyeblinks while viewing video stories. *Proc. Biol. Sci. R. Soc.* 276, 3635–3644. doi: 10.1098/rspb.2009.0828

O'Dea, C. J., Castro Bueno, A. M., and Saucier, D. A. (2018). Social vigilantism and the extremity, superiority, and defense of attitudes toward climate change. *Personal. Individ. Differ.* 130, 83–91. doi: 10.1016/j.paid.2018.03.040

Peltier, J. W., and Schibrowsky, J. A. (1994). Need for cognition, advertisement viewing time and memory for advertising stimuli. *Adv. Consum. Res.* 21, 244–250.

Pertzov, Y., Avidan, G., and Zohary, E. (2009). Accumulation of visual information across multiple fixations. *J. Vis.* 9, 2.1–2.12. doi: 10.1167/9.10.2

Petty, R. E., and Wegener, D. T. (1998). "Attitude change: multiple roles for persuasion variables" in *The handbook of social psychology*. eds. D. Gilbert, S. Fiske and G. Lindzey. *4th* ed (New York: McGraw-Hill)

Pomerantz, E. M., Chaiken, S., and Tordesillas, R. S. (1995). Attitude strength and resistance processes. *J. Pers. Soc. Psychol.* 69, 408–419. doi: 10.1037/0022-3514.69.3.408

Raimi, K. T., and Jongman-Sereno, K. P. (2020). General belief superiority (GBS): personality, motivation, and interpersonal relations. *Self Identity* 19, 546–571. doi: 10.1080/15298868.2019.1640785

Raimi, K. T., and Leary, M. R. (2014). Belief superiority in the environmental domain: attitude extremity and reactions to fracking. *J. Environ. Psychol.* 40, 76–85. doi: 10.1016/j.jenvp.2014.05.005

Ryu, S., and Vargas, P. (2021). Product visuals and consumers' selective exposure: the role of thought generation and cognitive motivation. *J. Mark. Commun.* 27, 780–798. doi: 10.1080/13527266.2021.1923556

Saucier, D. A., and Webster, R. J. (2010). Social vigilantism: measuring individual differences in belief superiority and resistance to persuasion. *Pers. Soc. Psychol. Bull.* 36, 19–32. doi: 10.1177/0146167209346170

Saucier, D. A., Webster, R. J., Hoffman, B. H., and Strain, M. L. (2014). Social vigilantism and reported use of strategies to resist persuasion. *Personal. Individ. Differ.* 70, 120–125. doi: 10.1016/j.paid.2014.06.031

Saucier, D. A., Webster, R. J., O'Dea, C. J., and Miller, S. S. (2017). "The role of individual differences in inciting anger and social action" in *Understanding angry groups: Multi-disciplinary perspectives on their motivations and effects on society*. eds. S. C. Cloninger and S. A. Leibo (Santa Barbara, CA: Praeger), 3–27.

Schmalhofer, F., and Glavanov, D. (1986). Three components of understanding of a programmer's manual: verbatim, propositional, and situational representations. *J Memory Lang* 25, 279–294. doi: 10.1016/0749-596X(86)90002-1

Skitka, L. J., Bauman, C. W., and Sargis, E. G. (2005). Moral conviction: another contributor to attitude strength or something more? *J. Pers. Soc. Psychol.* 88, 895–917. doi: 10.1037/0022-3514.88.6.895

Smith, T. J., and Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *J. Vis.* 13:16. doi: 10.1167/13.8.16

Tatler, B. W., Gilchrist, I. D., and Land, M. F. (2005). Visual memory for objects in natural scenes: from fixations to object files. *Q. J. Exp. Psychol.* 58, 931–960. doi: 10.1080/02724980443000430

Tsfati, Y., and Cappella, J. N. (2005). Why do people watch news they do not trust? The need for cognition as a moderator in the association between news media skepticism and exposure. *Media Psychol.* 7, 251–271. doi: 10.1207/S1532785XMEP0703_2

Van Dijk, T. A., and Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Visser, P. S., Bizer, G. Y., and Krosnick, J. A. (2006). Exploring the latent structure of strength-related attitude attributes. *Adv. Exp. Soc. Psychol.* 38, 1–67. doi: 10.1016/S0065-2601(06)38001-X

Visser, P. S., Krosnick, J. A., and Simmons, J. P. (2003). Distinguishing the cognitive and behavioral consequences of attitude importance and certainty: a new approach to testing the common-factor hypothesis. *J. Exp. Soc. Psychol.* 39, 118–141. doi: 10.1016/S0022-1031(02)00522-X

Westerwick, A., Johnson, B. K., and Knobloch-Westerwick, S. (2017). Confirmation biases in selective exposure to political online information: source bias vs. content bias. *Commun. Monogr.* 84, 343–364. doi: 10.1080/03637751.2016.1272761

Wright, D. B., Horry, R., and Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behav. Res. Methods* 41, 257–267. doi: 10.3758/BRM.41.2.257

Zelinsky, G. J., and Loschky, L. C. (2005). Eye movements serialize memory for objects in scenes. *Percept. Psychophys.* 67, 676–690. doi: 10.3758/BF03193524

Zuwerink, J. R., and Devine, P. G. (1996). Attitude importance and resistance to persuasion: It's not just the thought that counts. *J. Pers. Soc. Psychol.* 70, 931–944. doi: 10.1037/0022-3514.70.5.931

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 07 September 2023
DOI 10.3389/fnins.2023.1222472

OPEN ACCESS

EDITED BY
Celia Andreu-Sánchez,
Autonomous University of Barcelona, Spain

REVIEWED BY
Luz Maria Alonso-Valerdi,
Monterrey Institute of Technology and Higher Education (ITESM), Mexico
Danna Pinto,
Bar-Ilan University, Israel

*CORRESPONDENCE
Nunzio Langiulli
✉ nunzio.langiulli@unipr.it

RECEIVED 14 May 2023
ACCEPTED 14 August 2023
PUBLISHED 07 September 2023

CITATION
Langiulli N, Calbi M, Sbravatti V, Umiltà MA and Gallese V (2023) The effect of Surround sound on embodiment and sense of presence in cinematic experience: a behavioral and HD-EEG study. *Front. Neurosci.* 17:1222472.
doi: 10.3389/fnins.2023.1222472

The effect of Surround sound on embodiment and sense of presence in cinematic experience: a behavioral and HD-EEG study

Nunzio Langiulli[1]*, Marta Calbi[2], Valerio Sbravatti[3], Maria Alessandra Umiltà[4,5] and Vittorio Gallese[1,5]

[1]Unit of Neuroscience, Department of Medicine and Surgery, University of Parma, Parma, Italy, [2]Department of Philosophy "Piero Martinetti", State University of Milan, Milan, Italy, [3]Department of History, Anthropology, Religions, Arts and Performing Arts, Sapienza University of Rome, Rome, Italy, [4]Department of Food and Drug, University of Parma, Parma, Italy, [5]Italian Academy for Advanced Studies in America at Columbia University, New York, NY, United States

Although many studies have investigated spectators' cinematic experience, only a few of them explored the neurophysiological correlates of the sense of presence evoked by the spatial characteristics of audio delivery devices. Nevertheless, nowadays both the industrial and the consumer markets have been saturated by some forms of spatial audio format that enrich the audio-visual cinematic experience, reducing the gap between the real and the digitally mediated world. The increase in the immersive capabilities corresponds to the instauration of both the sense of presence and the psychological sense of being in the virtual environment and also embodied simulation mechanisms. While it is well-known that these mechanisms can be activated in the real world, it is hypothesized that they may be elicited even in a virtual acoustic spatial environment and could be modulated by the acoustic spatialization cues reproduced by sound systems. Hence, the present study aims to investigate the neural basis of the sense of presence evoked by different forms of mediation by testing different acoustic space sound delivery (Presentation modes: Monophonic, Stereo, and Surround). To these aims, a behavioral investigation and a high-density electroencephalographic (HD-EEG) study have been developed. A large set of ecological and heterogeneous stimuli extracted from feature films were used. Furthermore, participants were selected following the generalized listener selection procedure. We found a significantly higher event-related desynchronization (ERD) in the Surround Presentation mode when compared to the Monophonic Presentation mode both in Alpha and Low-Beta centro-parietal clusters. We discuss this result as an index of embodied simulation mechanisms that could be considered as a possible neurophysiological correlation of the instauration of the sense of presence.

KEYWORDS

HD-EEG, sense of presence, Surround sound, spatial audio, Alpha, Beta, ERD

Introduction

Cinema is a highly complex art form that combines visual and aural elements to create a cohesive and immersive experience. While the visual component of cinema has traditionally been the focus of both popular understanding and neuroscientific research (Heimann et al., 2014, 2019; Calbi et al., 2019), the role of sound in the cinematic experience has been largely overlooked. This bias toward the visual aspect can be attributed to a cultural tendency

(Sterne, 2003) to prioritize sight over hearing as well as the fact that the human brain is wired to process visual information more efficiently than auditory information (Kitagawa and Ichihara, 2002; Sbravatti, 2019). Previous research has demonstrated that when participants are simultaneously presented with movies depicting facial emotions and emotional sounds (such as crying and laughing) that are incongruent with each other, the electromyography (EMG) signals recorded from their facial muscles are activated in accordance with the visual stimuli and not with the auditory stimuli (Sestito et al., 2013).

On the other hand, even if empirical research on the relationship between moving images and sounds in cinema is limited, some authors have suggested that sound could enhance the immersive qualities of the two-dimensional cinematic experience (moving images) by creating a sense of three-dimensional reality (Elsaesser and Hagener, 2015). This concept is also supported by the idea that Surround sound formats, such as 5.1 channel configurations, have the capability to envelop the viewer in a 360-degree auditory space as opposed to the traditional 180-degree visual space (DiDonato, 2010). Some studies have been conducted to investigate the relationship between Surround sound and the sense of presence (see below for a definition) in the cinematic experience. For example, Västfjäll found that 6-channel audio reproductions received a significantly higher presence and emotional realism scores than stereo (2-channels) and mono (1-channel) reproductions (Västfjäll, 2003). Kobayashi et al. (2015) examined the influence of spatialized sounds (reproduced by a 96-channel system) on the sense of presence in virtual environments by using both physiological and psychological measures. Results showed that the presence ratings for sounds in the spatialized sounds condition were higher. Furthermore, physiological measures such as heart rate and skin conductance level indicated that the sympathetic nervous system was activated to a greater extent by sounds in the spatialized sounds condition similar to the responses elicited during intrusions into peri-personal space in real-world scenarios (such as clapping near the participant) (Kobayashi et al., 2015).

In a 1997 study, Slater and Wilbur critically examined for the first time the often confused concepts of immersion and presence, suggesting a way to disambiguate their meanings. The two authors defined immersion as an objective property of the technological playback system and presence as the subjective psychological experience of feeling situated in a mediated environment (Slater and Wilbur, 1997). The spatial situational model framework suggests that the experience of presence in a mediated environment is achieved through a two-step process (Wirth et al., 2007). The first step is the construction of a spatialized mental model of the mediated environment, in which participants can perceive the environment as a space and locate themselves within it. Certain features of the mediated environment are particularly important for the formation of a spatialized mental model, and one of these features is Surround sound among others such as stereoscopy and field of view (Wirth et al., 2003). The second step is the embodiment of the mediated environment. Gallese proposes that "film experience and film immersion do not depend just on concepts and propositions, but rely on sensory-motor schemas, which get the viewer literally in touch with the screen, shaping a multimodal form of simulation, which exploits all the potentialities of our brain–body system" (Gallese, 2019), referring to embodied simulation, a cognitive process described as the ability to simulate the actions, emotions, and sensations of others by activating the same neural circuits that are used to perceive one's own experiences. This mechanism allows individuals to recognize the meaning of others' behaviors and experiences by directly relating to them through the activation of sensory-motor representations in the bodily format (Gallese, 2009). The neural substrate of the embodied simulation mechanism for actions corresponds to a particular functional group of neurons called "mirror neurons," first discovered in area F5 of macaques during an intracortical recording of the premotor cortex that responds both during action execution and action observation (DiPellegrino et al., 1992). Mirror neurons allow for the internal representation of observed actions, which in turn facilitates understanding and imitation. According to Keysers et al., mirror neurons encode actions in an abstract manner, independent of the source of information (auditory or visual). This abstraction allows for multisensory integration, which is essential for generating meaningful representations and recognizing relevant actions within the environment (Keysers et al., 2003). In human beings, the mirror neuron mechanism is commonly associated with the mu rhythm typically recorded over sensorimotor centro-parietal cortical areas (Muthukumaraswamy and Johnson, 2004a,b; Muthukumaraswamy et al., 2004). The mu rhythm is an EEG measure of motor neuron activity considered to belong to the alpha band, generally ranging from 8 to 13 Hz, and the beta band, typically ranging between 14 and 32 Hz (Hari, 2006). When Gastaut and Bert (1954) initially observed the mu rhythm using EEG, they detected that this rhythm became less active, and there was an event-related desynchronization (ERD), when participants watched video clips of movements, but without exhibiting any visible motor movements themselves (Gastaut and Bert, 1954). Many subsequent studies observed a mu rhythm ERD, occurring during both voluntary movements, motor imagery, and action observation (Pfurtscheller et al., 1994; Toro et al., 1994; Pfurtscheller and Neuper, 1997; Neuper et al., 2006; Perry et al., 2010), and it has been proposed that this mu rhythm desynchronization represents activity in the mirror neuron system (e.g., Caetano et al., 2007; Perry and Bentin, 2009; Press et al., 2011).

The only study that investigated the effect of acoustic spatialization on the sense of presence using electroencephalography (EEG) was by Tsuchida et al. (2015). They used a surround sound reproduction system called BoSC (62 speakers), designed to simulate the presence of other individuals or objects by providing a highly realistic sound field, to deliver an acoustic stimulus under two experimental conditions: spatialized condition and monophonic condition (1-channel). EEG results showed that mu rhythm suppression occurred for action-related sounds but not for non-action-related sounds. Furthermore, this suppression was significantly greater in the Surround (62-channels) condition, which generates a more realistic sound field, than in the one-channel speaker condition. Additionally, the motor cortical activation for action-related sounds was influenced by the sense of presence perceived by the study participants as they perceived a significantly higher sound realism in the Surround condition (Tsuchida et al., 2015). It should be noted that this study had small participants and stimuli sample size, but only six action-related and non-action-related sounds were recorded and reproduced by an unconventional custom spatialized sound field system; hence, its results should be considered in light of the limitations of the study design. Further research with larger sample sizes and variegated stimuli is needed to fully understand the effect of acoustic spatialization on the sense of presence. Furthermore, the use of a standard surround sound reproduction system setup (such as 5.1-channel configurations) could ensure consistency and replicability compared to an unconventional setup.

Hence, this study aimed to investigate the time course and neural correlates of the sense of presence as evoked by different audio Presentation modes during cinematic immersion. We selected a diverse set of naturalistic stimuli, consisting of validated cinematic excerpts, which were presented to participants in different audio Presentation modes (Monophonic, Stereo, and Surround), while their neural and behavioral responses were measured. We first designed a behavioral experiment (Experiment 1) and subsequently a high-density electroencephalographic (HD-EEG) experiment (Experiment 2). Initially, in the context of the behavioral experiment, the sense of presence was rated by participants through explicit questions formulated to reflect its different aspects. The behavioral experiment was specifically designed to offer an initial investigation of the sense of presence with the aim to use results to guide the design of a subsequent EEG experiment. We hypothesized that participants exposed to the Surround presentation mode would report significantly higher subjective ratings compared to those exposed to the Monophonic and Stereophonic Presentation modes. Afterward, in the EEG experiment, we investigated the neural correlate of the sense of presence elicited by different acoustic Presentation modes. We hypothesized that the greater spatialization of sound in the Surround presentation mode, which more closely resembles a real-life hearing environment, would lead to a greater sense of embodiment as reflected by a higher ERD in the mu rhythm frequency band, compared to the Monophonic and Stereophonic presentation modes. This embodied simulation mechanism would be interpreted as a potential neurophysiological correlate of the rise of the sense of presence.

Experiment 1

Materials and methods

Participants

Thirty-two participants ($N = 32$, 14 men and 18 women, with a mean age M of 28.7 years and standard deviation SD of ±6.3, within a range of 22 to 42 years) were selected using an adaptation of the generalized listener selection (GLS) procedure described by Zacharov et al. (Mattila and Zacharov, 2001; Bech and Zacharov, 2006). The GLS procedures included six questionnaires, an audiometric test, and two screening tasks about loudness discrimination and localization of the sound source. For more information about GLS procedures, questionnaires, and descriptive statistics, see Supplementary material. Participants had a high education level ($M = 15.5$ years, SD $= \pm$ 2.3 years), had no prior history of neurological or psychiatric disorders, were right-handed as determined by the Italian version of the Edinburgh Handedness Inventory (Oldfield, 1971), had discriminative abilities of both loudness and sound source localization, had normal hearing acuity, and were "un-trained/naive subjects" as described in ITU-T Recommendation P.800 (ITU-R, 1996). All participants provided written informed consent to participate in the studies, which were approved by the local ethical committee "Comitato Etico Area Vasta Emilia Nord" and were conducted in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards (World Medical Association, 2013).

Acoustic apparatus

A silent audiometric cabin (IAC-Acoustics) 2 m high, 2.5 m wide, and 2.1 m deep was set up with a 5.1-channel surround sound reproduction system consisting of five APART MASK4C speakers (impedance 8 Ohms) and one APART SUBA165 sub-woofer (impedance 4 Ohms), all driven by a DENON AVR-X1600H amplifier. The participant was positioned at the center of the silent audiometric cabin, while the six speakers channels ("L" = left, "R" = right, "C" = center, "Ls" = left Surround, "Rs" = right Surround, "LFE" = low-frequency effects or sub-woofer) were positioned and oriented following the ITU-R BS.1116-1 recommendation so as to direct the sound to a central point that identified the reference listening position (ITU-R, 1997). Audio reproduction was room-corrected using the Audyssey software (Paul, 2009). Sound pressure levels (SPL) were recorded with a sound level meter (Gain Express, applied standard IEC651 type 2, type ANSI 2 SI 0.4) placed at the listening position, and the reproduction level was set below the hazardous hearing threshold (85 dB, A-weighted, for eight consecutive hours) defined and standardized by the National Institute for Occupational Health and Safety (NIOSH) in the ONE (Occupational Noise Exposure) recommendation (Murphy and Franks, 2002).

Stimuli

Twenty-seven cinematic excerpts (10 s long) without music and dialogues were chosen through an online validation experiment (see Supplementary material). We selected stimuli that had high dynamism, high emotional intensity, and negative emotional valence because these characteristics can elicit stronger arousing responses in the participants. Previous studies have demonstrated that negative audio-visual stimuli from feature films can increase arousal levels (Fernández-Aguilar et al., 2019). Our 27 stimuli were used in three Presentation modes: Surround, Stereo, and Monophonic for a total of 81 experimental stimuli, repeated twice. Stimuli were reproduced in the silent audiometric cabin by all six channels in the Surround reproduction mode, by "L" and "R" channels in the Stereo reproduction mode, and only by the "C" channel in the Monophonic reproduction mode. For more information about the stimuli selection procedure, see Supplementary material.

Procedure

Participants listened to 27 cinematic excerpts (10 s long) reproduced in three Presentation modes, played randomly twice for a total of 162 trials. The experiment was divided into three blocks of 54 trials each, with a break between blocks for a total experiment duration of ∼45 m. Each trial consisted of a black fixation cross on a gray background (1.5 s), followed by the auditory stimulus presented for 10 s on a black screen. After viewing the stimulus, each time participants had 5 s to respond to two questions, randomly selected from a pool of four questions, on a Visual Analog Scale (VAS) from 0 to 100. The questions were formulated by the authors to measure four potential aspects of the cinematic immersion and sense of presence induced by the sound excerpt: Enjoyment (EN)—“How much did you like the scene?”; Emotional Involvement (EI)—“How much did you feel emotionally involved?”; Physical Immersion (PI)—“How much did you feel physically immersed?”; and Realism (RE)—“How realistic did you judge the scene?” (for more information, see Supplementary material). Before the experiment, we trained participants with six trials, two per each Presentation mode, using stimuli previously excluded through the validation process. A gray background was used as an inter-trial interval (ITI) with a duration of 3.5 s. At the end of the experimental session, the participant was asked to fill out the Film Immersive Experience Questionnaire (F-IEQ) (Rigby et al., 2019). For descriptive statistics, see Supplementary material. Stimuli were presented with MATLAB extension Psychtoolbox-3 (Brainard, 1997).

Analysis

In order to investigate whether Enjoyment (EN), Emotional Involvement (EI), Physical Immersion (PI), and Realism (RE) were modulated by the experimental conditions, a linear mixed-effect analysis was performed. Following a hierarchical approach, we initially created a simple model using one parameter, and we progressively added others with the aim to evaluate whether their inclusion improved model fit. Likelihood ratio tests, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) were used to rigorously choose which parameters improved model fit. We entered participants’ scores as dependent variables, and Questions (EN, EI, PI, and RE, respectively) and Presentation modes (three levels: Surround, Stereo, and Monophonic) as independent fixed variables. Participants were included as a random intercept and Presentation mode as a random slope. This approach accounted for the within-subject and between-subject variability in the data. Outliers were identified and excluded from the analysis based on the standardized model residuals and a threshold value of Cook’s distance (threshold = 1). *Post-hoc* tests were conducted using Tukey’s correction for multiple comparisons and Kenward–Roger degrees-of-freedom approximation method. Statistical analyses were performed using R software (R Core Team, 2022), lme4 (Bates et al., 2015), effects (Fox and Weisberg, 2019), and emmeans (Lenth, 2022) packages. For data plotting, we used the ggplot2 (Wickham, 2016) package.

Results

The model explained 85% of the variance in the dependent variable taking into account the random effects ($R^2_{\mathrm{m}} = 0.22$; $R^2_{\mathrm{c}} = 0.85$). The model revealed a significant main effect of Presentation modes [$\chi^2_{(2)} = 65.16$, $p < 0.001$), showing that participants attributed significantly higher absolute scores when stimuli were presented in the Surround Presentation mode than when they were presented in the Stereo Presentation mode [$t_{(31)} = 7.76, p < 0.001$] or in the Monophonic Presentation modes [$t_{(30.9)} = 7.76, p < 0.001$; Surround: $M = 59.44$, CIs = 54.93, 63.95; Stereo: $M = 51.4$, CIs = 47.35, 55.44; Monophonic: $M = 41.15$, Cis = 36.01, 46.29]. At the same time, participants attributed significantly higher scores when stimuli were presented in the Stereo Presentation mode than when they were presented in the Monophonic Presentation mode [$t_{(31)} = 6.26$, $p < 0.001$].

The model also revealed a significant main effect of Question [$\chi^2_{(3)} = 71.57$, $p < 0.001$] showing that participants attributed

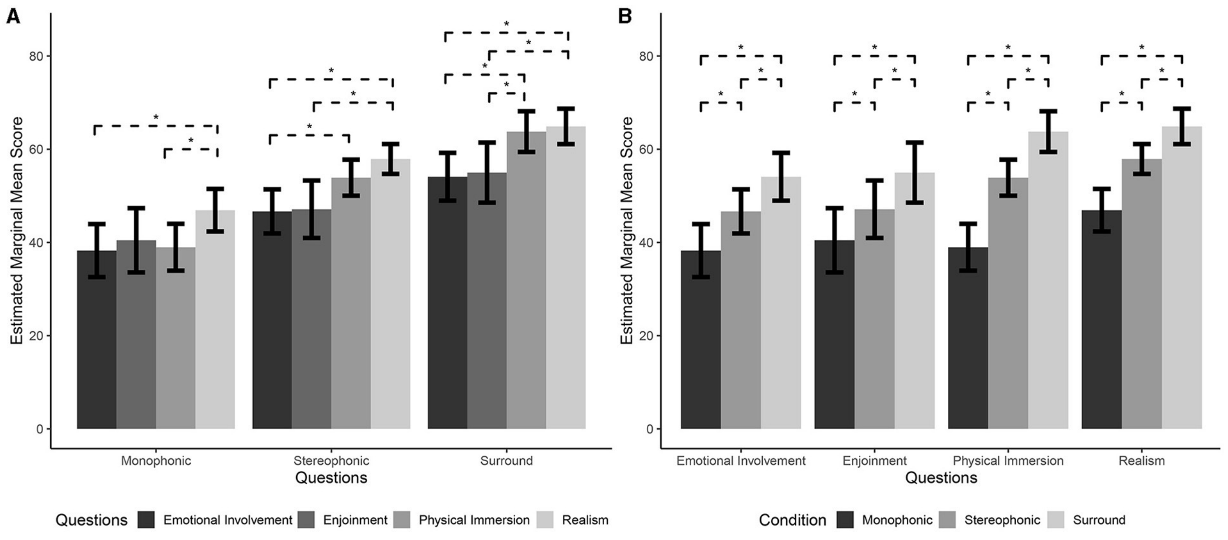


FIGURE 1
(A) Interaction effect *post-hoc* pairwise comparisons by Presentation modes. **(B)** Interaction effect *post-hoc* pairwise comparisons by Question. Error bars represent 95% confidence intervals of the mean (CI); asterisks (*) represents $p < 0.05$.

higher scores on Realism than on Enjoyment [$t_{(31)} = 3.63$, $p < 0.01$], Emotional Involvement [$t_{(31)} = 5.7$, $p < 0.001$], and Physical Immersion [$t_{(31)} = 3.23$, $p < 0.01$; EI: $M = 46.33$, CIs = 41.56, 51.11; EN: $M = 47.53$, CIs = 41.34, 53.71; PI: $M = 52.21$, CIs = 48.29, 56.14; RE: $M = 56.57$, CIs = 53.31, 59.84]. In addition, participants attributed higher scores to Physical Immersion than to Emotional Involvement [$t_{(30.9)} = 5.55$, $p < 0.001$].

Additionally, the model revealed significant Presentation modes*Question interaction [$\chi^2_{(6)} = 269.36$, $p < 0.001$]. Interaction *post-hoc* comparisons showed that in Monophonic Presentation mode (Figure 1A), participants attributed significantly higher scores on Realism than on Emotional Involvement [$t_{(34.6)} = 4.69$, $p < 0.001$] and Physical Immersion [$t_{(38.1)} = 5.59$, $p < 0.001$; Monophonic EI: $M = 38.25$, CIs = 32.57, 43.94; Monophonic EN: $M = 40.47$, CIs = 33.58, 47.36; Monophonic PI: $M = 38.97$, CIs = 33.94, 44.00; Monophonic RE: $M = 46.92$, CIs = 42.36, 51.48]. In the Stereo Presentation mode (Figure 1A), participants attributed significantly higher scores on Realism than on Emotional Involvement [$t_{(33.8)} = 6.13$, $p < 0.001$] and Enjoyment [$t_{(32.6)} = 4.28$, $p < 0.001$; Stereo EI: $M = 46.66$, CIs = 41.93, 51.39; Stereo EN: $M = 47.12$, CIs = 40.9, 53.28; Stereo PI: $M = 53.89$, CIs = 50.02, 57.76; Stereo RE: $M = 57.91$, CIs = 54.71, 61.11]. In addition, participants attributed significantly higher scores to Physical Immersion than to Emotional Involvement [$t_{(39.4)} = 6.43$, $p < 0.001$]. In the Surround Presentation mode (Figure 1A), participants attributed significantly higher scores on Realism than on Emotional Involvement [$t_{(34.4)} = 5.86$, $p < 0.001$] and Enjoyment [$t_{(32.7)} = 3.93$, $p < 0.001$; Surround EI: $M = 54.09$, CIs = 48.96, 59.22; Surround EN: $M = 54.99$, CIs = 48.53, 61.44; Surround PI: $M = 63.78$, CIs = 59.41, 68.15; Surround RE: $M = 64.90$, CIs = 61.10, 68.70]. Furthermore, participants attributed significantly higher scores to Physical Immersion than to Emotional Involvement [$t_{(41.4)} = 8.5$, $p < 0.001$) and Enjoyment [$t_{(33.3)} = 4.02$, $p < 0.001$]. Moreover, in all questions (Figure 1B) participants always attributed significantly higher absolute scores when stimuli were presented in the Surround Presentation mode than when they were presented in the Stereo Presentation mode [EI: $t_{(42.1)} = 6.64$, $p < 0.001$; EN: $t_{(41.9)} = 7.03$, $p < 0.001$; PI: $t_{(42.7)} = 8.81$, $p < 0.001$; RE: $t_{(42.5)} = 6.23$, $p < 0.001$] or in the Monophonic Presentation mode [EI: $t_{(33.18)} = 6.61$, $p < 0.001$; EN: $t_{(32.9)} = 6.07$, $p < 0.001$; PI: $t_{(33.5)} = 10.33$, $p < 0.001$; RE: $t_{(33.3)} = 7.5$, $p < 0.001$]. Moreover, independently from the question, participants attributed significantly higher scores when stimuli were presented in the Stereo Presentation mode than when they were presented in the Monophonic Presentation mode [EI: $t_{(35.2)} = 4.97$, $p < 0.001$; EN: $t_{(35)} = 3.94$, $p < 0.001$; PI: $t_{(35.8)} = 8.8$, $p < 0.001$; RE: $t_{(35.6)} = 6.49$, $p < 0.001$].

Discussion

In this first behavioral experiment, we used a diverse set of naturalistic stimuli consisting of validated cinematic audio excerpts. This approach allowed for a more diverse range of stimuli and more generalizable results (Sonkusare et al., 2019) compared to previous studies (Lipscomb and Kerins, 2004). We investigated how different audio Presentation modes affect the emotional and bodily involvement and audio perception of participants. Results showed that participants consistently gave higher ratings when stimuli were presented in the Surround Presentation mode compared to the Monophonic or Stereo Presentation modes. Specifically, we found that the Surround Presentation mode was particularly effective in eliciting a sense of Realism, Emotional Involvement, and Physical Immersion among participants. These data are in line with the meta-analysis by Cummings and Bailenson, who reported that the spatial presence experience, evoked by the Surround Presentation mode, correlates positively with the level of immersion of the system (Cummings and Bailenson, 2015). We also corroborate, with more robust results and heterogeneous and ecological stimuli, previous results confirming that the sense of presence can be heightened by the spatialized sound Presentation mode (Lessiter and Freeman, 2001; Västfjäll, 2003; Kobayashi et al., 2015).

Experiment 2

Materials and methods

Participants

Twenty-four participants ($N = 24$, 11 men and 13 women, with a mean age M of 24.3 years and standard deviation SD of ±2.4, within a range of 21 to 30 years) were selected using an adaptation of the generalized listener selection (GLS) procedure (Mattila and Zacharov, 2001; Bech and Zacharov, 2006). For questionnaire descriptive statistics, see Supplementary material. Participants had a high education level ($M = 15.2$ years, SD = ± 1.5 years), had no prior history of neurological or psychiatric disorders, were right-handed as determined by the Italian version of the Edinburgh Handedness Inventory (Oldfield, 1971), had discriminative abilities of both loudness and sound source localization, had normal hearing acuity, and were "un-trained/naive subjects" as described in ITU-T Recommendation P.800 (ITU-R, 1996). All participants provided written informed consent to participate in the studies, which were approved by the local ethical committee "Comitato Etico Area Vasta Emilia Nord" and were conducted in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards (World Medical Association, 2013).

Stimuli

We used the same set of 27 cinematic excerpts without music and dialogues used in Experiment 1. A set of 27 control stimuli were also generated by phase-scrambling the original audio excerpts making them unintelligible to the participant but retaining all the acoustic characteristics on the frequency level and the same duration (10 s). Based on the results of the first experiment and for the purpose of simplifying the experimental paradigm, we chose the most and least effective among the three Presentation modes, Surround and Monophonic, for a total of 108 experimental stimuli.

Procedure

Participants listened to 27 cinematic excerpts and 27 control excerpts reproduced in two Presentation modes, played randomly twice for a total of 216 trials. The experiment was divided into four blocks of 54 trials each, with a break between blocks for a total experiment duration of ∼60 m. Each trial consisted of a black fixation cross on a gray background (1.5 s), followed by the auditory stimulus presented for 10 s on a black screen. After viewing the stimulus, participants had 5 s to respond to a question on a Visual Analog Scale (VAS) from 0 to 100: "How much did you feel physically immersed?" (Physical Immersion, PI). We used only one question compared to Experiment 1 in order to reduce the complexity of the task and we chose one of the questions that better characterizes the spatialized sound experience as exposed in Experiment 1. Before the experiment, we trained participants with six trials, three per each Presentation mode, using stimuli previously excluded through the validation process. A gray background was used as an inter-trial interval (ITI) with a duration of 3.5 s. After the experimental task, participants were asked to indicate if they recognized any action-related sound. If they stated that there was an action-related sound present, they were asked to write down what action sound they recognized. This information was used to verify that participants were able, on average, to recognize one action-related sound in each stimulus. At the end of the experimental session, the participant was asked to fill out the Film Immersive Experience Questionnaire (F-IEQ) (Rigby et al., 2019). For questionnaire descriptives, see Supplementary material. Stimuli were presented with MATLAB extension Psychtoolbox-3 (Brainard, 1997).

EEG and EMG recording and pre-processing

The electromyography (EMG) signal was acquired by an AD Instruments PowerLab 35 (ADInstruments, U.K.), and LabChart 8 Pro software was used for recording. EMG activity was bipolarly recorded on the left Extensor Digitorum Communis and left tibialis anterior with 4 mm standard Ag/Ag-Cl electrodes. Before being attached to the muscle regions, the participants' skin was cleaned with an alcohol solution and the electrodes were filled with gel electrode paste (Fridlund and Cacioppo, 1986). EMG was sampled at 2 kHz and recorded with an online Mains Filter (adaptive 50 Hz filter). A band-pass filter (20–500 Hz) was applied, and data were arithmetically rectified (Abs). We calculated the EMG root-mean-square (RMS) response in microvolts (μV) by subtracting the baseline activity (average activity during fixation cross) from the activity during each stimulus divided into 20 segments of 500 ms each. EMG recording was done to exclude that the desynchronization recorded during stimulus presentation was influenced by participants' movements. Hence, outliers (segments with EMG activity ±3 SDs from baseline RMS) were considered movement artifacts, leading to trial exclusion during EEG pre-processing.

EEG data were acquired by a Geodesic Sensor System which includes the Net Amps 300 amplifier and a 128-channel HydroCel Geodesic Sensor Net (HCGSN-128) and recorded at a sampling rate of 500 Hz with the vertex (Cz) as an online reference while sensor-skin impedances were maintained below 50 kΩ for each sensor using Net Station 5.4 EGI software (Electrical Geodesic Inc., Eugene, OR). We applied a high-pass filter (0.5 Hz, transition window of 0.25 Hz) and ZapLine line noise removal (50 Hz Notch) using MATLAB (MathWorks, 2022) toolbox EEGLAB v2022.1 (Delorme and Makeig, 2004). Bad channels were identified with *Clean Rawdata* EEGLAB plug-in (v2.0) using flatline criterion (max 5s), line noise criterion (cutoff SD = 4), and minimum channel correlation (cutoff $r = 0.7$) and were interpolated using the spherical interpolation method. We removed 24 channels that were located at the periphery or the frontal region of the sensor net as they were likely to show residual muscle (13 peripheral channels: Ch48, Ch49, Ch56, Ch63, Ch68, Ch73, Ch81, Ch88, Ch94, Ch99, Ch107, Ch113, and Ch119) and eye artifacts (11 frontal channels: Ch1, Ch8, Ch14, Ch17, Ch21, Ch25, Ch32, Ch125, Ch126, Ch127, and Ch128), reducing the number of channels from 128 to 104. Continuous EEG data were divided into 12 s epochs, which included 2 s of baseline and 10 s of activity during the presentation of the stimulus. Epochs with muscle activity, identified using EMG (see above), were removed. Independent component analysis (ICA) was applied, and an automated recognition algorithm (MARA) was used to identify ocular, cardiac, and muscular artifacts (Winkler et al., 2011, 2014). A mean number of 16.7 (SD = 1.8) independent components (ICs) were removed. Finally, EEG data were re-referenced to the common average (Bertrand et al., 1985).

Analysis

In order to investigate the dynamic changes in spectral power over time and temporal patterns of neural activity related to the perception of audio Presentation modes, we performed a time-frequency analysis using the Hanning taper method. The window length was fixed at 0.5 s, with a frequency resolution of 1 Hz, spanning from 3 to 32 Hz. This allowed for the examination of event-related spectral perturbation (ERSP) in Alpha (8–13 Hz) and Beta (14–32 Hz) frequency bands. Data were baseline corrected by division considering as a baseline the 500 ms window before the stimulus onset. We averaged the Monophonic Control Presentation mode and the Surround Control Presentation mode with the aim to obtain a control condition independent from the perceptual differences due to Presentation modes and considered hereinafter as a Control Presentation mode.

In order to address the multiple comparisons problem (MCP) that arises from the multidimensional EEG data structure and to control for family-wise error rate (FWER), we used a cluster-based non-parametric test for within-subjects experiments implemented in FieldTrip (Oostenveld et al., 2011). The cluster-based test statistics is calculated by comparing experimental conditions at the sample level, selecting samples with t-values above a certain threshold, clustering them based on temporal, spatial, and spectral adjacency, and taking the sum of t-values within each cluster. The significance probability is then calculated using the Monte Carlo permutation method with 500 random draws. A p-value is calculated by comparing the observed test statistic to the distribution of test statistics obtained through random partitions of the data. A cluster is considered significant if its p-value is less than the critical Alpha level of 0.05. This data-driven approach allows one to identify specific time windows and electrode clusters, where there is a significant difference in neural activity between experimental conditions without any spatial cluster and frequency band assumption, and highlight regions of interest for further analysis. From electrodes in the identified clusters, we then extracted the log-ratio frequency power within the significant time window/frequencies range, and a linear mixed effect analysis was performed. Following a hierarchical approach, we initially created a simple model using one parameter, and we progressively added others with the aim to evaluate whether their inclusion improved model fit. Likelihood ratio tests, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) were used to rigorously choose which parameters improved model fit. We entered log-ratio frequency power as the dependent variable and Presentation modes (three levels: Surround, Monophonic, and Control) as independent fixed variables. The participants were included as a random intercept. This approach accounted for the within-subject and between-subject variability in the data. Outliers were identified and excluded from the analysis based on the standardized model residuals and a threshold value of Cook's distance (threshold = 1). *Post-hoc* tests were conducted using Tukey's correction for multiple comparisons and Kenward–Roger degrees-of-freedom approximation method.

For behavioral analysis and its results, see Supplementary material. Statistical analyses were performed using R software (R Core Team, 2022), lme4 (Bates et al., 2015), effects (Fox and Weisberg, 2019), and emmeans (Lenth, 2022) packages. For data plotting, we used the ggplot2 (Wickham, 2016) package.

Results

Alpha band range

A significant cluster in central and parietal areas (see Supplementary Figure S3 for cluster channels) was identified in the time window from 3 to 7 s in the Alpha frequency band (8–10 Hz). This reflects a difference in neural activity in this frequency band and time window between the experimental conditions. Specifically, this cluster is characterized by a significantly higher event-related desynchronization (ERD) during the Surround Presentation mode compared to both the Monophonic (Figure 2A) and Control Presentation modes (Figure 2B), with a peak difference of ∼5 s from the stimulus onset.

The linear mixed model on log-ratio frequency power explained 57% of the variance in a dependent variable taking into account the random effects ($R^2_{\mathrm{m}} = 0.11$; $R^2_{\mathrm{c}} = 0.57$). The model revealed a significant main effect of Presentation modes [$\chi^2_{(2)} = 43.87$, $p < 0.001$], showing that there was a significantly greater ERD in the Surround Presentation mode when compared to both the Monophonic [$t_{(16,\,777)} = 3.79$, $p < 0.001$] and Control Presentation modes [$t_{(16,\,777)} = 6.62$, $p < 0.001$; Surround: $M = 1.5$, CIs = 0.83, 2.17; Monophonic: $M = 1.6$, CIs = 0.93,

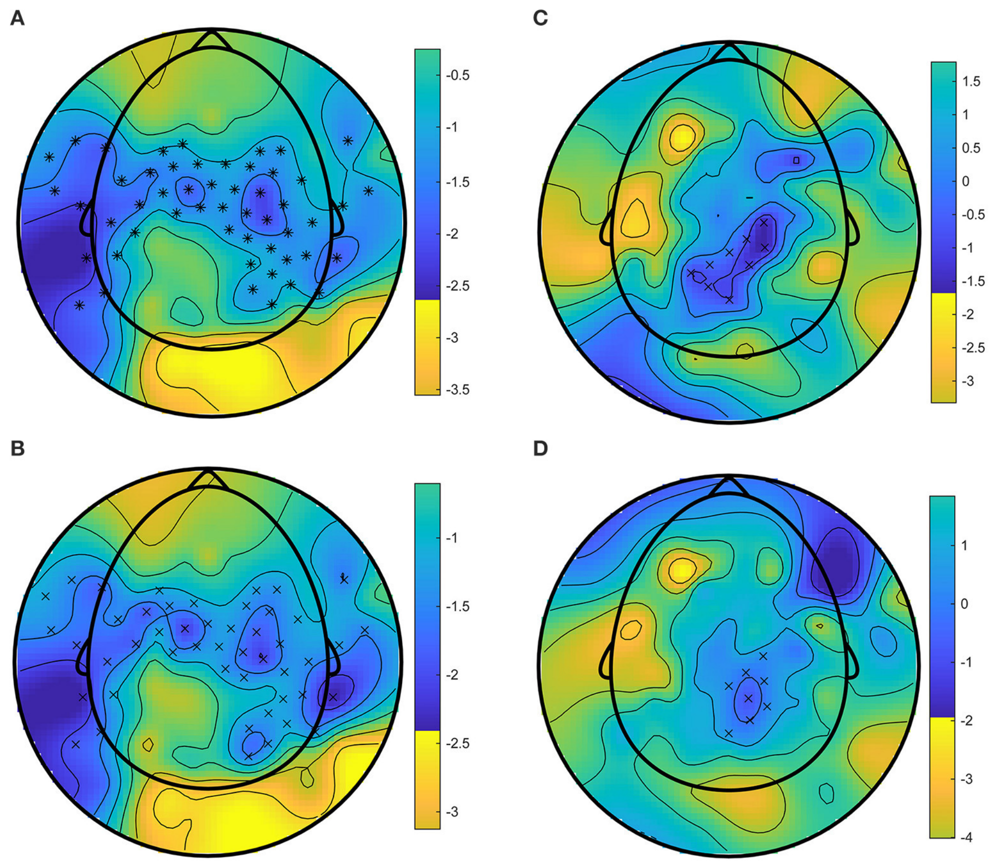


FIGURE 2
(A) Surround—Monophonic Alpha band (8–10 Hz) cluster (peak time 5 s from stimulus onset). **(B)** Surround—Control Alpha band (8–10 Hz) cluster (peak time 5 s from stimulus onset). **(C)** Surround—Monophonic Low-Beta band (16–18 Hz) cluster (peak time 4.5 s from the stimulus onset). **(D)** Surround—Control Low-Beta band (16–18 Hz) cluster (peak time 4.5 s from the stimulus onset). The Crosses (+) symbol indicates channels with $p <$ 0.01 and asterisks (*) indicates channels with $p < 0.001$.

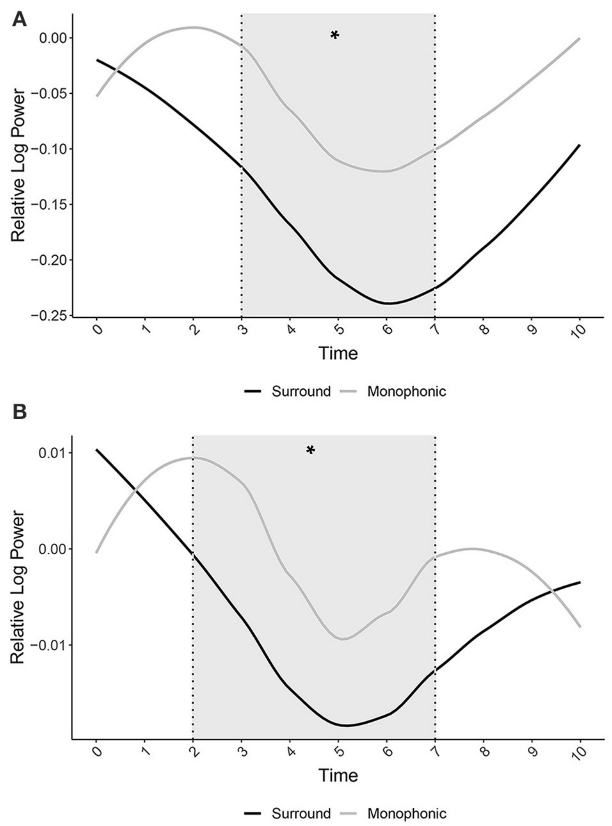


FIGURE 3
(A) Log-ratio frequency power extracted from the Surround Presentation mode and the Monophonic Presentation mode relative to the Control in the Alpha cluster over time. **(B)** Log-ratio frequency power extracted from the Surround Presentation mode and the Monophonic Presentation mode relative to the Control in the Low-Beta cluster over time. Gray area highlights indicate significant differences between Presentation modes, asterisk (*) = $p < 0.05$.

2.28; Control: $M = 1.66$, CIs $= 0.98$, 2.33]. This means that there was an increase in neural activity in the centro-parietal areas during the Surround Presentation mode when compared to the Monophonic and Control Presentation modes, with a peak difference of ∼5 s after the stimulus onset. Furthermore, the *post-hoc* comparisons also showed that there was a significantly greater ERD in the Monophonic Presentation mode when compared to the Control Presentation mode [$t_{(16, 777)} = 2.25$, $p < 0.001$]. In order to better visualize the time course and patterns of ERD, we normalized Surround and Monophonic Presentation mode power by subtracting Control Presentation mode power. We detected a different ERD pattern between the Surround and Monophonic Presentation modes (Figure 3A). Even if a significant difference is detected in the time windows between 3 and 7 s after the stimulus onset, the ERD in Surround starts after the stimulus onset followed by a power rebound 6 s after the stimulus onset, while in Monophonic, we distinguish an ERS in the first 2 s after the stimulus onset and an ERD 2 s after the stimulus onset followed by a power rebound 6 s after the stimulus onset (Figure 3A). The findings demonstrated that the temporal progression of cortical activation differed between the two Presentation modes.

Low Beta band range

The second significant cluster in the central area (see Supplementary Figure S4 for cluster channels) was identified in the time window from 2 to 7 s in the Low-Beta frequency band (16–18 Hz). This reflects a difference in neural activity in this frequency band and time window between the experimental conditions. Similar to the first Alpha cluster, this cluster is also characterized by an ERD during the Surround Presentation mode compared to both the Monophonic (Figure 2C) and Control Presentation modes (Figure 2D), with a peak difference of ∼4.5 s from the stimulus onset.

The linear mixed model on log-ratio frequency power explained 89% of the variance in a dependent variable taking into account the random effects ($R^2_m = 0.2$; $R^2_c = 0.89$). The model revealed a significant main effect of Presentation modes [$\chi^2_{(2)} = 9.79$, $p < 0.001$], showing that there was a significant ERD in the Surround Presentation mode when compared to both Monophonic [$t_{(5017)} = 2.58$, $p < 0.05$] and Control Presentation modes [$t_{(5017)} = 2.94$, $p < 0.001$; Surround: $M = 0.29$, CIs $= 0.17$, 0.40; Monophonic: $M = 0.31$, CIs $= 0.18$, 0.41; Control: $M = 0.30$, CIs $= 0.18$, 0.41]. This means that there was an increase in neural activity in the centro-parietal areas during the Surround Presentation mode when compared to the Monophonic and Control Presentation modes with a peak difference of ∼4.5 s after the stimulus onset. In order to better visualize the time course and patterns of ERD, we normalized Surround and Monophonic Presentation mode power by subtracting Control Presentation mode power. Moreover, in this cluster, we detect a partially different ERD pattern between the Surround and Monophonic Presentation modes (Figure 3B). Even if the significant difference is detected 2 s after the stimulus onset, the ERD in the Surround starts after the stimulus onset followed by a power rebound 6 s after the stimulus onset, while in Monophonic, we distinguish an ERS in the first 2 s after the stimulus onset and an ERD 2 s after the stimulus onset followed by a power rebound 5 s after the stimulus onset.

Discussion

The objective of this HD-EEG experiment was to explore the neural cortical mechanisms and the temporal specificity of sound perception when presented in two distinct acoustic Presentation modes, namely, Monophonic and Surround. The main focus was to compare the neural activity in the Surround Presentation mode to that in the Monophonic and Control Presentation modes, with the hypothesis that the enhanced spatialization of sound in the Surround Presentation mode would lead to greater activation of embodied simulation mechanisms viewed as the physiological index of the sense of presence. Using a data-driven approach that allowed us to identify specific time windows and electrode clusters where there is a significant difference in neural activity between experimental conditions without any spatial cluster and frequency band assumption, we identified two significant centro-parietal clusters: the first in the Alpha frequency band (8 to 10 Hz) and in the time window from 3 to 7 s and the second in the Low-Beta frequency band (16 to 18 Hz) and in the time window from 2 to 7 s. Since the Rolandic Alpha frequency band of interest (8–13 Hz) overlaps with the occipital Alpha band, recordings in central areas might be affected by this posterior activity. However, given that significant clusters were detected only in central and parietal areas, we can exclude that our results were related to attentional/vigilance factors originating from the parieto-occipital cortex. Further analysis revealed a significant ERD in the Surround Presentation mode when compared to both Monophonic and Control Presentation modes both in Alpha and Low-Beta centro-parietal clusters, confirming previous results (Tsuchida et al., 2015) using a more robust analysis approach. We observed a late significant ERD peak (∼4.5/5 s) compared with the typical time course of mu rhythm desynchronization (Avanzini et al., 2012).

General discussion

The results of the present study provide novel insights into the relationship between virtual acoustic spatial environments and the sense of presence providing evidence that Surround Presentation mode enhances the sense of presence by activating embodied simulation mechanisms. In Experiment 1 and consistently with previous research on the relationship between Surround sound and the cinematic experience (Lessiter and Freeman, 2001; Västfjäll, 2003; Pettey et al., 2010; Kobayashi et al., 2015), participants explicitly reported that the Surround Presentation mode significantly enhances the sense of presence, Emotional Involvement, and Physical Immersion, thus showing that the Surround Presentation mode enhances immersion by more closely approximating real-world auditory experience. These findings are consistent with previous research showing that Surround sound formats can envelop the viewer in a 360-degree auditory space unlike the 180-degree space of stereo or mono sound (DiDonato, 2010). Cummings et al., using a meta-analytic approach, investigated the relationship between the immersive quality of a mediated environment and the level of presence experienced by the participant. Several immersive features that offer high-fidelity simulations of reality such as Surround sound had a significant effect on presence (Cummings and Bailenson, 2015). Additionally, these results offer some interesting theoretical implications, supporting the formation of presence as outlined by the spatial situational model framework proposed by Wirth et al. (2007).

The level of similarity between the perceptual experience elicited by video clips and the visual experience during real-life movements is believed to depend on the filming technique. The results of Heimann et al. indeed suggest that there may be a relationship between the perception of approaching stimuli and the feeling of involvement in the scene (Heimann et al., 2014, 2019). This may be due to the presence of more depth cues, which more closely resemble real-life vision. A similar mechanism can be hypothesized for the audio component in cinematic immersion, where the Surround Presentation mode can more closely resemble real-life hearing and activate embodied simulation processes. The EEG results of Experiment 2 further support this hypothesis. The Surround Presentation mode elicited a higher peak of Alpha rhythm desynchronization, reflecting greater activation of the mirror mechanism which represents the neural basis of embodied simulation. The desynchronization peak was delayed from the onset of stimulus presentation likely because of the naturalistic and, to some degree, heterogeneous stimuli used, which lacked a clearly time-locked goal-oriented action onset. This may have influenced the timing of the neural response observed in our study as the sound of the actions that were present in the stimuli had a high and diverse temporal dynamicity. Overall, these findings highlight the importance of considering the nature and characteristics of stimuli used in experiments, particularly when investigating the temporal specificity of neural responses. Furthermore, it is also possible that the use of dynamic and naturalistic stimuli, as opposed to experimental stimuli created *ad hoc*, may have led to a more complex and nuanced neural response. Our findings are consistent with previous studies that have reported different EEG topographies for the Alpha and Beta components of the mu rhythm (Pfurtscheller et al., 1994; McFarland et al., 2000). Previous research revealed different source locations and reactivity for the Alpha and Beta subcomponents of the mu rhythm desynchronization active during action execution and action observation, supporting the idea that they serve distinct functions (Hari and Salmelin, 1997; Pfurtscheller et al., 1997; Hari, 2006; Press et al., 2011; Hobson and Bishop, 2016). The Alpha subcomponent is thought to reflect a sensorimotor function, while the Beta component is more closely linked to motor cortical control. Indeed, further research is needed to fully understand the underlying mechanisms and factors that contribute to this neural response, the different functions of Alpha and Beta subcomponents and how they arise from different neural networks, and the functional significance of the activation of embodied simulation mechanisms in acoustic cinematic immersion. Regardless of these future developments, we can state that immersion, as an objective property of the playback system, a defining characteristic of our stimuli delivery setup, was reflected by the instauration of the sense of presence revealed by a stronger embodiment of spectators. These findings support the idea that cinematic experience is unique and directly connected to sensory-motor patterns that connect the viewer with the screen, allowing for a form of immersive simulation that exploits the full potential of our brain-body system (Freedberg and Gallese, 2007; Gallese and Guerra, 2012, 2019, 2022; Fingerhut and Heimann, 2022). The result is an intersubjective relationship between the viewer and the film that blurs the boundary between the real and imaginary worlds (Gallese and Guerra, 2019).

Conclusion

This study provides new data on how increasing the spatial complexity of virtual environments mediated by cinematic sequences can increase the participant's sense of presence. The greater neural activity recorded in the centro-parietal areas can contribute to the understanding of the neural mechanisms of embodied spatialized auditory perception. In future, by further understanding how the integration between sound and visual information in the cinematic experience occurs, we can gain insight into how the brain processes this information and how it can be used to enhance the viewers' immersive experience. Furthermore, deeper comprehension can also be applied to other areas such as virtual reality and augmented reality, which also rely on the integration of sound and visual information to create immersive experiences. Filmmakers and sound designers may also leverage this knowledge to precisely manipulate audiovisual elements, resulting in a heightened emotional impact and greater engagement with film scenes. The knowledge gained from this exploration should also have broader implications in fields beyond entertainment. Fields such as education and therapy can benefit from a deeper insight into how the brain processes and integrates sound and visual information. These applications can range from designing effective educational multimedia content to developing immersive training for therapy/rehabilitation purposes.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/ntemf/?view_only=a6f1d81e84d940f9a31513b9ed3f8c67.

Ethics statement

The studies involving human participants were reviewed and approved by "Comitato Etico Area Vasta Emilia Nord" and were conducted in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards (World Medical Association, 2013). Participants provided their written informed consent to participate in these studies.

Author contributions

NL and VS conceptualized the idea and discussed it with all the authors. NL edited the stimuli, performed data acquisition, performed analyses, and wrote the manuscript. All authors designed the experiment, interpreted the results, contributed, and approved the manuscript.

Funding

The study was supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006)—A multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022).

Acknowledgments

The authors wish to thank Davide Bonini (AudioB S.a.s.) for his help with acoustic system arrangement and calibration, and Leonardo Fazio and Rosalia Burrafato for their help in data recording.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1222472/full#supplementary-material

References

Avanzini, P., Fabbri-Destro, M., Volta, R. D., Daprati, E., Rizzolatti, G., Cantalupo, G., et al. (2012). The dynamics of sensorimotor cortical oscillations during the observation of hand movements: an EEG study. *PLoS ONE* 7, e37534. doi: 10.1371/journal.pone.0037534

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bech, S., and Zacharov, N. (2006). *Perceptual Audio Evaluation–Theory, Method and Application*. Hoboken, NJ: John Wiley & Sons, Ltd. doi: 10.1002/9780470869253

Bertrand, O., Perrin, F., and Pernier, J. (1985). A theoretical justification of the average reference in topographic evoked potential studies. *Electroencephalogr. Clin. Neurophysiol.* 62, 462–464. doi: 10.1016/0168-5597(85)90058-9

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Caetano, G., Jousmäki, V., and Hari, R. (2007). Actor's and observer's primary motor cortices stabilize similarly after seen or heard motor actions. *Proc. Nat. Acad. Sci.* 104, 9058–9062. doi: 10.1073/pnas.0702453104

Calbi, M., Siri, F., Heimann, K., Barratt, D., Gallese, V., Kolesnikov, A., et al. (2019). How context influences the interpretation of facial expressions: a source localization high-density EEG study on the "Kuleshov effect." *Sci. Rep. Int. Dev. Res. Cent. Can.* 9, 2107. doi: 10.1038/s41598-018-37786-y

Cummings, J. J., and Bailenson, J. N. (2015). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychol.* 19, 272–309. doi: 10.1080/15213269.2015.1015740

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

DiDonato, M. (2010). *La Spazializzazione Acustica Nel Cinema Contemporaneo. Tecnica, Linguaggio, Modelli di Analisi*. Leesburg, VA: Onyx, Ed.

DiPellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180. doi: 10.1007/BF00230027

Elsaesser, T., and Hagener, M. (2015). *Film Theory an Introduction through the Senses*. London: Routledge. doi: 10.4324/9781315740768

Fernández-Aguilar, L., Navarro-Bravo, B., Ricarte, J., Ros, L., and Latorre, J. M. (2019). How effective are films in inducing positive and negative emotional states? A meta-analysis. *PLoS ONE* 14, e0225040. doi: 10.1371/journal.pone.0225040

Fingerhut, J., and Heimann, K. (2022). Enacting moving images: film theory and experimental science within a new cognitive media theory. *Projections* 16, 105–123. doi: 10.3167/proj.2022.160107

Fox, J., and Weisberg, S. (2019). *An R Companion to Applied Regression,* 3rd ed. London: Sage Publications. Available online at: https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html (accessed May 13, 2023).

Freedberg, D., and Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends Cogn. Sci.* 11, 197–203. doi: 10.1016/j.tics.2007.02.003

Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology* 23, 567–589. doi: 10.1111/j.1469-8986.1986.tb00676.x

Gallese, V. (2009). Mirror neurons, embodied simulation, and the neural basis of social identification. *Psychoanal. Dialogues* 19, 519–536. doi: 10.1080/10481880903231910

Gallese, V. (2019). Embodied simulation. its bearing on aesthetic experience and the dialogue between neuroscience and the humanities. *Gestalt Theory* 41, 113–127. doi: 10.2478/gth-2019-0013

Gallese, V., and Guerra, M. (2012). Embodying movies: embodied simulation and film studies. *Cinema J. Philos. Moving Image* 3, 183–210.

Gallese, V., and Guerra, M. (2019). *The Empathic Screen: Cinema and Neuroscience.* Oxford: Oxford University Press. doi: 10.1093/oso/9780198793533.001.0001

Gallese, V., and Guerra, M. (2022). The neuroscience of film. *Projections* 16, 1–10. doi: 10.3167/proj.2022.160101

Gastaut, H. J., and Bert, J. (1954). EEG changes during cinematographic presentation (moving picture activation of the EEG). *Electroencephalogr. Clin. Neurophysiol.* 6, 433–444. doi: 10.1016/0013-4694(54)90058-9

Hari, R. (2006). Action–perception connection and the cortical mu rhythm. *Prog. Brain Res.* 159, 253–260. doi: 10.1016/S0079-6123(06)59017-X

Hari, R., and Salmelin, R. (1997). Human cortical oscillations: a neuromagnetic view through the skull. *Trends Neurosci.* 20, 44–49. doi: 10.1016/S0166-2236(96)10065-5

Heimann, K., Uithol, S., Calbi, M., Umiltà, M. A., Guerra, M., Fingerhut, J., et al. (2019). Embodying the camera: an EEG study on the effect of camera movements on film spectators' sensorimotor cortex activation. *PLoS ONE* 14, e0211026. doi: 10.1371/journal.pone.0211026

Heimann, K., Umiltà, M. A., Guerra, M., and Gallese, V. (2014). Moving mirrors: a high-density EEG study investigating the effect of camera movements on motor cortex activation during action observation. *J. Cogn. Neurosci.* 26, 2087–2101. doi: 10.1162/jocn_a_00602

Hobson, H. M., and Bishop, D. V. M. (2016). Mu suppression – a good measure of the human mirror neuron system? *Cortex* 82, 290–310. doi: 10.1016/j.cortex.2016.03.019

ITU-R (1996). *Recommendation ITU-R P. 800, Methods for Subjective Determination of Transmission Quality*. Geneva: International Telecommunication Union.

ITU-R (1997). *Recommendation BS.1116-1, Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*. Geneva: International Telecommunication Union.

Keysers, C., Kohler, E., Umiltà, M. A., Nanetti, L., Fogassi, L., Gallese, V., et al. (2003). Audiovisual mirror neurons and action recognition. *Exp. Brain Res.* 153, 628–636. doi: 10.1007/s00221-003-1603-5

Kitagawa, N., and Ichihara, S. (2002). Hearing visual motion in depth. *Nature* 416, 172–174. doi: 10.1038/416172a

Kobayashi, M., Ueno, K., and Ise, S. (2015). The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study. *Presence* 24, 163–174. doi: 10.1162/PRES_a_00226

Lenth, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.8.3. Available online at: https://CRAN.R-project.org/package=emmeans (accessed May 13, 2023).

Lessiter, J., and Freeman, J. (2001). "Really hear? The effects of audio quality on presence," in *4th International Workshop on Presence*, 288–324.

Lipscomb, S. D., and Kerins, M. (2004). "An empirical investigation into the effect of presentation mode in the cinematic and music listening experience," in *8th International Conference on Music Perception* and *Cognition*.

MathWorks (2022). *Statistics and Machine Learning Toolbox Documentation, Natick, Massachusetts: The MathWorks Inc*. Available online at: https://www.mathworks.com/help/stats/index.html (accessed May 13, 2023).

Mattila, V.-V., and Zacharov, N. (2001). "Generalized listener selection (GLS) procedure," in *AES 110th CONVENTION.*

McFarland, D. J., Miner, L. A., Vaughan, T. M., and Wolpaw, J. R. (2000). Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topogr.* 12, 177–186. doi: 10.1023/A:1023437823106

Murphy, W. J., and Franks, J. R. (2002). Revisiting the NIOSH criteria for a recommended standard: Occupational noise exposure. *Int. Congress Exposit. Noise Cont. Engg.* 111, 19–21. doi: 10.1121/1.4778162

Muthukumaraswamy, S. D., and Johnson, B. W. (2004a). Changes in rolandic mu rhythm during observation of a precision grip. *Psychophysiology* 41, 152–156. doi: 10.1046/j.1469-8986.2003.00129.x

Muthukumaraswamy, S. D., and Johnson, B. W. (2004b). Primary motor cortex activation during action observation revealed by wavelet analysis of the *EEG. Clin. Neurophysiol.* 115, 1760–1766. doi: 10.1016/j.clinph.2004.03.004

Muthukumaraswamy, S. D., Johnson, B. W., and McNair, N. A. (2004). Mu rhythm modulation during observation of an object-directed grasp. *Brain Res. Cogn. Brain Res.* 19, 195–201. doi: 10.1016/j.cogbrainres.2003.12.001

Neuper, C., Wörtz, M., and Pfurtscheller, G. (2006). ERD/ERS patterns reflecting sensorimotor activation and deactivation. *Prog. Brain Res.* 159, 211–222. doi: 10.1016/S0079-6123(06)59014-4

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 156869. doi: 10.1155/2011/156869

Paul, A. (2009). *Audyssey DSX 10.2. Surround Sound Overview*.

Perry, A., and Bentin, S. (2009). Mirror activity in the human brain while observing hand movements: a comparison between EEG desynchronization in the μ-range and previous fMRI results. *Brain Res.* 1282, 126–132. doi: 10.1016/j.brainres.2009.05.059

Perry, A., Troje, N. F., and Bentin, S. (2010). Exploring motor system contributions to the perception of social information: evidence from EEG activity in the mu/alpha frequency range. *Soc. Neurosci.* 5, 272–284. doi: 10.1080/17470910903395767

Pettey, G., Bracken, C. C., Rubenking, B., Buncher, M., and Gress, E. (2010). Telepresence, soundscapes and technological expectation: putting the observer into the equation. *Virtual Real.* 14, 15–25. doi: 10.1007/s10055-009-0148-8

Pfurtscheller, G., and Neuper, C. (1997). Motor imagery activates primary sensorimotor area in humans. *Neurosci. Lett.* 239, 65–68. doi: 10.1016/S0304-3940(97)00889-6

Pfurtscheller, G., Pregenzer, M., and Neuper, C. (1994). Visualization of sensorimotor areas involved in preparation for hand movement based on classification of μ and central β rhythms in single EEG trials in man. *Neurosci. Lett.* 181, 43–46. doi: 10.1016/0304-3940(94)90556-8

Pfurtscheller, G., Stancák, A., and Edlinger, G. (1997). On the existence of different types of central beta rhythms below 30 Hz. *Electroencephalogr. Clin. Neurophysiol.* 102, 316–325. doi: 10.1016/S0013-4694(96)96612-2

Press, C., Cook, J., Blakemore, S.-J., and Kilner, J. (2011). Dynamic modulation of human motor activity when observing actions. *J. Neurosci.* 31, 2792–2800. doi: 10.1523/JNEUROSCI.1595-10.2011

R Core Team (2022). R: *A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed May 13, 2023).

Rigby, J. M., Brumby, D. P., Gould, S. J. J., and Cox, A. L. (2019). "Development of a questionnaire to measure immersion in video media: the film IEQ," in *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, 35–46. doi: 10.1145/3317697.3323361

Sbravatti, V. (2019). *La cognizione dello Spazio Sonoro Filmico: Un Approccio Neurofilmologico*. Rome: Bulzoni.

Sestito, M., Umiltà, M. A., Paola, G. D., Fortunati, R., Raballo, A., Leuci, E., et al. (2013). Facial reactions in response to dynamic emotional stimuli in different modalities in patients suffering from schizophrenia: a behavioral and EMG study. *Front. Hum. Neurosci.* 7, 368. doi: 10.3389/fnhum.2013.00368

Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence* 6, 603–616. doi: 10.1162/pres.1997.6.6.603

Sonkusare, S., Breakspear, M., and Guo, C. (2019). Naturalistic stimuli in neuroscience: critically acclaimed. *Trends Cogn. Sci.* 23, 699–714. doi: 10.1016/j.tics.2019.05.004

Sterne, J. (2003). *The Audible Past Cultural Origins of Sound Reproduction*. Durham, NC: Duke University Press. doi: 10.1515/9780822384250

Toro, C., Deuschl, G., Thatcher, R., Sato, S., Kufta, C., Hallett, M., et al. (1994). Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG. *Electroencephalogr. Clin. Neurophysiol.* 93, 380–389. doi: 10.1016/0168-5597(94)90126-0

Tsuchida, K., Ueno, K., and Shimada, S. (2015). Motor area activity for action-related and nonaction-related sounds in a three-dimensional sound field reproduction system. *NeuroReport* 26, 291–295. doi: 10.1097/WNR.0000000000000347

Västfjäll, D. (2003). The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. *CyberPsychol. Behav.* 6, 181–188. doi: 10.1089/109493103321640374

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. Available online at: https://ggplot2.tidyverse.org (accessed May 13, 2023).

Winkler, I., Brandl, S., Horn, F., Waldburger, E., Allefeld, C., Tangermann, M., et al. (2014). Robust artifactual independent component classification for BCI practitioners. *J. Neural Eng.* 11, 035013. doi: 10.1088/1741-2560/11/3/035013

Winkler, I., Haufe, S., and Tangermann, M. (2011). Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behav. Brain Funct.* 7, 30–30. doi: 10.1186/1744-9081-7-30

Wirth, W., Hartmann, T., Bocking, S., Vorderer, P., Klimmt, C., Schramm, H., et al. (2003). Constructing *Presence: A Two-Level Model of the Formation of Spatial Presence Experiences*.

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., et al. (2007). A process model of the formation of spatial presence experiences. *Media Psychol.* 9, 493–525. doi: 10.1080/15213260701283079

World Medical Association (2013). World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013.281053

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 19 October 2023
DOI 10.3389/fpsyg.2023.1153968

OPEN ACCESS

EDITED BY
Celia Andreu-Sánchez,
Autonomous University of Barcelona, Spain

REVIEWED BY
Javier Sanz-Aznar,
University of Barcelona, Spain
Corinne Jola,
Abertay University, United Kingdom

*CORRESPONDENCE
Janne Kauttonen
✉ janne.kauttonen@haaga-helia.fi

RECEIVED 30 January 2023
ACCEPTED 04 October 2023
PUBLISHED 19 October 2023

CITATION
Kauttonen J, Paekivi S, Kauramäki J and Tikka P (2023) Unraveling dyadic psycho-physiology of social presence between strangers during an audio drama – a signal-analysis approach.
Front. Psychol. 14:1153968.
doi: 10.3389/fpsyg.2023.1153968

Unraveling dyadic psycho-physiology of social presence between strangers during an audio drama – a signal-analysis approach

Janne Kauttonen[1,2,3]*, Sander Paekivi[4], Jaakko Kauramäki[2,5,6] and Pia Tikka[2,7]

[1]Competences, RDI and Digitalization, Haaga-Helia University of Applied Sciences, Helsinki, Finland, [2]School of Arts, Design and Architecture, Aalto University, Espoo, Finland, [3]Aalto NeuroImaging, Aalto University, Espoo, Finland, [4]Max Planck Institute for the Physics of Complex Systems, Dresden, Germany, [5]Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland, [6]Cognitive Brain Research Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland, [7]Enactive Virtuality Lab, Baltic Film, Media and Arts School (BFM), Centre of Excellence in Media Innovation and Digital Culture (MEDIT), Tallinn University, Tallinn, Estonia

A mere co-presence of an unfamiliar person may modulate an individual's attentive engagement with specific events or situations to a significant degree. To understand better how such social presence affects experiences, we recorded a set of parallel multimodal facial and psychophysiological data with subjects ($N = 36$) who listened to dramatic audio scenes alone or when facing an unfamiliar person. Both a selection of 6 s affective sound clips (IADS-2) followed by a 27 min soundtrack extracted from a Finnish episode film depicted familiar and often intense social situations familiar from the everyday world. Considering the systemic complexity of both the chosen naturalistic stimuli and expected variations in the experimental social situation, we applied a novel combination of signal analysis methods using inter-subject correlation (ISC) analysis, Representational Similarity Analysis (RSA) and Recurrence Quantification Analysis (RQA) followed by gradient boosting classification. We report our findings concerning three facial signals, gaze, eyebrow and smile that can be linked to socially motivated facial movements. We found that ISC values of pairs, whether calculated on true pairs or any two individuals who had a partner, were lower than the group with single individuals. Thus, audio stimuli induced more unique responses in those subjects who were listening to it in the presence of another person, while individual listeners tended to yield a more uniform response as it was driven by dramatized audio stimulus alone. Furthermore, our classifiers models trained using recurrence properties of gaze, eyebrows and smile signals demonstrated distinctive differences in the recurrence dynamics of signals from paired subjects and revealed the impact of individual differences on the latter. We showed that the presence of an unfamiliar co-listener that modifies social dynamics of dyadic listening tasks can be detected reliably from visible facial modalities. By applying our analysis framework to a broader range of psycho-physiological data, together with annotations of the content, and subjective reports of participants, we expected more detailed dyadic dependencies to be revealed. Our work contributes towards modeling and predicting human social behaviors to specific types of audio-visually mediated, virtual, and live social situations.

KEYWORDS

audio narratives, non-verbal dyadic interaction, multimodal data, signal analysis, machine learning, time series analysis, recurrence quantification analysis

1. Introduction

Dramatized narratives, such as movies, audio stories, and literate texts, have a strong tendency to create intersubjectively shared experiences between people who engage with them. Such experiential situations intertwine the human brain and body with its environment in a holistic manner that can be argued to be systemic and complex by nature (Varela et al., 1991; see Tikka et al., 2023, for review). In this study, we applied a novel data analysis approach that takes into account the systemic complexity of the experimental setting, in line with Turkheimer et al. (2022). The context-dependent time-locked synchronization has been demonstrated for neural signals in various neuroimaging studies using functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) when subjects have watched feature films (Hasson et al., 2004; Jääskeläinen et al., 2008; Kauttonen et al., 2015, 2018; Lankinen et al., 2018; Tikka et al., 2018; see Jääskeläinen et al., 2021 for review), recorded dance performance (Jola et al., 2013), or listened to audio narratives (Boldt et al., 2013, 2014; Koskinen and Seppä, 2014; Simony et al., 2016; Koskinen et al., 2020). Synchronization of brain functions across different viewers is more extensive when the stimulus represents socio-emotionally meaningful contexts (e.g., Stephens et al., 2010; Saarimäki, 2021). Yet, in these studies, due to the experimental conditions in neuroimaging labs, the so-called synchronization is to some extent hypothetical, taken that the subjects experience the content alone, and not in real-time interaction with others (e.g., Ames et al., 2014). It can be assumed that the social presence of another person, perhaps even a stranger, affects this synchronization in some way.

To emphasize the presence of another listener, who is unfamiliar to the other participant, we designed an experimental condition where two persons (later, "partners") were listening to the same audio drama while purposefully unnaturally facing one another. In the control condition a person was listening to the audio drama alone. This setting allowed us to collect facial expressions and a set of psycho-physiological signals from a dramatically contextualized dyadic co-presence between two strangers and compare that data with single listeners (see Methods section for details). During a long-duration behavioral experiment, the attention of a person who is alone conducting the task may shift from task-related thoughts to non-task-related thoughts, a cognitive phenomenon referred to as "mind-wandering" (see Smallwood and Schooler, 2015). In social settings such as in the dyadic listening task where one is facing a stranger, such a shift of attention from the tasks to the other person can be assumed to be more frequent than when listening to a story alone, not only due to the mind-wandering phenomena but also due to the awareness of the attention of the other in the shared situation that may "modulate brain regions involved in regulating behavioral motivation" Chib et al. (2018), thus adding socially induces variations to the complex behavioral dynamics, for instance, enhancing the performance ('social facilitation', Zajonc, 1965).

Any socially conditioned situation is characterized by modifications in physiological and motor responses, such as eye-blinks and eye-contact, indicating the shift of attention from the shared object of attention (narrative) to the other person (Nakano, 2015; Chauhan et al., 2017; Shapiro et al., 2017). The co-presence of another may also "influence both the top-down and bottom-up attention-related processes guiding the decision to move the eyes" (Tricoche et al., 2020; for a review, see Stephenson et al., 2021). Shifting one's gaze to the eyes of the other person might indicate a social action, for instance, a confirmation of joint attention, or a search for a confirmation of a joint affective response (Hamilton, 2016). If the other person is a stranger, eye-contact may also have an important role in establishing a non-verbal agreement of mutual co-presence in the shared space (Miura and Noguchi, 2022). The human skill to recognize and make judgments of an unfamiliar person relies to a great extent on unconscious neural processing of dynamic features of the new face (George, 2016). The presence of another person can be expected to change the properties of the physiological signals of an individual in a complex manner (Chib et al., 2018; Miura and Noguchi, 2022; see Hamilton, 2016; Stephenson et al., 2021, for reviews).

In a face-to-face situation even without verbal interaction, following (Hasson and Frith, 2016), the behavior of the subjects is understood as similarly dynamically coupled, as during a dialogue where interacting parties take turns. The term 'synchronization' is applied here in line with Pikovsky et al. (2001) for describing the tendency of connected systems to organize their cognitive oscillations and motor movement together. Importantly, instead of time-locked synchrony of the *same* simultaneous behavior (e.g., imitation), we take the synchrony of signals to describe the presence of some functional relation (coupling) between two *different* states of the coupling systems (Pikovsky et al., 2001). Furthermore, synchrony or coupling may rely on linear, nonlinear, instantaneous, or delayed coupling of two systems.

Dynamical coupling of social signals is considered an elementary part of human-human interaction, such as with synchronized smiles that already babies show (Ruvolo et al., 2015). While computing synchronization mathematically is often straightforward, e.g., using correlation, interpretation of its meaning for physiological signals is far from straightforward. To what extent social synchronization can be assumed to happen between two people, is dependent on the contextual factors and individual personality traits of the co-present persons (Lumsden et al., 2012; Neufeld et al., 2016), detectable in the physiological signals (Stephens et al., 2010; Canento et al., 2011; Soleymani et al., 2012; Kragel and Labar, 2013).

How and when do social interactions tend towards synchrony? While individuals are listening to the same narrative, their personal attitudes or worldviews are to produce some context-related differences. For example, our audio narrative included humorous, tragic, and sexually colored moments. In this case, two persons could interact either through spontaneous alignment in response to the stimulus content (e.g., simultaneously shared humor or embarrassment), in a cognitive synergy in terms of the conducted task of listening (not so much responding to the narrative content but to the joint task), or the situation could elicit a spontaneous coupled response to the other person's reaction and not so much to the narrative ("I smile because she smiles"). These dramatically "lifted" moments in the narrative could produce observable, significant physiological markers in the data (Stuldreher et al., 2020).

To have more insights to such social systems, we studied the effect of dyadic interactions on the synchronization of individuals' time-locked response signals, which can take rather complex forms. Addressing dynamical modifications of contextually situated human interaction from holistic complex systems point of view called for application of a range of computational methodologies, incorporating signal analysis and machine learning (ML) techniques. We investigated how the co-presence of an unfamiliar individual modulates attentive and physiological responses during shared audio narrative experiences. To facilitate this, we employed conventional tools based on linear relationships, such as correlation and normality assuming tests, as well as non-linear tools such as recurrence quantification and classification algorithms. By employing a multi-methodological approach, we offer new insights into the complexities of social synchronization and individual variability in shared contexts.

To initiate the investigation of the dynamical differences between two groups of data and to further demonstrate the suitability of methods employed, we posed two hypotheses. Firstly, we hypothesized that the paired setup would modulate the physiological signals in a way that would be detected in the higher synchrony (e.g., correlation) of signals between paired subjects (e.g., simultaneously smiling) compared to subjects, who listen to the narrative alone. Based on data from the body of neuroscientific literature, most likely, the single listener group could be expected to result with higher intersubject synchrony than the paired group as the latter group would present more behavioral variations due to the unavoidable social interaction in the experimental setting. However, ISC may not allow such high temporal resolution that would allow detecting the possible delayed responses of the individuals of the pairs to the facial ("dialogical") behavior of the partner, thus any temporally close shared facial behaviors (gaze, smile, eye-brow movements) of the pairs could theoretically show higher synchrony. Due to these possibilities, although unlikely, we chose to test the hypothesis that the paired setup would show higher intersubject synchrony than the single listener setup; a negative correlation coefficient would support the opposite version. Secondly, we hypothesized that signals obtained from our paired setup would exhibit more distinctive and dynamically rich variations compared to those from a single-subject setup. In the latter, we expected the signals to be primarily driven by auditory stimuli, rather than the interactive dynamics present in the paired setting. Here richer dynamics of pairs correspond to increased specific types of activity and variability in the signal. Based on these assumptions, we further made a third hypothesis that one can identify signals recorded from paired and individual subjects based on their signal segments recurrent properties.

1.1. Meeting the methodic challenges of signal analysis of dyadic coupling

The interest in identification of physiological signals related to distinct mental phenomena, such as emotions, has generated a range of methodical approaches (e.g., Stephens et al., 2010; Canento et al., 2011; Soleymani et al., 2012; Kragel and Labar, 2013). For addressing the complex nature of the cognitive phenomena under study, we employed ISC and RSA methods with two novel additions. First, a sliding-window based scan type RSA analysis of the audio drama, and secondly a combination of RQA and machine learning, to provide a deeper understanding of the effect of dyadic interaction on the subjects.

The inter-subject correlation analysis (ISC; Hasson et al., 2004; Nastase et al., 2019) has become a standard method particularly in the field of neuroscience as it can be applied even when the generating process of the signal is unknown. ISC assesses the similarity (often referred to as synchrony) of time series from separate subjects exposed to the same time-locked paradigm, typically using Pearson correlation metrics, which is a linear measure of synchrony. ISC can be further leveraged in Representational Similarity Analysis (RSA; Kriegeskorte et al., 2008), a method for comparing groups of pairwise similarity measures. The main limitation of ISC is, however, that it requires two time-locked signals from different individuals and cannot be applied to an individual signal.

To analyze individual signals, we applied Recurrence Quantification Analysis (RQA), formalized by Zbilut and Webber (1992), and expanded on over the years (see, e.g., Marwan et al., 2007, for overview). Recurrence is an important feature of dynamic systems, characterizing repetitions of values or ranges of values, whose frequencies reveal behavioral regimes of stochastic signals, without any assumption of an underlying model. Quantifying recurrent patterns in signals has been found useful in extracting distinguishing and descriptive information from dynamic systems (Webber and Marwan, 2015), including physiological signals (Fusaroli et al., 2014; Guevara et al., 2017; Xu et al., 2020; Stevanovic et al., 2021). This information can be further applied in building features for predictive machine learning analysis.

We demonstrate the effectiveness of RQA as a fundamental research tool into the effect dyadic interaction has on complex recurrent patterns in facial response signals. With predictive power being the benchmark, the impact and order of importance of RQA quantifiers reveals prioritization of complex behaviors for this target variable, which in turn can tell us much about the interactions taking place in dyadic interactions. This approach is (to the author's knowledge) novel in the field and enabled specifically by the interpretability of recurrence quantifiers, and as such makes up a core motivation of this study. Our analyses do not rely on manual annotations, hence they can be applied in data-driven exploration of dyadic physiological data.

2. Materials and methods

2.1. Subjects

All subjects signed an informed consent prior to participation. The study was approved by the local institute ethics committee (Aalto University). Subjects ($N = 36$; 20–47 years, mean 27.5) were all female, and spoke Finnish as their native language. The subjects were divided into two equal-sized groups by the experimenters depending on subject availability, and they either participated in the experiment alone ("subject single," denoted as SS) or together with another participant ("subject paired," denoted as SP) where we employed a hyperscanning paradigm, collecting data simultaneously from both participants. The experiment type was not informed to the subjects prior to their arrival to the experiment site.

2.2. Measurement setup

After informing of the procedures, the subjects were guided to the measurement room. The room, depicted in Figure 1, was 2.5 m by 3.2 m in size, soundproof, well-lit, and included a central table with two chairs on opposite sides. The two high-quality loudspeakers (Genelec 2029A monitors, Genelec Oy, Iisalmi, Finland) were located symmetrically beside the side walls with equal distance to both subjects (~1.2 m from subjects' ears). The central table housed the two Kinect 2.0 sensors (Microsoft, WA, USA), directed to both subjects, but allowing direct eye contact in the paired condition without interference. In the single subject setup, the subject was seated in "Subject 1″ position, opposite to the door. Physiological measures and depth + traditional video streams were collected via Kinect. In addition, our setup allowed collection of EEG, respiration, galvanic skin response and infrared video data. In noting, due to the limitations of the space, and for not complicating the explication of the method, we decided to exclude this data from the current paper. The collected multisource data is reserved for further validation of the proposed method in the future.

The subjects were instructed in the audiotrack to sit comfortably but relatively still, attend to the auditory stimuli, and at the same time, either pay attention (1) their feelings during the "single" condition, (2) to the other subject's feelings in the "paired" condition without talking. Instructions were pre-recorded to minimize experimental variability, and to allow the subjects to familiarize with the setup, and to allow the experimenters to verify smooth collection of input signals in the control room before the actual experiment sounds. These instructions were used to encourage the subjects to have eye contact with the other participant and not acting such that it would disturb data collection (e.g., talk aloud, do major change in position or stance). During single condition, there was a fixation mark on the opposite wall, roughly at the level of the other subject's head during paired condition, and the subjects were instructed to keep their head fixed to that direction.

2.3. Stimuli

The whole setup consisted of a pre-recorded set of trigger sounds, instructions, affective sound localizer (10 clips) and an excerpt of 27-min soundtrack from a Finnish episode film "Kohtaamisia" (dir. Saara Cantell). The audio parts were presented at 16-bit, 44,100 Hz using a dedicated computer with Presentation program (v17.2, Neurobehavioral systems, Albany, CA, USA), sending digital triggers to Nexus-10 devices. Besides digital triggers, we used short trigger sounds both at the start and in the end to allow manual video synchronization. The sound level was fixed to a comfortable listening level, peaking at about 75 dB SPL at the loudest portions, and was kept constant across subjects. After a one-minute instruction period, the experiment continued automatically to the localizer part. The waveform of the whole stimulus is depicted in Supplementary Figure S1 using both linear and decibel scale.

We used a subset of 10 sounds from the IADS-2 affective sound dataset (Bradley and Lang, 2007) as a reference stimulus, which were played before the audio drama. We selected the 10 sounds to have examples of extreme valence and arousal values of the whole dataset continuum. The IADS-2 sound codes were 110, 424, 817, 286, 216, 261, 246, 262, 810, and 277, presented in this order. Content of these sounds were: [Child's laughter], [Car crash], [African music], [Man yelling before sounds of shooting], [Making love], [Child crying], [Heartbeat], [Yawning], [Classical instrumental music], and [Woman

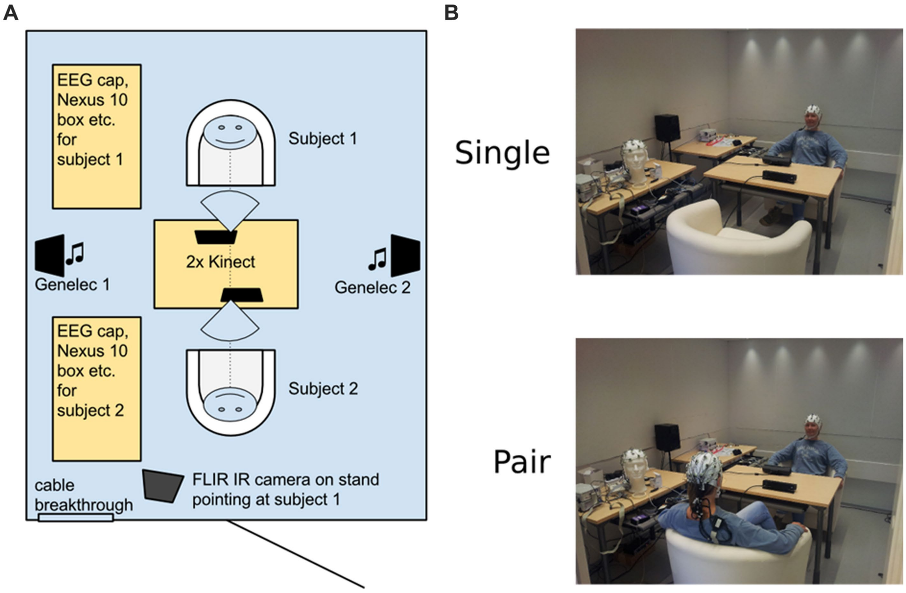


FIGURE 1
The measurement room setup. **(A)** Schematic diagram of the room and placement of periphery. **(B)** Exemplary photos of the actual measurement room in the two different conditions (one of the authors working as a model). The set-up image shows one of the authors.

screaming]. These sounds, originally edited to 6 s, were each looped to a total of 12 s presentation time to allow time for the physiological signal changes. After each presentation of IADS-2 sound, there was a 10 s silent period. The segments were characterized by different socio-emotional content and serve in this paper as an informative introduction to the ISC analysis.

After all the IADS-2 sounds, the experiment continued directly to the soundtrack phase. Notably, the soundtrack was presented in stereo, and on certain occasions the sounds came dominantly from either speaker, possibly promoting joint attention of the subjects.

2.4. Data collection

2.4.1. Video data

We collected both 512 × 424 pixel resolution depth video and 960 × 540 color video at 30 Hz from the one or two Kinect sensors. Further, we included microphone signals to allow exact synchronization of the video onset. The data were initially stored using iPI recorder software (iPi Soft LLC, Russia) in a proprietary video format, and later exported to standard MP4 video. Again, we employed two independent computers for the recording due to bandwidth requirements of the Kinect video stream.

2.4.2. Web ratings

We asked the subjects to rate the soundtrack, at their own pace, using an online web-based tool. The rating was done by intuitive up/down mouse cursor movements while listening to the audio drama. Subjects annotated both valence and arousal-related experiences. The web rating tool records the movements at a 5 Hz sampling rate, allowing us to get a continuous rating of the long soundtrack.

2.4.3. Expert annotations

Audio drama content was annotated by an external expert using ELAN annotation tool (Max Planck Institute[1]).

2.5. Signal processing and feature extraction

2.5.1. Video (RGB, IR)

After the Kinect video collection, we first exported the whole recording soundtrack as plain audio. The soundtrack was used to extract the onset and offset of the video recording based on audio trigger sounds, and this information (video frame numbers) was used to export color/RGB video from the proprietary combined depth + color video for further processing.

2.5.2. Facial muscle activity

Facial activity is typically measured in terms of Facial Action Coding System (FACS) (Ekman and Friesen, 1978). Originally, FACS was developed as a tool for manually quantifying observed facial actions, but it has turned out to be useful in automatic facial expression detection as well. We used the IntraFace software (De la Torre et al., 2015) for detecting the movement of a set of landmarks on the face and performed an approximate FACS coding based on distances between the landmarks. Figure 2 depicts the numbers and locations of FACS, which are listed in Table 1 with the corresponding muscles involved.

Along with the facial feature extraction, IntraFace software provided gaze angle estimation from the video. The video was obtained with a Kinect wide angle lens, so the accuracy was not comparable to for instance an eye-tracking device that processes a more defined video stream, often taken with a dedicated telephoto lens. Due to this, for instance, direct eye contact with the other participant in a paired setup could not be estimated with certainty. Still, as the angle was a continuous measure of interest that seemed robust enough, we chose it for further analysis.

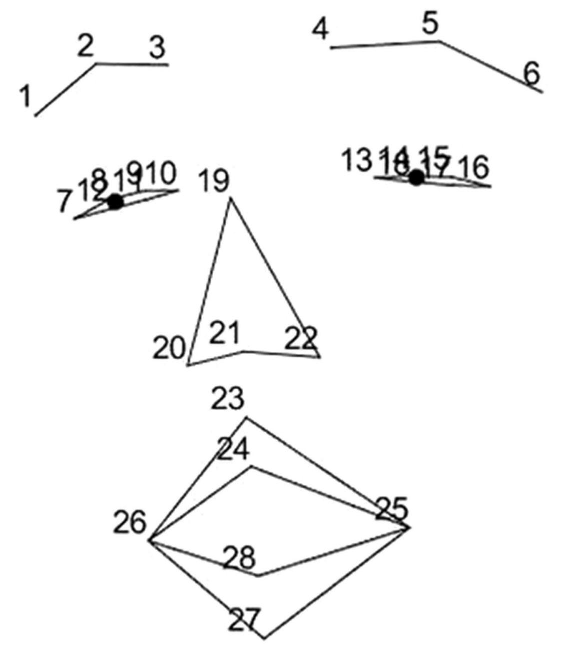


FIGURE 2
Facial landmarks as they were captured from videos.

TABLE 1 Action units in facial analysis.

AU	Name	Muscle	Landmarks
AU 1	Inner Brow Raiser	Frontalis, pars medialis	3, 10, 4, 13
AU 2	Outer Brow Raiser	Frontalis, pars lateralis	1, 7, 6, 16
AU 4	Brow Lowerer	Corrugator supercilii	3, 4
AU 5	Upper Lid Raiser	Levator palpebrae superioris	8, 9, 11, 12, 14, 15, 17, 18
AU 6	Cheek Raiser	Orbicularis oculi	8, 9, 11, 12, 14, 15, 17, 18
AU 9	Nose Wrinkler	Levator labii	10, 13, 20, 22
AU 12	Lip Corner Puller	Zygomaticus major	25, 26
AU 18	Lip Puckerer	Incisivii labii	25, 26
AU 25	Lips part	(several muscles)	24, 28

Landmarks are visualized in Figure 2.

---

[1] https://archive.mpi.nl/tla/elan

2.6. Signal analysis

Our analysis methods of choice were Inter-Subject Correlation (ISC) analysis, Representational Similarity Analysis (RSA) and gradient boosting classification (CatBoost algorithm) combined with Recurrence Quantification Analysis (RQA) as methods to see how individual listeners and those who listened with pairs differed. First, we focus on describing ISC and RSA (2.6.1) aimed to check the hypothesis of differences between the single and paired groups, using a sliding-window based scan type RSA analysis of the audio drama. The second section focuses on RSA and ML (2.6.2) aimed to check the hypothesis of differences between the single and paired groups, using a combination of RQA and ML to provide a deeper understanding of the effect of dyadic interaction on the subjects.

2.6.1. Representational similarity analysis and inter-subject correlations

RSA is an approach to visualize and quantify complex time-series data, irrespective of underlying models, based on a measure of distance describing the similarity between two time-series. We used the ISC as this distance measure, which is based on interpreting the signal as made up of three components: $c(t)$ reflecting processing triggered by the stimulus (the audio track), and should be consistent across subjects; $id_A(t)$ reflects the idiosyncratic response of subject A (respectively for the other subject, B); $\varepsilon_A(t)$ is an error term, reflecting spontaneous activity unrelated to either the stimulus or subject-specific response (Nastase et al., 2019). For a time-series segment, the signal $x_A(t)$ is given then as:

$$x_A(t) = \alpha\, c(t) + \beta\, id_A(t) + \varepsilon_A(t)$$

ISC then assumes that as the second subject B experiences the same, time-locked stimulus (e.g., a movie), their same-source signals will also be a mixture of c, id_B and ε_B. The component $c(t)$ then should be perfectly correlated for subjects A and B, while $id(t)$ and $\varepsilon(t)$ will not (Nastase et al., 2019). The Pearson correlation $r(x_A x_B)$ between the signals of subjects A and B then will increase with α and the average r is a proxy for the latter. Thus, ISC filters out subject-specific information and reveals the joint stimulus-induced component.

Between all individuals, paired or unpaired, we calculated a similarity measure based on the Pearson correlation coefficient, given simply as $d_P = 1 + r$. While this is just a shift of the regular correlation coefficient, it has been used as such in RSA (Kriegeskorte et al., 2008; Nili et al., 2014) previously and we keep the same convention for consistency. The measure as such is called the correlation distance, where the values can be read as $d_P = 0$ indicates strong dissimilarity, $d_P = 2$ indicates strong similarity, with $d_P = 1$ meaning $r = 0$ and thus indicating no negative or positive correlation. This way, an ISC matrix gives a quantifiable and visually interpretable result to determine differences between individual subjects and differences between groups of subjects. We are particularly interested in the difference of ISCs between the group of individual subjects and the group of paired subjects.

In RSA analysis, pair-wise ISC values resemble a time point Representational Dissimilarity Matrix (RDM), where each cell contains the ISC reflecting the respective similarity or dissimilarity between the individuals. From the RDMs, one can gather a visual overview of similarity patterns associated with the individuals, by conditions (specific emotional content, annotation) and modalities (different modes of response, i.e., time-series) (Kriegeskorte et al., 2008; Nili et al., 2014). A further comparison can then be carried out on separate RDMs via for example the Kendall's Tau correlation coefficient, enabling one to identify narrative inputs that produce similar RDMs within or outside of measurement modalities. Kendall's Tau correlation is a non-parametric measure of similarity and it's a recommended method to be applied with RDMs (Nili et al., 2014). By comparing RDMs of different modalities, we could identify stimuli that produce similar, in structure and magnitude, responses in the subjects.

In this work, we performed RSA using specific windows of signals for localizers and applying a sliding window method for audio drama. In the latter, the full-length signals are first analyzed by computing RSA matrices in short (20s) overlapping windows, and then tested against a model matrix with the respective statistical significance filtered by the false discovery rate (Benjamini and Hochberg, 1995) procedure. The former provides a baseline understanding of the intuitive meaning of ISC and RSA analysis, while the latter reveals a general understanding of the relations between audio drama narrative content and ISCs.

Based on data from the body of literature (section #), one may assume that the paired and the single listener groups would show differences in synchrony when compared. However, for arbitrary testing using RSA, we hypothesized detecting an increased correlation of signals in the paired setup when compared against single listeners. That is, in the models that the RDM is compared against, ISC values for the SP group are assumed higher than those for SS or mixed SS-SP subject groups.

2.6.2. Recurrence quantification analysis

Recurrence quantification analysis (RQA) is a non-linear method for characterizing individual time-series (Thiel et al., 2002; Zbilut et al., 2002; Webber et al., 2009), being robust against non-stationarities and noise. The method has been widely applied in many fields (see, e.g., Marwan et al., 2007), and in analyzing dynamics of physiological signals in social interaction (Fusaroli et al., 2014; Guevara et al., 2017; Xu et al., 2020). RQA is based on recurrence plots, which visualize temporal repetitions (recurrences) of values as points on a time–time coordinate system. This visualization is depicted in Figure 3 for our data. That is to say, both the horizontal and vertical axis take values of the time values of a time-series, and thus a repeated sequence appears as a diagonal line. Analysis of the distribution of these points and especially the diagonal lines they form, can yield a variety of interpretable quantities through RQA.

Often the first step of this technique is time delay embedding (Webber and Zbilut 2005), aiming to reconstruct the state space dynamics of higher dimensionality, lost by considering only a scalar time-series, through supplementary delayed coordinates (Takens, 2006; Goswami, 2019). This is achieved by successively delaying the time-series by some value τ, for a number of times D, called the delay, and embedding dimension, respectively. Embedded recurrence plots thus do not consider as a recurrence only a single value repeating, but a sequence, D of them separated from each other by τ. An embedding dimension $D > 1$ may not itself be stationary however (Brick et al., 2017), as the sequence of recurrence might change its character in a non-stationary series – a feature that RQA without embedding ignores.

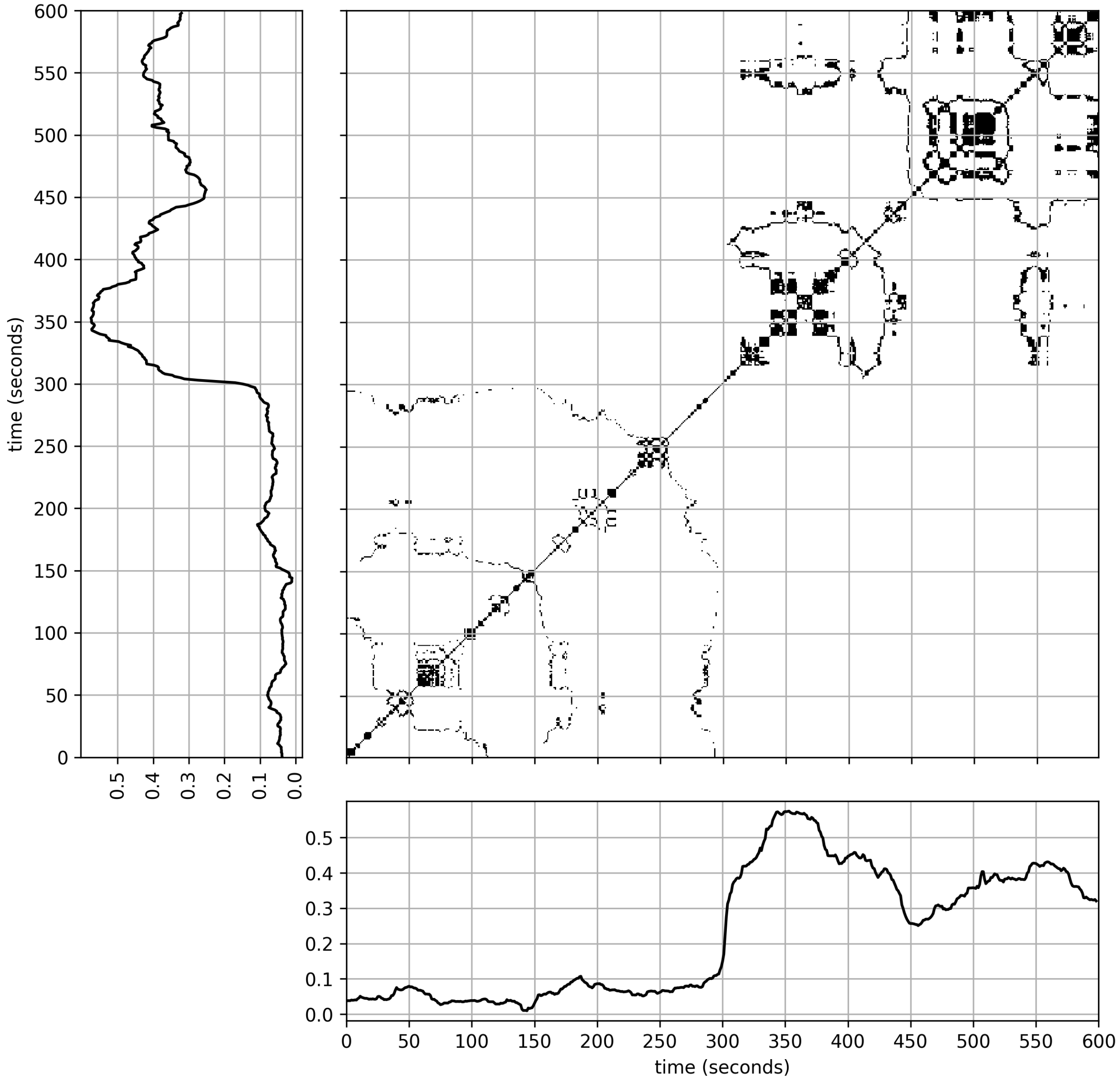


FIGURE 3
A recurrence plot based on the time-series shown on both the vertical and horizontal axis, generated with parameters: recurrence rate = 3%, $D = 1$, and $\tau = 1$. Off-diagonal formations show periods of retained or repeated values as horizontal/vertical and diagonal shapes, respectively. This is not included in recurrence quantifier calculations.

For continuous time-series, what constitutes two values being "the same" is not as straightforward as for discrete series. Here a radius E is considered, by which we consider two values to be recurrent if $\left(x(t_1)-x(t_2)\right)<E$.

Once the parameters have been chosen, the simplest recurrence quantifier obtainable from recurrence plots is the recurrence rate (RR), which is the percentage of points that fulfill the condition $\left(x(t_1)-x(t_2)\right)<E$

$$RR = 100 \times \frac{N_{rec}}{N_p},$$

where N_{rec} is the number of recurrent points and N_p is the number of points on the recurrence plot in total. Based on this ratio, many quantifiers can be defined, of which we employ three:

$$DET = 100 \times \frac{N_{diag}}{N_{rec}} \quad \text{(Determinism)}$$

$$ENTR = -\sum_{l=l_{\min}}^{N_p} p(l)\log_2\left(p(l)\right) \quad \text{(Entropy)}$$

$$LAM = 100 \times \frac{N_{vert}}{N_{rec}} \quad \text{(Laminarity)}$$

The first, determinism (DET) describes the proportion of recurrent points that form diagonal lines, i.e., not just repeating values but a sequence that is repeating. N_{diag} is the number of recurrent points forming diagonal lines. It's worth pointing out that the minimal number of points required to consider a diagonal can vary, but mostly is chosen as the minimal possible, i.e., 2. The second, entropy (ENTR) describes the complexity of the deterministic structure in recurrent points. As $p(l)$ is the probability that a diagonal line has length l, its entropy quantifies the complexity of a time series - higher values indicate a wider distribution of line lengths and translates to the deterministic segments being more varied in their duration. How deterministic a system is, is another thing, as DET only considers the number of points that form diagonal lines, irrespective of their lengths. Lastly, laminarity (LAM) is analogous to DET, but counts the percentage of recurrent points comprising vertical line structures, N_{vert} being the number of points forming such lines. Thus, LAM describes a rate of stagnancy, periods of no change (within the radius threshold) in the time-series. In fact, for an auto-recurrence plot, meaning recurrence analysis on one time-series with itself, vertical lines are the same as horizontal lines, simply at different perspectives.

To calculate the quantifiers, RQA parameters must be chosen, and for this purpose, many methods have been used in the past. The embedding dimension D can potentially be estimated by the method of false nearest neighbors (Kennel et al., 1992), however this also needs some parameter fitting. The delay τ can potentially be estimated as the first minimum of either the auto-correlation or mutual information functions, but not all time-series will provide a reasonable value through such methods. There is also evidence that it is not always important to set these parameters, especially if reconstructing the nonlinear state space is not of interest (Iwanski and Bradley, 1998; March et al., 2005). This can be due to the RQA outputs remaining independent with respect to D. Furthermore, if their numerical value does not have to be the "true" value, meaning if only comparisons are of interest, RQA without embedding ($D = 1$) performs just as well on experimental series (Iwanski and Bradley, 1998). Such an approach has yielded satisfactory results in previous research (Piskun and Piskun, 2011; Arunvinthan et al., 2020; Soloviev et al., 2020) and was found suitable for our data as well.

Lastly, the threshold radius E, defining recurrences can be chosen through empirical or statistical considerations. Larger radii yield more recurrent points, possibly cluttering the recurrence plot, making interpretation difficult and the obtained results uninformative. Among multiple suggestions, a popular one is choosing E that yields a low (1–5%) recurrence rate while the latter scales with E linearly in the log–log sense (Webber and Zbilut 2005; Wallot and Leonardi, 2018). Another suggestion for facilitating comparisons between diverse signals is instead fixing the recurrence rate, letting E vary accordingly (Curtin et al., 2017; van den Hoorn et al., 2020).

We utilized RQA as a measure of temporal dynamics inside windows of signals, to be used in a machine-learning classifier model. For the RQA parameters, we choose $D = \tau = 1$ in all recurrence quantification analyses. For E, in order to compare groups we employed an algorithm to find a suitable E conditioned on fixing the recurrence rate at 3%, as this yielded good model accuracy for our data.

The main reason to employ RQA derives from its malleability by these parameters. Indeed, the recurrence plots and values for recurrence quantifiers change when different parameters are used, but to what extent depends a lot on the system under study. The main important issue here is that the results are interpretable and comparable. For example, when working with N sets of a single measurement dimension, say a single individuals smile patterns on different occasions, then a fixed recurrence radius makes sense, as we would assume the persons underlying characteristics to be consistent, i.e., the radius of recurrence reflects their personality or related properties. This is not comparable however to other people, whose baseline for repetition of a smile can be different and while within their own dataset, a single radius is suitable, the radius is not the same as for the first individual. That is why a fixed recurrence rate is fixed instead, enabling everyone to fit with their own radius, while retaining a comparable set of quantifiers based on the percentage of repetitions. As for the value of 3%, the general best practice is to set this value as low as possible while retaining a vivid recurrence plot. For a stricter rule of thumb, the scaling between quantifier values and the recurrence rate should be linear. Thus, increasing the rate to 4% would increase or decrease the entropy, determinism, etc. linearly. In our dataset 3% was well in a linear regime of scaling and also a small value, enabling recurrence plots to be descriptive and thus the quantifiers as well. There are unfortunately no more rigorous methods to set the value of the recurrence rate, which is common for model-free tools of analysis, and a certain level of expertise must be employed.

In terms of the dimension and delay of embedding being set to 1, meaning no embedding is used, this is due to us only being interested in comparisons of recurrence patterns, not a "true" value of any of the quantifiers. Embedding is used to reconstruct the system's true phase space and is necessary if the study of the latter is of interest, however, the lack of embedding does not take away from a uniform comparison in a lower, non-reconstructed phase space. For further clarification, we refer the reader to the references cited above, at the respective parameter descriptions.

We carried out RQA in a non-overlapping windowed framework for the whole audio drama, with a window length of 20s, yielding 600 data points. With RQA we address the second hypothesis about the data and phenomenon at hand – that the SP group shows richer, more unique dynamics than the SS group. With RQA we could capture the dynamic and non-linear patterns of behavior for individual signals via determinism, entropy and laminarity measures. Then, utilizing these quantifiers as features for a classification algorithm, the ML model can better identify in what manner and how much the quantifiers differ between the groups helps us answer the proposed question in quite large detail. This has the double purpose of also providing an estimate of predictive power for RQA of facial features alone, in the context of dyadic interaction. We tested this hypothesis by training a binary classifier to separate SS and SP signals.

2.6.3. Gradient boosting classification

We employed ML to evaluate information content of physiological signals related to the presence of another person, utilizing the CatBoost gradient boosting classifier (Prokhorenkova et al., 2018). The aim of this analysis was to see if dyadic interaction leaves an imprint on an individual's response signals, identifiable without the need to compare to the responses of others, as is the case for ISC. While there can be many predictors derived from these signals, we employed RQA, which yields single-valued quantifiers for features of determinism, entropy and laminarity, that are capable of capturing nonlinear properties of the underlying time-series. Furthermore, RQA quantifiers are interpretable and simple in their design, making them useful in not only classification but giving them a descriptive dimension.

CatBoost relies on building an ensemble of models in sequence using the whole data, where each model reduces the previous model's error. While there are many different gradient boosting algorithms applicable for similar goals, CatBoost features symmetric decision tree generation, resulting in faster computations and less risk of overfitting (Dorogush et al., 2018), and sees lots of application in scientific inquiry (Hancock and Khoshgoftaar, 2020). The algorithm also readily provides an importance score to all prediction features, which indicates how much on average the prediction accuracy changes, if the feature value changes. We normalized this metric between 0 and 100. It gives an intuitive overview of the impact any given predictor has on the model's performance, and we used it to interpret the latter to a degree, which we use in the context of recurrence quantifiers. We employed the open-source Python implementation via the package 'catboost' available at https://github.com/catboost. The label for the classifier was group belonging (i.e., SP and SS groups) and the predictor variables were RQA quantifiers (determinism, entropy, and laminarity) of data segments. With this analysis, we could test our third hypothesis related to the identification of the group identity of a subject on the basis of signal properties.

3. Results

In the following sections, we report our results regarding three data sources: Gaze direction, and two facial action units (AU2, AU12) which relate to facial signals. These three signals can be linked to socially motivated facial movements, here, listening to an audio drama with a pair against listening to it alone. Next, we discuss our reasoning for selecting these three signals. Our analysis codes can be found in the following repository: https://github.com/SanderPaekivi/ISC_RQA_Toolkit.

3.1. Initial analyses and selection of eyebrow, smile, and gaze responses

To identify facial behavior/expressions of individual participants, the Initial analysis was carried out on FACS signals from all subjects: AU1, AU2, AU4, AU5, AU6, AU9, AU12, AU18, AU25, Gaze Angle. However, in this paper, for the purpose of clarity regarding the applied methods, we specifically focused on reporting the results related to three facial modalities, AU2, AU12 and the Gaze Angle data. This choice for these three facial signals was motivated by considering overlap and performance between signals that are gathered from the same facial area. As an example, in FACS the AU2 signal informs about the movement of the Outer Brow, the *Frontalis, pars lateralis* muscle, that may be interpreted as an indication of a positively valenced, yet to a great extent automated, uncontrollable facial muscle movement. AU1 overlaps to an extent with modalities of AU1 and AU4 (Inner Brow Raiser and Brow Lowerer respectively). Considering their performance with the above-described analysis methods then, the descriptor that most clearly differentiated between SS and SP groups was chosen.

Regarding AU5 through AU9, they were not found informative with ISC and RQA in differentiating between paired and single subjects. Similarly, as AU12, AU18 and AU25 all relate to the motion of the mouth, AU12 was found to best track the dynamical facial expressions that can be related to the initiation of positive social contact. From henceforth we refer to the chosen signals, AU2, AU12 and Gaze Angle, as the "eyebrow-," "smile-" and "gaze" responses, respectively. Examples of the signals are depicted in Figure 4, demonstrating notable burstiness and zero-segments for AUs compared to the gaze signal. Note that the signals are given as normalized values from the facial tracking apparatus, reflecting the facial features movement in one arbitrary dimension. The means and standard deviations of these three signals are listed in Supplementary Tables S1, S2 for the audio clips and long narrative segments, respectively.

3.2. Representational similarity analysis for pairwise coupling

We calculated the ISC matrices for the 10 initial localizing segments each lasting 12 s. To capture lingering effects, we performed the analysis on 20s long non-overlapping windows, as this size proved

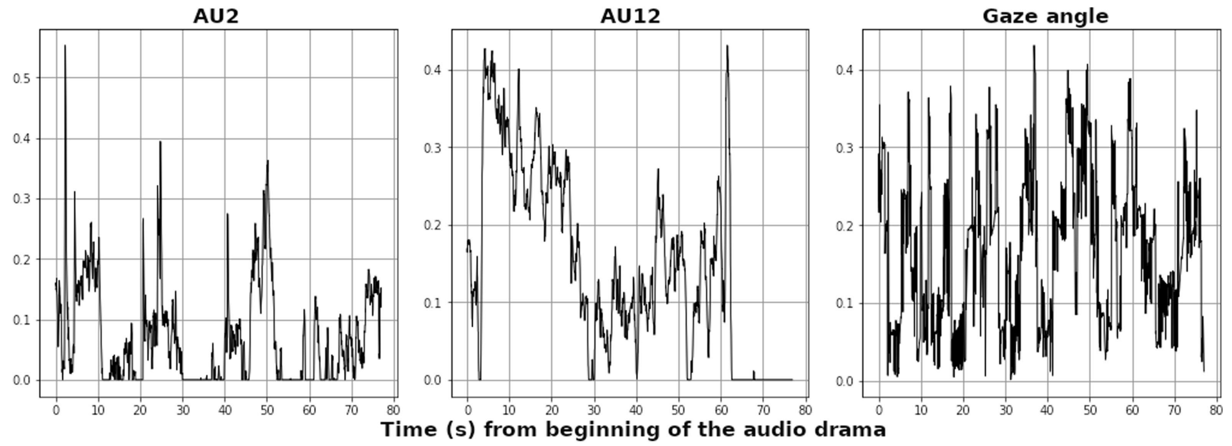


FIGURE 4
Examples of the main analyzed modalities, AU2 ("eyebrow"), AU12 ("smile") and gaze angle from a representative individual (sp07-1), taken for the first 80 s from the onset of the audio drama. All signals are in arbitrary units given by the face tracking apparatus, and scaled between 0 and 1, reflecting the magnitude of the participants' motion.

informative. ISCs were calculated for all pairwise combinations and ordered in the RDM matrices seen in Figures 5–7.

Figure 5 depicts the RDM matrices for the smile response (the Zygomaticus major muscle, AU12 in FACS). Row and column indexes are presented over-one for clarity: starting at individual listener *SS*01-1 (label hidden), followed by *SS*02-1, followed by *SS*03-1 (label hidden) etc. Paired listeners (SP) follow a similar notation: *SP*01-1 (label hidden) and their partner is *SP*01-2. The structure of matrices is illustrated in the last subplot with mixed pairs located in upper-left and lower-right blocks of the matrix.

When observing the visualizations of these matrices, the stimuli [Child crying], [African music], and to some extent [Making love] all exhibit clear separation of ISCs between *SS* individuals (lower left corner) and SP individuals (upper right corner). These stimuli elicit different reactions in the participants in the dyadic condition compared to the individual listeners. Specifically, we see the upper right corner of the ISC matrix shaded darker, representing lower similarity. The distinct difference between the SS and SP groups shows a decrease in synchrony as defined by correlation. This can be interpreted so that the localizing sounds induce more unique responses in subjects who have a partner, while individual listeners yield a more uniform response. These unique responses, regardless of the true pair, can be assumed to emerge due to a dialogical yet silent facial interaction elicited by the drama stimulus. Such social interaction can be detected as non-correlating response signals. This is not the same as the second hypothesis, however, as correlating signals can be more or less complex (the latter is tested via RQA). The ISCs between paired and single listeners showed on average a medial response, indicative of a lack of correlation, as could be expected due to the different listening conditions.

The RDM matrices were calculated also for the eyebrow and gaze responses (AU2 and Gaze Angle) presented in Figures 6, 7 respectively. For these modalities, it is harder to visually identify strong differences in reaction between the studied groups to any stimuli in particular.

In these latter two cases, the difference of ISCs between SS and SP individuals seems to be in their homogeneity. Namely, the values of ISCs between SP individuals range from the lower to the upper bound (see top right quarter of panels), while ISCs between SS individuals show less deviation from a general mean as do ISCs between individual and paired subjects. Next, we use RSA to quantify above visual observations.

3.3. RSA with coupling models

To quantify the similarities of ISCs for different stimuli, we performed Kendall Tau rank correlation analysis on the RDM matrices discussed above. Furthermore, for a description of the

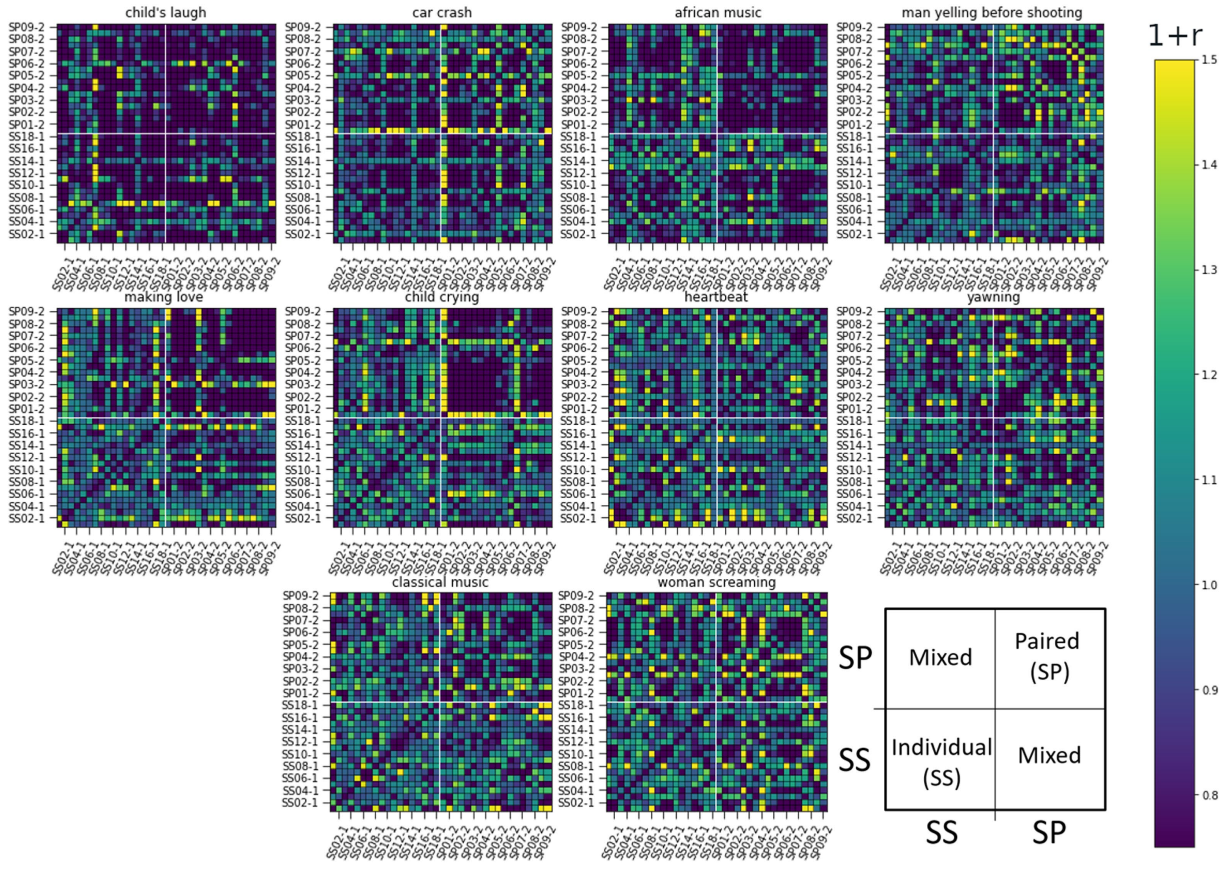


FIGURE 5
RDM matrices of the smile response (the Zygomaticus major muscle, AU12) for all 10 sound-clip stimuli. The figure presents pair-wise ISC values between all subjects. Row and column indexes are presented over one for clarity: Individual listeners ss01-1 to ss17-1 and paired listeners sp01-1 to sp17-1. Darker color corresponds to lower similarity. The arrangement of subjects and related ISC values are illustrated in the last subplot.

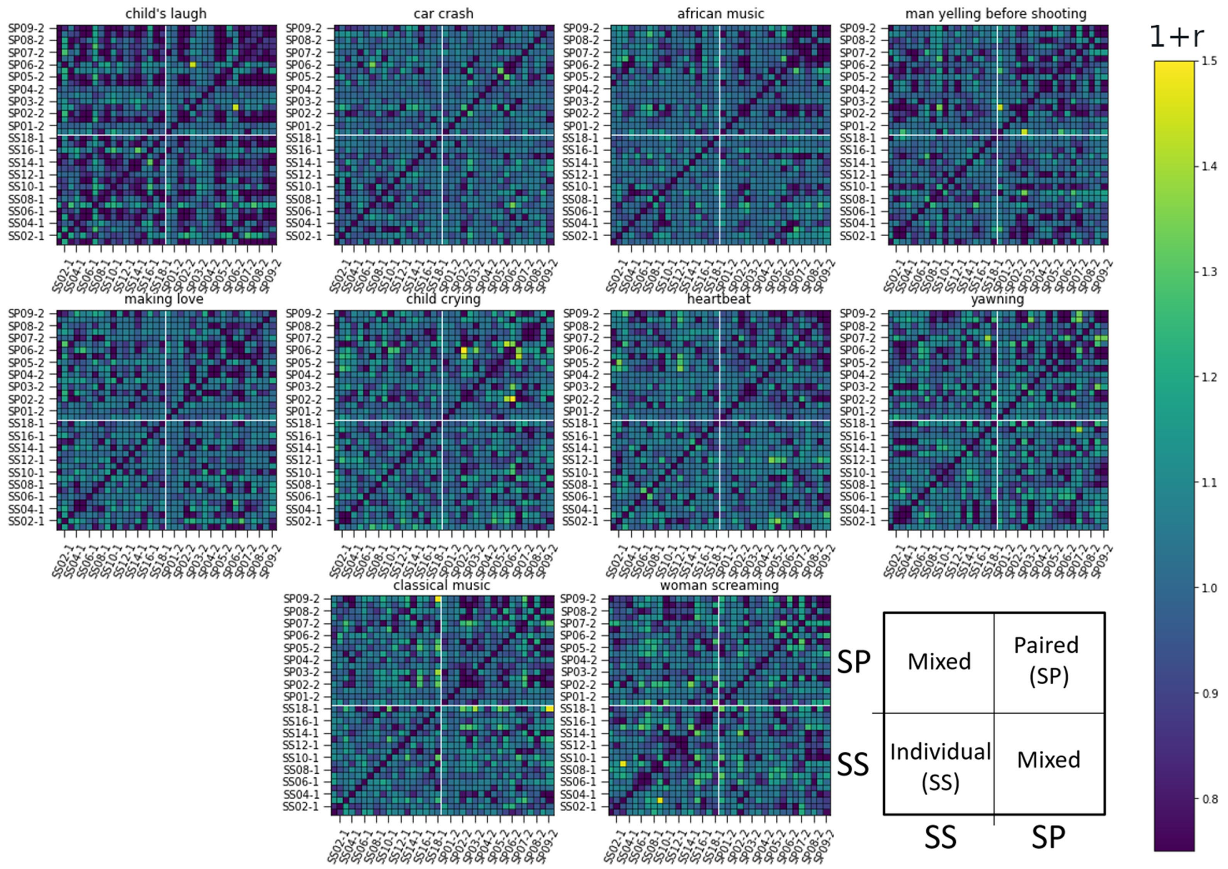


FIGURE 6

RDM matrices of the eyebrow response (the Frontalis, pars lateralis muscle, AU2) for all 10 sound-clip stimuli. The figure presents pair-wise ISC values between all subjects. Row and column indexes are presented over one for clarity: Individual listeners ss01-1 to ss17-1 and paired listeners sp01-1 to sp17-1. Darker color corresponds to lower similarity. The arrangement of subjects and related ISC values are illustrated in the last subplot.

differences between the SS and SP groups specifically, we included in this analysis two model matrices, Model-1 and Model-2 based on our first hypothesis related to synchrony differences.

3.3.1. Model-1 ("the strong coupling model")

Assume higher similarity for only true paired listeners. This means that for example, SP01-1 and SP01-2 are assumed to have an ISC value of 2 (i.e., perfect correlation), while SP01-1 and SP02-2 have a value 1 (i.e., zero correlation). This model essentially assumes that dyadic interaction presents itself in the response signals only when the true dyadic couples are compared.

3.3.2. Model-2 ("the weak coupling model")

Assume higher similarity for all paired listeners, regardless of true couples, i.e., all SP group members' ISC value is assumed value 2. This model essentially assumes that dyadic interaction induces a shared component in response signals regardless of the specific partner (true or not).

Above two models are illustrated in Figure 8 as RDMs.

We computed Kendall Tau rank correlations between all 10 stimuli and 2 model matrices, i.e., total $(12-1) \times 12/2 = 66$ values (i.e., half-diagonal of symmetric matrices). The resulting Kendall Tau rank correlation matrices of the participants are presented below in Figures 9–11 for the smile-, eyebrow- and gaze responses, respectively. Correlations coefficients surpassing statistical significance $p < 0.05$ are shown.

For the smile and eyebrow responses, 6 out of 10 localizer stimuli correlated statistically significantly and relatively strongly with the Model 2, however with a negative coefficient. Firstly, this means that differences in ISC values are higher within SP subjects (true pairs or not) versus SS or mixed SS-SP pairs. That is, *regardless of specific partners in the* SP *group*, ISCs calculated between any individual from the SP group will on the whole yield different results than those calculated between individuals from the SS group. Secondly, the direction of our original assumption was reversed and in fact, individuals in the SP group respond on average *less similarly* to the stimuli than individuals from the SS group.

We can identify other connections as well. For example, in Figure 9 for the smile response we notice a relatively strong correlation between [Making love] and [Car crash] on one hand, and on the other, between [African music] and [Child crying]. These correlations represent a relationship of the pair-by-pair ISC values, and as such a positive correlation between [Making love] and [Car crash] for example tells us that their ISC matrices are significantly similar (something tentatively visible by inspection of the ISC matrices themselves). In this case, there is a significant overlap in the smile response to the mentioned stimuli, likely due to their emotionally provocative content.

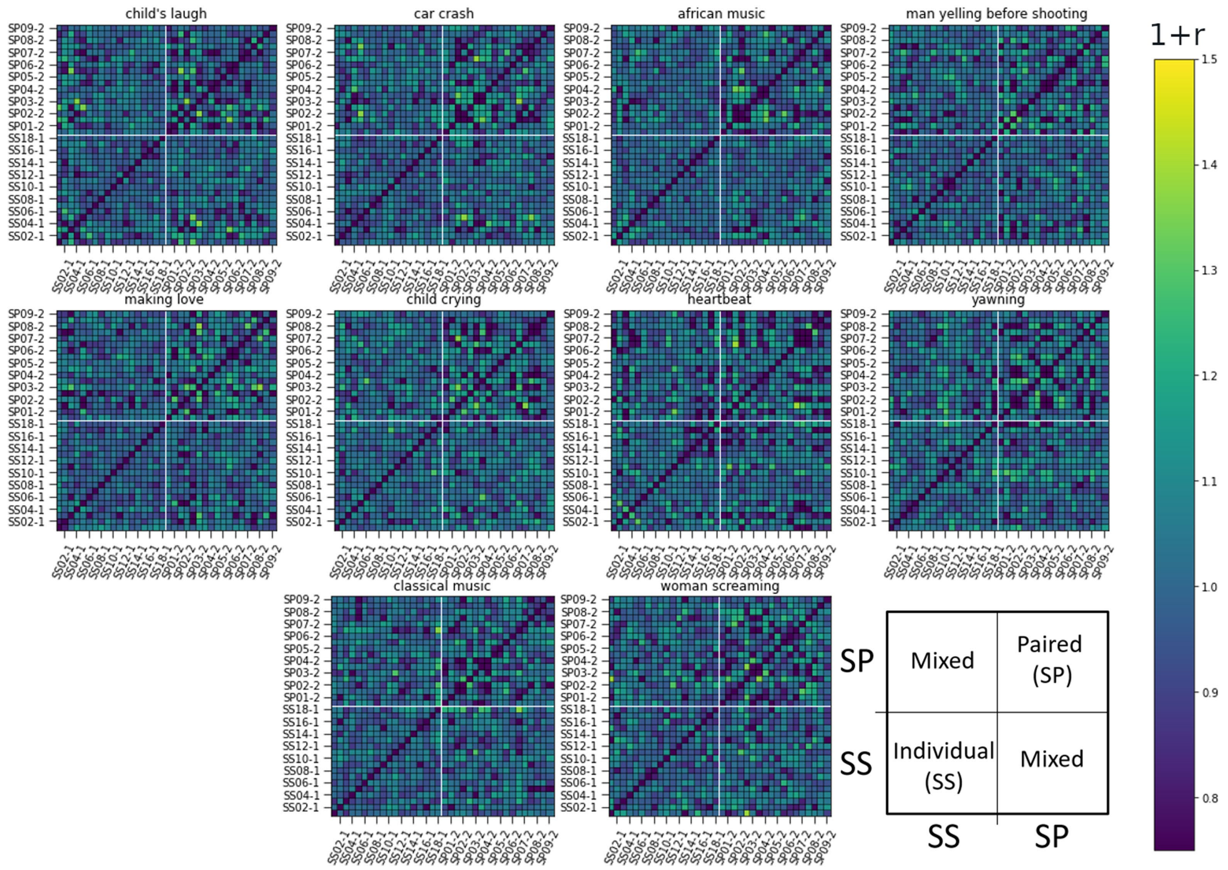


FIGURE 7
RDM matrices of the gaze response (Gaze Angle in FACS) for all 10 sound-clip stimuli. The figure presents pair-wise ISC values between all subjects. Row and column indexes are presented over one for clarity: Individual listeners ss01-1 to ss17-1 and paired listeners sp01-1 to sp17-1. Darker color corresponds to lower similarity. The arrangement of subjects and related ISC values are illustrated in the last subplot.

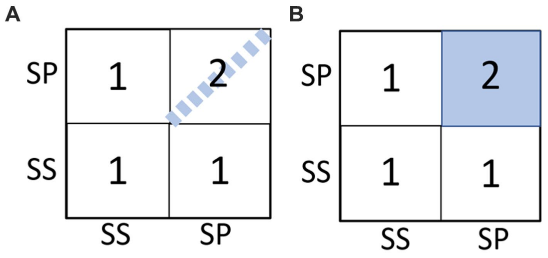


FIGURE 8
Illustration of the two coupling models. **(A)** Model-1 aka the strong coupling model (true pairs in SP) and **(B)** Model-2 aka the weak coupling model (all pairs in SP).

The matrix in Figure 10 for the eyebrow response also shows significant correlations, however for different stimuli. We note a significant overlap in the eyebrow movement to stimuli such as [Man yelling before shooting] (shortened to [Yelling-Shooting] in the label), [Woman screaming], and [Making love].

Lastly, we present the matrix for the gaze response in Figure 11. Regarding the gaze, very few stimuli follow either Model 1 or 2, or are similar to each other. To an extent, this can arise from a stringent confidence level, but more likely is that this particular modality is not easily studied through the ISC framework. Next, we apply ISC and RSA for the signals recorded during the long narrative.

3.4. Sliding-window analysis of the temporally unfolding narrative

The previous analysis demonstrated the capability of ISCs to be used in analyzing dyadic interactions, so in this section we extend it to the whole audio drama, listened to by SS and SP individuals. We carried out a sliding window ISC analysis across the audio drama (27 min), seeking specific intervals where statistically significant ($p < 0.05$) Kendall Tau correlation with Model 2 (the weak coupling model) was present. The window size was selected similarly to the localizing segment lengths including their buffer times, 20 s, with a step of 1 s to capture the relevant highly influential segments. Upon generating the relevant statistics, a false discovery rate (FDR; Benjamini and Hochberg, 1995) adjustment was performed on the series of p-values. In order to avoid spurious results, we considered an interval to show a significant correlation with the model 2 only, if multiple windows in a row were significant.

All intervals with statistically significant correlations are presented in Figure 12 with color-coded and overlaid on the mean valence and arousal graphs of the audio drama, based on manual annotations of

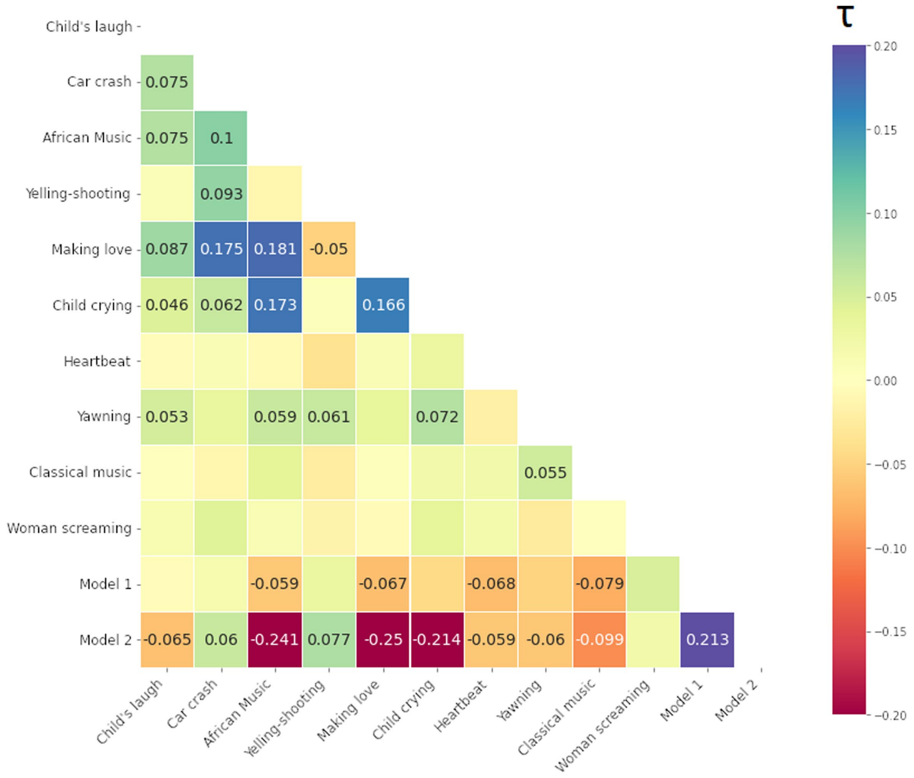


FIGURE 9
Kendall Tau correlation coefficients of RDMs for the smile response (the Zygomaticus major muscle, AU12 in FACS). Correlations between all 10 audio-clip stimuli and two hypothesis coupling models (1 being the strong coupling model). Blue color indicates more similarity, red more dissimilarity (marked with − minus). Only coefficients with statistical significance at $p < 0.05$ are shown.

the participants. Note that because the participants annotated valence and arousal in separate listening sessions on their own, and this portion of the experiment was done voluntarily after the laboratory sessions, the N was different ($N = 11$ for valence; $N = 14$ for arousal). Here the annotated valence/arousal graph was used to describe temporally unfolding emotional content of the film, instead of applying annotations with more complicated labeling.

In Figure 12, intervals that differ notably based on modality; both seem to often precede or align with large changes in the mean valence. Also, all the significant segments of Kendall tau between the RDM and Model 2 yielded *negative* coefficients, as was visible in the RDM comparisons. Gaze Angle again showed little correspondence with Model 2, yielding only one statistically significant interval. Although this is a coarse-grained overview, averaging out a variety of subject-specific responses, the results indicate usefulness of using gaze, smile and eyebrow signals with RSA to pinpoint narrative moments associated with strong dyadic coupling. The analysis suggests that drama moments revealing socially embarrassing information or improper character behavior elicit the strongest dyadic facial interaction (see Section 4.1.2. for details).

3.5. Identifying dyadic interaction by machine learning on recurrence quantifiers

Lastly, we present results for the classification between SP and *SS* groups, i.e., the presence of dyadic interaction or lack thereof. The dataset was generated from the eyebrow-, smile- and gaze responses over the audio drama, split into non-overlapping 20-s segments. Such a setup yielded, for each of the 36 subjects in each modality, $3 \times 3 = 9$ features with total 81 segments corresponding to a total $36 \times 81 = 2{,}916$ samples for CatBoost.

The data was split into a training and a testing set in two ways. First, by randomly splitting the data with an 80% (training) to 20% (testing) ratio yielding 2,332 and 584 samples, respectively. This process was repeated 40,000 times, in order to gauge the impact of individual subject differences on the results. The optimal model parameters (L2 regularization and tree depth) were identified during training with the built-in randomized search function, and estimated in the same framework, with the optimal parameters being chosen as the average over all iterations. The second method of data splitting was based on subjects, having the training and testing set composed of different participants. The training and testing data split of 80 and 20% was maintained here as closely as possible while keeping different subjects within the two sets.

In both training-test split types (i.e., random and subject-wise) the model performances were compared against a baseline (aka dummy) classifier, which was a naive predictor that always predicts the most frequent label found in the training set. The resulting distributions of accuracy percentages for both the model and a dummy classifier are depicted in Figure 13. Mean accuracies of 67% (subject-wise split) and 75% (random split) were above the chance level of 54%, indicating that models learned successfully. In other words, even short segments of our recorded physiological signals carry information on the presence

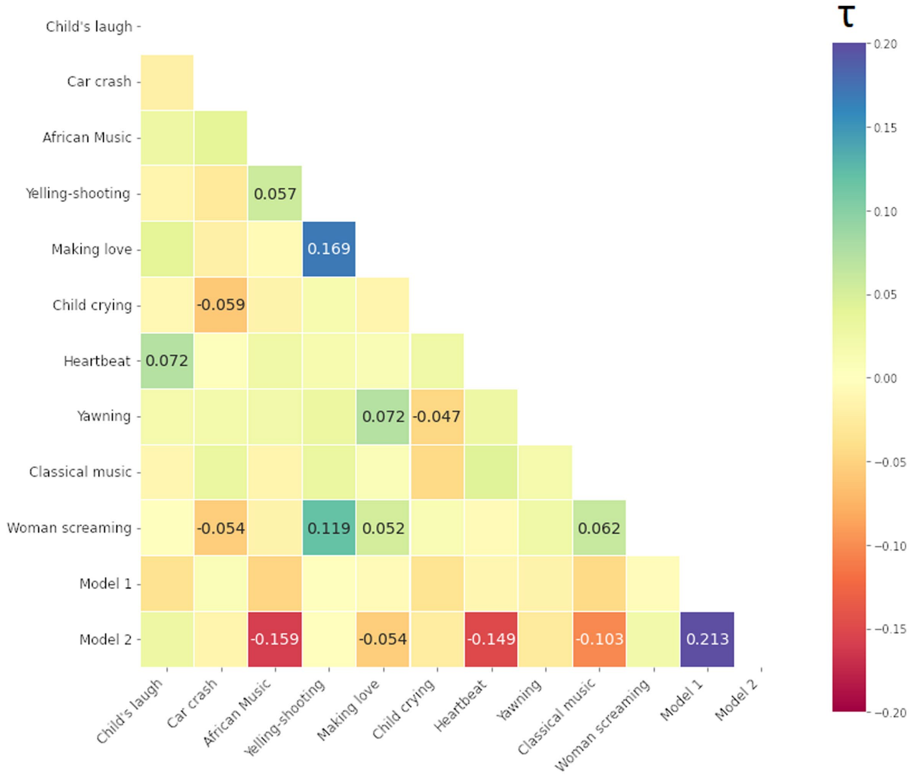


FIGURE 10
Kendall Tau correlation coefficients of RDMs for the eyebrow response (the Frontalis, pars lateralis muscle, AU2 in FACS). Correlations between all 10 audio-clip stimuli and two hypothesis models (1 being the strong coupling model). Blue color indicates more similarity, red more dissimilarity (marked with − minus). Only coefficients with statistical significance at $p < 0.05$ are shown.

of dyadic interactions, and we can classify if a specific subject belongs to the SP and SS group with better than chance probability using features obtained using RQA. This result is in line with our third hypothesis.

In Figure 13B, we notice that the subject-wise split method can generate accuracies not only overlapping with the dummy classifier, but also significantly worse than the latter. This indicates that the training and test set for those splits behave in a contrary manner to each other. The widespread, poor predictions likely relate to overfitting on the training subjects and, considering their individual differences, the test set simply does not follow the same patterns. At the same time, the subject-wise splitting procedure can also yield testing sets that perform better than any combination of the random split method, most likely containing individuals responding closer to average facial behaviors, which the model could learn from the training set individuals, irrespective of outliers.

We further investigated the relative importance of the three features, determinism, entropy and laminarity, for the model. CatBoost feature importances indicate how much the prediction changes on average, if the feature value changes. Feature importances for the random sample-wise split and for subject-wise splits are presented in Figure 14, separately for each data type. For the three modalities, gaze was the most important with mean importance scores between 10 and 30, while smile and eyebrow are equally important at means around 6–10. Therefore, in general, all RQA features were considered relevant for the classifier (i.e., none close to zero).

When considering the random-split procedure, the following can be noticed: While gaze response carries the most importance in classifying dyadic interaction, it is based mainly on laminarity, which essentially measures stagnancy in the time-series, whereas for the eyebrow- and smile responses, entropy was the most important. As entropy is based on the recurrent points that already form diagonal lines (the basis for calculating determinism), i.e., a sequence of values being repeated, it tells us about the distribution of these sequence lengths. This in turn describes the variability in repeated sequences, as a small entropy means repeated sequences were mainly of a similar length. A high entropy smile signal for example includes many repeated smiles that vary in their motion. This can be interpreted as the complexity in the recurrent sequences being associated with dyadic interaction, while eye movement itself or lack thereof is indicative of another person being present.

There was notable variability in importance particularly for the subject-wise splitting. For the gaze response, laminarity is on average even more prioritized over other quantifiers for the subject-wise split, and laminarity also takes priority (albeit not as greatly) for the smile response. This hints at how a model that does not learn to value the entropy of the individual's signal will make worse predictions on the whole. More information on how specifically this is the case can be found in the Supplementary Figure S2 for subject-wise splitting. Interestingly, eyebrow response retains its ordering of importance for both splitting methods, the subject-wise variant increasing the variability notably. This was also the case for further divisions by

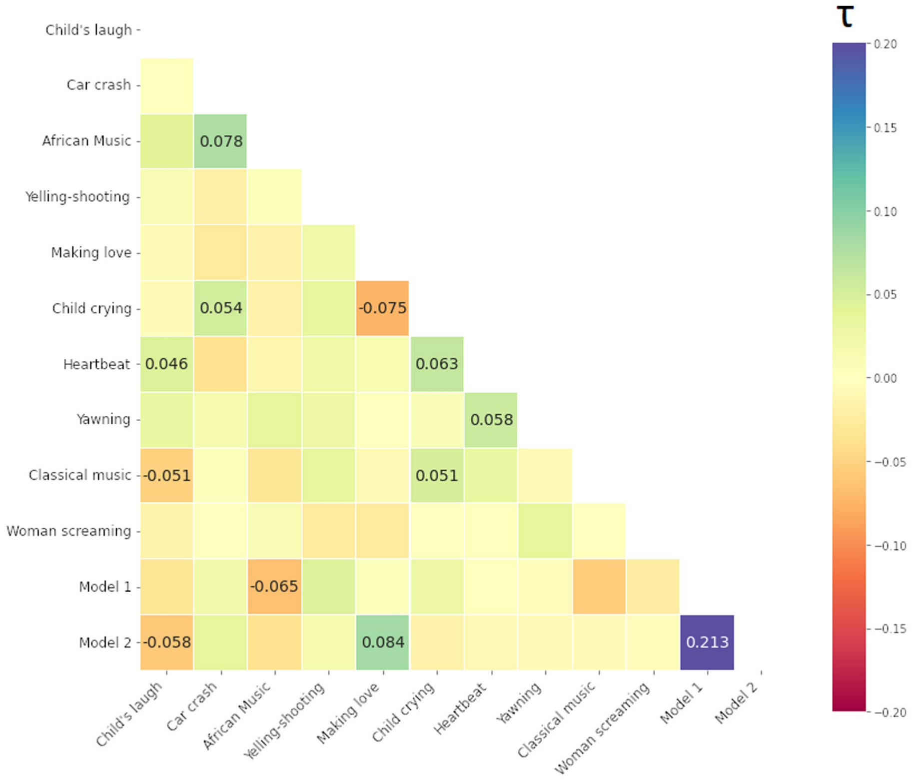


FIGURE 11
Kendall Tau correlation coefficients of RDM matrices for the gaze response (Gaze Angle). Correlations between all 10 stimuli and two hypothesis models (1/2 being the strong/weak coupling model). Blue color indicates more similarity, red more dissimilarity (marked with - minus). Only coefficients with statistical significance at $p < 0.05$ are shown.

accuracy, indicating an interesting robustness of this particular signal, perhaps being more uniformly complex even for individuals whose other signals can vary widely and lead to spurious classifications. Overall, the importance scores highlight the impact of complexity in the smile and eyebrow responses for predicting dyadic interactions, in line with our second hypothesis. A detailed discussion can be found in Supplementary material.

4. Discussion

We report a methodic pathway for analysis of the multiplicity of psycho-physiological signals in individual and paired settings in the context of dynamically unfolding audio stimulus.

We first analyzed signals obtained for 10 sound clips from the affective sound database IADS-2 (Bradley and Lang, 2007) that is widely used in psychological studies (Yang et al., 2018), thus providing us additional support for interpretation. We found that gaze direction estimated from the video and two facial action units, AU2 (eyebrow) and AU12 (smile), provided the most robust information on pairwise dynamics. These three signals could be linked to socially motivated facial movements during listening to an audio drama. They also allowed relatively straight-forward visual validation of data analysis in terms of their social functions: Consider, for instance, movement of the gaze towards or away from the other, or simultaneous smile and/or eyebrow movements. The audio drama part allowed validating the proposed method with longer durational and dynamically more varying physiological signals. Next, we discuss our key results and their interpretations in more detail.

4.1. Social presence affects signal synchrony

How the social presence is shown in the analyzed data is discussed here, first, in terms of the short audio clips, and, secondly, reflected on the 27-min audio drama.

4.1.1. Dyadic behaviors detected in emotional sounds

In our intersubject correlation (ISC) analysis of short localizing audio clips, the single listener groups (SS) showed uniform ISCs, representing the audio stimuli-induced component as theorized. Interestingly, we found that ISC values between any participants in the paired group (SP) were, in general, lower compared to those in the SS group, or mixed SS-SP pairs (subjects in different groups). In other words, regardless of specific partners in the SP group, ISCs calculated between any individual from the SP group yielded different results than those calculated between individuals from the SS group. This could be interpreted so that dyadic interactions do not induce a specific time-locked response but rather a communication with different dynamics of turn-taking patterns between the pairs. The latter could be assumed to consist of temporally delayed socially determined responses between the members of specific pairs, not

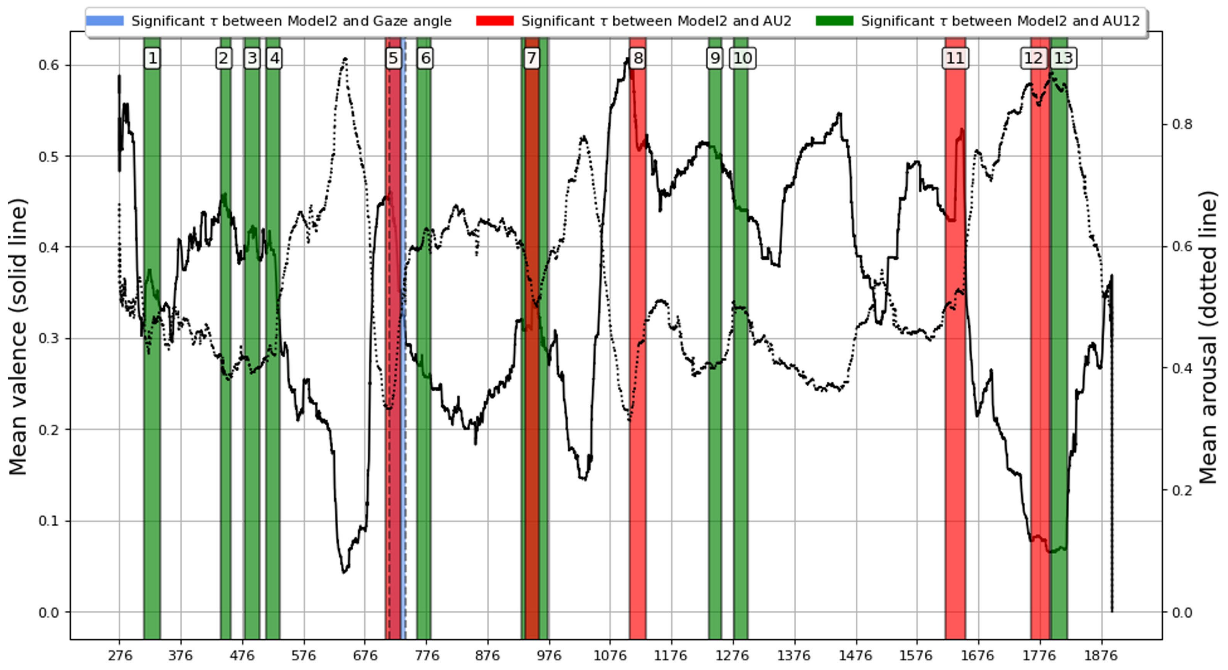


FIGURE 12
Sliding window ISC analysis across the audio drama (27 min, starting at 275 s). Lines correspond to mean valence (solid) and arousal (dotted) annotated by subjects. Colored intervals pinpoint moment with statistically significant ($p < 0.05$, FDR adjusted) Kendall Tau correlation with the weak coupling model (model 2) that corresponds to the assumption of higher similarity between paired (SP) subjects in comparison to single (SS) subjects. Model 2 assumes that dyadic interaction induces a shared component in response signals regardless of the specific partner.

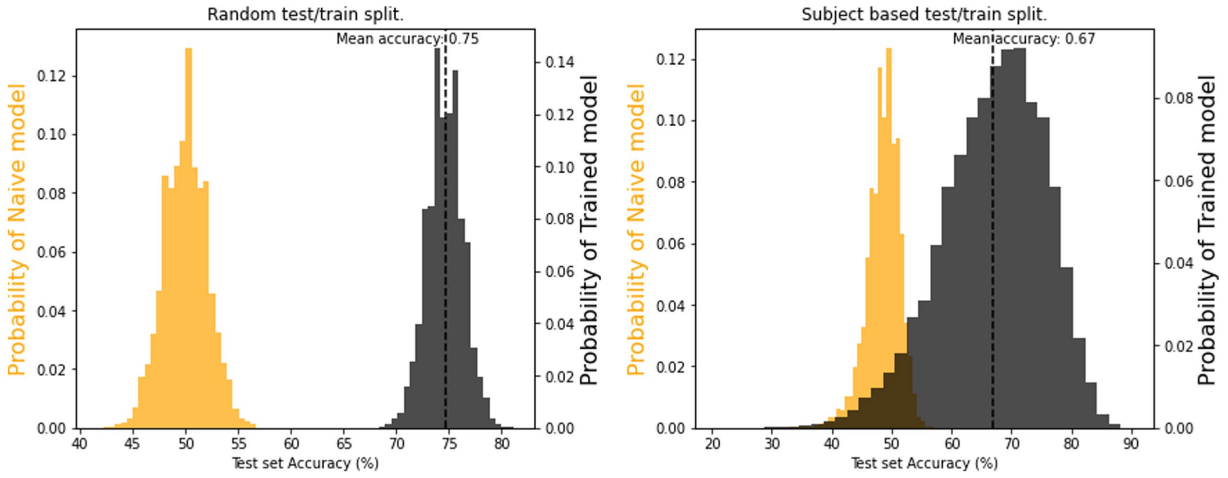


FIGURE 13
Distributions for SP/SS classification model accuracies for the test set using **(A)** random splitting and **(B)** subject-wise splitting validation strategy. The two distributions are the results for a constant-only (dummy) model (orange color) and a catboost classifier model (black color) computed with 40,000 cross-validation iterations.

imitation, thus yielding lower time-locked ISC. Lower ISC values for pairs that for the SS group could be due to communicative turn taking (see, e.g., turn taking in question reply setting, Bögels, 2020). In this case, one's response to the stimuli (e.g., a spontaneous smile or a glance toward the partner) is met with a unique reaction from the other person, in turn provoking potentially further nonverbal interchange based on subtle facial expressions, i.e., dyadic reaction emotion (Sham et al., 2022).

The lower ISC values of SP than SS groups was especially evident for the modality of "smile" (AU12). This may indicate that single subjects showed a tendency to smile even alone when hearing, for example, a child laughing. This positively valenced behavior has been shown to emerge almost automatically (Stark et al., 2020), reflecting innate "caregiving instinct" in adults (Young et al., 2016). For smile (AU12), the strongest distinction between the groups when comparing the ISC results was found in the stimuli [African

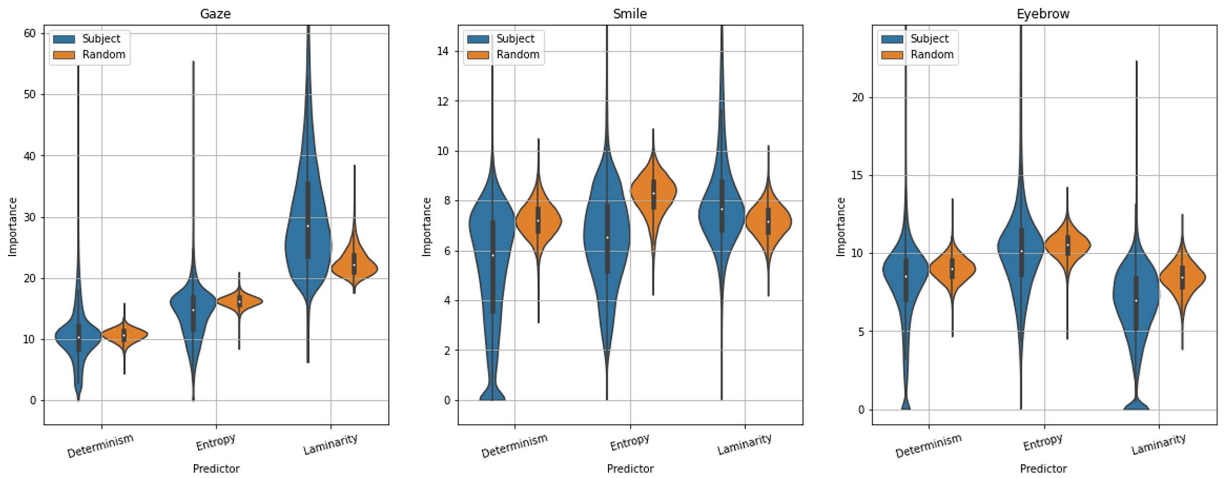


FIGURE 14
Violin plots of feature importance percentages for **(A)** gaze, **(B)** smile, and **(C)** eyebrow features in the binary classification model between paired (SP) and individual (SS) group subjects, for two train/test splitting procedures: random and subject-wise. The violin plots are bounded by the maximum 99th percentile of data within the window. Distributions were computed with 40,000 cross-validation iterations.

music], [Making love], and [Child crying]. For the eyebrow response (AU2), visually the ISC matrixes were much less distinct, however separation of SP and SS groups could be seen [Making love], [Man yelling before shooting], [Heartbeat], and to a lesser extent in others. In the case of gaze angle, a similarly weaker distinction was visible when compared to the smile response, however, [African music] and [Child crying] remain tentatively distinct, however [Making love] less so.

To handle the patterns more rigorously within the ISCs, we employed RSA with two coupling models. We also compared RDMs between stimuli. Besides finding out which stimuli produce similar results, we could also identify which matrices were entirely or almost unique in their presentation. For the smile signal, [Classical music] had a significant albeit weak similarity to [Yawning] and with the two coupling models (see Section 3.2), indicating distinguishability between two groups. The [Yawning] however was similar to other stimuli, but neither model 1 (strong coupling) or 2 (weak coupling). This could indicate that *the similarity arises from particular pairs and their interaction dynamics* rather than group differentiation. Furthermore, the stimuli [Woman screaming] seemed to be unique, following neither model nor having significant similarity to any other stimuli. Regarding the eyebrow response, a correlation with Model 1 was seen, although [African music], [Making love], [Heartbeat] and [Classical music] again fit the Model 2 assumptions, showing the separation of SS and SP in general.

The analysis of smile (AU12) provided the strongest distinction between the groups and stimuli, while eyebrow and gaze angle modality were less useful. This is possibly due to the noisy nature of the data itself and could have been expected from a visual inspection of the RDMs themselves. Furthermore, since ISC only measures linear correlation between signals, it cannot take into account any non-linear or delayed forms of synchrony that might be present. However, while not evident from our ISC analyses, all responses turned out to be informative of the presence of dyadic interaction between the pairs.

4.1.2. Context-dependent dyadic behaviors detected in audio drama

The short audio clips in the IADS-2 database involve emotional situations categorized based on the emotional valence and level of arousal they elicit in participants (Yang et al., 2018). These clips, for instance, [woman screaming] or [child laughing], are effective due to their generality in terms of basic emotions, such as fear, sadness, or happiness. However, they lack time-dependent contextualizations characteristic for long durational narratives, not to mention contexts related to the lived experience of each participant. Our analysis allowed for detecting socially relevant events that reflect dyadic interactions (SP) and are not similarly present in the data from single listeners (SS).

We detected 13 moments in the audio drama (see Figure 12) where AU2, AU12, and gaze data suggested dyadic interactions between the subjects who had a pair. A description of these moments is listed in Supplementary Table S3. These moments take place in three environments, around the family dinner table (timepoints 1–4), in the Elderly Nursing Home (5–7), and at the stairway of an apartment building and an apartment in the same building (8–13). The detected dyadically active moments mainly reveal socially relevant information or behaviors that deviate from generally accepted social norms, such as infidelity, racist naming, or swearing in rage. The content of many of these 13 moments also implies internal references to previous events in the audio drama, carrying context-determining knowledge that is expected to affect reactions at later time points.

Figure 12 depicts the continuous annotations of arousal and valence by the subjects while listening again to the audio drama in their own time at home after the experiment. Interestingly, based on visual inspection, all 13 moments are located around time-locked (negative or positive) peaks of either arousal, valence, or both. Although mainly tragic and painful for the characters in the audio drama, these events could also elicit humorous, sarcastic, or even empathic responses in the subjects, this in turn, creating social pressure to share such feelings or experiences with the other person in the room. It's plausible that besides the audio drama content itself, the enhanced memory of socially relevant dyadic interaction with the partner was occurring during those moments.

4.2. Social presence and complexity of recurrence patterns

The ISC allows observing correlations across multiple subjects but does not provide information on the dynamics of an individual's data. Therefore, we employed RQA to generate descriptors of individuals' responses that are robust against non-stationarities and also interpretable in the scope of non-linear time series analysis.

We used the quantifiers yielded by RQA as predictors in a gradient boosting binary classification algorithm (CatBoost). The aim was to see if single and paired subjects could be identified based on their individual eyebrow-, smile- and gaze properties. Using the CatBoost model, the classification was successful with mean accuracies of 67% (subject-wise split) and 75% (random split), both above chance level, thus confirming that paired setup had a fundamental effect on the autorecurrence properties of the signals. It's noteworthy that for the subject-wise splitting procedure, where the testing and training sets were composed of distinct individuals, produced a wide range of accuracies; some lower than the chance level, some higher than the relatively tight spread of the random-split procedure. This highlights the impact of individual differences in predicting dyadic interaction from the given quantifiers and was investigated more in-depth via the relative importance of the predictors. This is discussed in more detail in the Supplementary material.

Looking at the relative importance of classifiers, the most important features were associated with gaze, and specifically its laminarity quantifier. The latter being the count of recurrent points forming vertical (or horizontal) lines describes a rate of stagnancy - periods of no change in the signal's value (*cf.* Figure 3). Another way to describe the situation is that laminarity indicates how likely the system is to be trapped in specific states at any given time, that is to say, the response value changes slowly or does not change at all during some time interval. The mean length of these states is known as the trapping time, which describes the average time the system remains trapped in some stable state. Thus, we identified that the simple absence of movement in the gaze response is a main predictor of dyadic interaction. This result is in hindsight expectable, as another person in the room giving rise to dyadic interactions is expected to alter the movement of the participants' eyes. However, this effect was notably stronger for the subject-wise splitting procedure. That is to say, a model trained on one set of individuals learned to over-value the simple absence of dynamic eye movement and thus underperformed in predicting dyadic interaction among individuals with a more varied range of movement in this signal. So, while the absence of eye movement is very important, it is not the be-all and end-all, as it is not solely responsible for the predictive capacity based on gaze.

For purely predictive purposes, it appears thus that gaze is behaving almost like a trigger, for whether some interaction is taking place or not. Meanwhile, more complex communication is seen in the smile and eyebrow responses, although the smile and eyebrow responses were both less important on the whole compared to gaze, the ordering of their RQA quantifiers by importance turned out informative. For the smile signal, entropy was most important for the random splitting procedure, indicating that the complexity in an individual's smile response was what best described dyadic interaction.

Entropy in RQA describes the complexity of the deterministic structure in the data and gives a measure of how much information is needed to recover the system dynamics from the noisy and limited data (in terms of a stochastic ensemble, we have one iteration). A low entropy indicates that the length of the longest deterministic (in RQA, an "exactly" repeated) segment is short and has low variability, characteristic of chaotic behavior, while high entropy is characteristic of periodic behaviors. We stress that while this is not as impactful in terms of predictive capability, the latter can be improved by fine-tuning and the inclusion of more data and context. The primary information obtained in this manner is however that while gaze dictates well a general presence of another person in this setup, the interaction causes a noticeable uptick in periodic behaviors of different lengths in smile and eyebrow movement.

In the subject-wise splitting procedure, the classification models were not always able to learn the value of entropy in the smile signal and thus performed worse. This was demonstrated in more detail in the Supplementary material, where the importance of the subject-wise splitting procedure was further divided into high and low accuracy percentage samples (see Supplementary Figure S2). Given that the importance scores are based on the training set, the lack of weight on entropy as a predictor of dyadic interaction thus yielded lower accuracies in the prediction task.

For the eyebrow signals, no change in ordering by importance occurred when changing the train/test splitting method, with the main difference being a wider spread in the importance around a mean, and entropy being most important overall. This indicates that the eyebrow response was relatively robust in its complexity in all individuals, in contrast to the smile response. Even given the wider spread in importance, the same structure is present in the further divided plot (see Supplementary Figure S2). It shows that while individuals might significantly vary in their personality and the SP and SS groups vary fundamentally, their eyebrow movement contains a consistent amount of variability regardless.

4.3. Socially meaningful audio drama events enhance dyadic interaction

The experiment had two factors that were assumed to affect the physiology of the participants, namely, (1) the dramatic content of the audio stimuli which was the same for all participants independent of the listening situation, and (2) the listening situation, which was either alone or with the presence of another listener. On one hand, it has been shown that when subjects are engaged with narratives, their responses tend to align with other subjects in synchrony with the specific narrative events (Regev et al., 2013; Tikka et al., 2018). On the other hand, in the presence of other people, the physiology and behavior of the person tend to align with those of others (Holler, 2022). While the recognition of the emotional states of other persons has shown to be especially contagious, for instance, regarding anxiety (Shaw et al., 2021), or the spontaneous smile of a co-stranger (Golland et al., 2019), affective responses to stimuli may not rely on unconscious perceptual processes only but require attentive awareness of the semantic content of the stimulus (Lähteenmäki et al., 2019).

Our results showed that the behaviors of those who listened alone were more aligned with other single listeners than the behaviors of those who had a pair. The listening situation was affected by the observed facial behavior of the pair on the other side of the table. The socio-emotional events in the audio drama triggered non-verbal responsive expressions between the pairs, which altered in a dialogical manner, likely depending on the personality of each participant. One interpretation of the lower synchronization could be that the engagement with the social face-to-face interaction during audio drama may have required socially contextualized efforts that could lead to redistribution of available affective-cognitive resources, which in the case of a single listener condition would be directed solely to the engagement with the narrative. It's also possible that dyads were doing more complicated interactions not captured by linear ISC or were simply distracted from listening to the narrative.

In face-to-face situations, following (Hasson and Frith, 2016), it may be assumed that during a physical co-presence with another person in a room, even without tasked bodily or verbal interaction, the behavior of the participants will gradually start to be dynamically coupled. In a motivated interaction or conversation situation, the co-present parties consciously take turns, adapting their behavior to the behavior of the other (Trujillo and Holler, 2023). Again, this type of synchrony could be more complicated than linear correlation used in computing ISC, hence resulting in low apparent synchrony using ISC. In our experimental situation, the two co-present participants (SPs) were instructed not to talk with each other, but only to attend to the auditory stimuli and pay attention to the other subject's feelings on the opposite side of the table. These instructions were used as a practical matter of preventing subjects from moving, acting or speaking aloud so that it would disturb data collection, however without biasing the experimental situation. The emotional narrative was considered a more powerful driver of emotions and social dynamics than the instructions we used.

Although our findings are based on three specific facial signals, we assume that a manifestation of this socially determined feature might be observable in any other multiple signal source combinations of the dyadic group, thus allowing to separate the people with pairs (SP) from the individual participants (SS), based on their ISC values and even the individual time-series based on RQA quantifiers.

4.4. Limitations of the study

In this study, we concentrated on three data sources from a range of multiple data sources collected in the experiment. This was a limitation considering the possible information embedded in the other signal sources that were not used. However, we also consider this as a strength, as pinpointing and working only with the three strongest facial behavioral modalities (AU2/eyebrow, AU12/smile, and gaze) allowed us to make more direct assumptions on the functions of each modality in terms of the socially contextualized study set-up. Consequently, interpretation of the findings is straightforward, in comparison to well-known interpretation challenges related, for instance, heart rate variances, or even more so with EEG and related neural signals. Yet, the interpretation of data of easy access modalities such as a smile or lift of an eyebrow may embed *ad hoc* interpretation biases inherited from 'folk psychology,' introducing limitations in its one right. While our experiment revealed notable differences in signals between individual and paired subjects, additional control variables and/or stimuli are needed to further explore the origin and factors driving those differences. For example, using more or less engaging narratives and introducing other types of interactions (see Future directions) could change the relative importance of the narrative versus social interaction.

We did not quantitatively measure the level of attention of participants towards the audio narrative, instead, we relied on the qualitative questionnaire ("What kind feelings did the story evoke?") and post-experiment discussions about the content of the story, which allowed us to ensure that all subjects recalled at least some events in the story. In our study, all SP subjects were assigned randomly with their partners instead of measuring and matching partner sympathy ratings explicitly. This may have influenced the higher variability in the partnered participants.

Finally, the number of participants was relatively small and limited to single listeners and to the pairs that were unfamiliar to one another. While the unfamiliarity in the pair condition was deliberately chosen as one of the participant selection factors, this due to the expectation that strangeness of the other person would generate sufficiently socially driven distraction to the dyadic setting compared to the single-person setting, to properly study the actual effect of unfamiliarity us such we should have expanded the paired participant group with the familiar pairs. Due to our orientation towards elaborating novel methodic pathways, the study of familiarity vs. unfamiliarity was excluded from this experiment in favor of focusing on the single vs. paired groups.

We furthermore acknowledge that additional data collection is necessary to link behavioral and personal traits with properties of signals, such as synchrony and autorecurrence. Specifically for the interpretation and analysis of recurrence quantifiers, a larger sample size can help add confidence and clarify the role of entropy vs. laminarity among the chosen modalities, based on individual differences, as current results could too strongly be influenced by outliers.

4.5. Future directions

We aim to eventually apply the analysis framework developed and tested here for large-scale multisource data acquired from participants' engaging with different social activities, not only listening to a narrative. Studying recurrence properties of physiological signals during interaction is important to better understand and model human responses to long-duration narratives in social contexts. Such information is valuable in modeling and predicting human responses to audio-visually mediated or live social situations. In addition to increasing social and psychological understanding of the dynamics of dyadic non-verbal interaction between two people, such knowledge contributes to the fields of human-computer interaction and social robotics. In this study, we limited the RQA analysis to individual signals, but extensions to multiple signals, e.g., between different modalities or individuals, including joint-RQA and cross-RQA exist (Webber and Zbilut, 2005). Here we proposed to use a classification algorithm that can yield a metric of importance for predictors, but application as numerous. For example, one might be interested in developing the RQA further on specific segments, reconstructing the proper phase space with time-delayed embedding, and perhaps even introducing cross-RQA, which is again a pairwise analysis of recurrence on two series. We focused on SP/SS group differentiation rather than individual identification; nevertheless, we highlight the ethical implications inherent in developing AI models that may be capable of individual-level physiological signal detection. Consequently, careful ethical consideration would be essential for any research moving in this direction. Finally, in our setup, we evaluated signal changes with face-to-face paired listening with another person. However, numerous other potentially interesting setups presumably induce signal changes. For example, what if another person is not face-to-face, but with a 45° angle, or there are multiple persons, an animated robot, or just a display? Each of these scenarios could leave their specific fingerprints on the physiological signals.

5. Conclusion

The face-to-face presence of another unfamiliar person in a shared context may be a strong driver for variations in the physiological signals. The question we tackled was to what extent this social phenomenon overrides the interpersonal synchronization induced by a dramatized context. Our experimental setting revealed differences between single listeners and dyadic pairwise patterns when listening to audio drama. We concentrated on two computational methods, ISC and RQA, that can be applied in the analysis of dyadic physiological signals without the need for manual annotations.

Our ISC analysis showed that the audio stimuli induced more unique responses in those subjects who were listening to it in the presence of another person, while single listeners tended to yield a more uniform group response as it was driven by the audio drama alone. In other words, the behaviors of paired individuals correlated less with one another, compared to the single listeners. The ISC values of pairs, whether calculated on true pairs or any two individuals who had a partner, are lower than the group with single individuals. Most likely the signals generated by pairs were not driven only by the audio drama stimulus, but also by their mutual facial interactions that were unique to particular pairs.

The strongest distinction in facial interaction between the two groups (paired vs. single listeners) was found with three sound localizers containing distinguished socially charged positive or negative valences. Out of the three facial signal sources that we studied, the 'smile' (Lip Corner Puller, AU12) may best indicate the momentary non-verbal facial dialogue in search of agreement on responses to the shared affective context. In the 27-min long audio drama we pinpointed 13 moments with distinctive dyadic interactions between the subjects who had a pair. Using RQA, we found that recurrence patterns in an individual's responses carry information about dyadic interactions, via a classification task, and accounting for the entropy or complexity of recurrent patterns was important for the latter, especially so for the smile response. While the gaze was mainly a laminarity-based predictor of the presence of another person, the undervaluing of entropy related to worse classifications in all cases, demonstrates the way RQA can describe nonlinear effects of dyadic interaction on one's response signals.

Regarding the two main tools used in the analysis, RSA and RQA, we demonstrated how the more conventional RSA is indeed effective in analyzing dyadic interactions, highlighting moments of a clear division between the two studied groups. However, this method is limited by linearity and requires pairwise comparisons. RQA however applies to a single time series with non-linear properties. The two methods target different aspects of the studied phenomenon. In short, RSA tells us where differences occur as a whole, and the latter can be compared easily to hypothetical similarity matrix structures. In our case this showed a reduction of correlation for dyadic interaction, most likely indicating a nonverbal dialogical communication. RQA tells us how individual time-series differ by their non-linear recurrent properties, which were found in the study to follow the second initial hypothesis, that dyadic interaction influences the complexity of recurrent dynamics. Specifically, the novel combination of RQA and classification tasks has many further avenues to explore regarding the data. The latter, however, should be investigated in the future in the context of dyadic interaction. This work serves as a methodological first step in this direction. Our work also contributes towards modeling and predicting human social behaviors to specific types of audio-visually mediated, virtual, and live social situations.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Local institute ethics committee (Aalto University). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JaK, PT, and JnK participated in planning the study, collecting the data and supervising. SP, JnK, and JaK participated in analyzing the data. All authors participated in discussing results, preparing manuscript and approval of the manuscript.

Funding

This work has been supported by Aalto Starting Grant (PT), the EU Mobilitas Pluss Top Researcher Grant MOBTT90 (PT) and the Tallinn University Research grant TF/1522 (PT).

Acknowledgments

We thank AALTO Neuroimaging, Veli-Matti Saarinen for assistance, Meeri Mäkäräinen for help in data collection and eye-gaze and facial tracking analysis and Balraj Rathod for help in data collection. The recordings were conducted at the Aalto Behavioral Laboratory (ABL), Aalto NeuroImaging, Aalto University School of Science.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1153968/full#supplementary-material

References

Ames, D. L., Honey, C. J., Chow, M. A., Todorov, A., and Hasson, U. (2014). Contextual alignment of cognitive and neural dynamics. *J. Cogn. Neurosci.* 27, 655–664. doi: 10.1162/jocn_a_00728

Arunvinthan, S., Gopal, R., Chandrasekar, V. K., and Nadaraja Pillai, S. (2020). Recurrence analysis of surface pressure characteristics over symmetrical aerofoil. *Chaos* 30:013116. doi: 10.1063/1.5121569

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bögels, S. (2020). Neural correlates of turn-taking in the wild: response planning starts early in free interviews. *Cognition* 203:104347. doi: 10.1016/j.cognition.2020.104347

Boldt, R., Malinen, S., Seppä, M., Tikka, P., Savolainen, P., Hari, R., et al. (2013). Listening to an audio drama activates two processing networks, one for all sounds, another exclusively for speech. *PLoS One* 8:E64489. doi: 10.1371/journal.pone.0064489

Boldt, R., Seppä, M., Malinen, S., Tikka, P., Hari, R., and Carlson, S. (2014). Spatial variability of functional brain networks in early-blind and sighted subjects. *Neuro Image* 95, 208–216. doi: 10.1016/j.neuroimage.2014.03.058

Bradley, M. M., and Lang, P. J. (2007). The international affective digitized sounds: affective ratings of sounds and instruction manual (technical report no. B-3). Gainesville, FL: University of Florida, NIMH Center for the sStudy of eEmotion and aAttention

Brick, T. R., Gray, A. L., and Staples, A. D. (2017). Recurrence quantification for the analysis of coupled processes in aging. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 73, 134–147. doi: 10.1093/geronb/gbx018

Canento, F., Fred, A., Silva, H., Gamboa, H., and Lourenco, A. (2011). Multimodal biosignal sensor data handling for emotion recognition. *IEEE Sensors*, Ireland: Limerick 647–650. doi: 10.1109/ICSENS.2011.6127029

Chauhan, V., di Oleggio, V., Castello, M., Soltani, A., and Gobbini, M. I. (2017). Social saliency of the Cue slows attention shifts. *Front. Psychol.* 8:738. doi: 10.3389/fpsyg.2017.00738

Chib, V. S., Adachi, R., and O'Doherty, J. P. (2018). Neural substrates of social facilitation effects on incentive-based performance. *Soc. Cogn. Affect. Neurosci.* 13, 391–403. doi: 10.1093/scan/nsy024

Curtin, P., Curtin, A., Austin, C., Gennings, C., Tammimies, K., Bölte, S., et al. (2017). Recurrence quantification analysis to characterize cyclical components of environmental elemental exposures during fetal and postnatal development. *PLoS One* 12:e0187049. doi: 10.1371/journal.pone.0187049

De la Torre, F., Chu, W.-S., Xiong, X., Vicente, F., Ding, X., and Cohn, J. (2015). Intra face. IEEE International Conference on Automatic Face & Gesture Recognition and Workshops

Dorogush, A.V., Ershov, V., and Gulin, A. (2018). Cat *Boost: Gradient boosting with categorical features support. Ar xiv, ar xiv: 1810.11363.*

Ekman, P, and Friesen, W. (1978). *Facial action coding system: A technique for the measurement of facial movement.* Palo Alto: Consulting Psychologists Press

Fusaroli, R., Konvalinka, I., and Wallot, S. (2014). "Analyzing social interactions: the promises and challenges of using cross recurrence quantification analysis" in *Translational recurrences: Proceedings in Mathematics & Statistics, 103.* eds. N. Marwan, M. Riley, A. Giuliani and WebberC. Jr. (Cham: Springer), 137–155.

George, N. (2016). A neuroimaging point of view on the diversity of social cognition: evidence for extended influence of experience- and emotion-related factors on face processing. *Cult. Brain* 4, 147–158. doi: 10.1007/s40167-016-0043-6

Golland, Y., Mevorach, D., and Levit-Binnun, N. (2019). Affiliative zygomatic synchrony in co-present strangers. *Sci. Rep.* 9:3120. doi: 10.1038/s41598-019-40060-4

Goswami, B. (2019). A brief introduction to nonlinear time series analysis and recurrence plots. *Vibration* 2, 332–368. doi: 10.3390/vibration2040021

Guevara, M., Cox, R. F. A., van Dijk, M., and van Geert, P. (2017). Attractor dynamics of dyadic interaction: A recurrence based analysis. *Nonlinear Dynamics Psychol. Life Sci.* 21, 289–317.

Hamilton, A. F. D. C. (2016). Gazing at me: the importance of social meaning in understanding direct-gaze cues. *Philos. Trans. R. Soc B Biol. Sci.* 371:20150080. doi: 10.1098/rstb.2015.0080

Hancock, J. T., and Khoshgoftaar, T. M. (2020). Cat boost for big data: an interdisciplinary review. *J. Big Data* 7:94. doi: 10.1186/s40537-020-00369-8

Hasson, U., and Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 371:20150366. doi: 10.1098/rstb.2015.0366

Hasson, U., Yuval, N., Levy, I., Fuhrmann, G., and Malach, R. (2004). Inter-subject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Holler, J. (2022). Visual bodily signals as core devices for coordinating minds in interaction. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 377:20210094. doi: 10.1098/rstb.2021.0094

Iwanski, J. S., and Bradley, E. (1998). Recurrence plots of experimental data: to embed or not to embed? *Chaos* 8, 861–871. doi: 10.1063/1.166372

Jääskeläinen, I. P., Koskentalo, K., Balk, M. H., Autti, T., Kauramäki, J., Pomren, C., et al. (2008). Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. *Open Neuroimage. J.* 2, 14–19. doi: 10.2174/1874440000802010014

Jääskeläinen, I. P., Sams, M., Glerean, E., and Ahveninen, J. (2021). Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage* 224:117445. doi: 10.1016/j.neuroimage.2020.117445

Jola, C., McAleer, P., Grosbras, M. H., Love, S. A., Morison, G., and Pollick, F. E. (2013). Uni- and multisensory brain areas are synchronised across spectators when watching unedited dance recordings. *i-Perception* 4, 265–284. doi: 10.1068/i0536

Kauttonen, J., Hlushchuk, Y., Jääskeläinen, I. P., and Tikka, P. (2018). Brain mechanisms underlying cue-based memorizing during free viewing of movie memento. *NeuroImage* 172, 313–325. doi: 10.1016/j.neuroimage.2018.01.068

Kauttonen, J., Hlushchuk, Y., and Tikka, P. (2015). Optimizing methods for linking cinematic features to fMRI data. *NeuroImage* 110, 136–148. doi: 10.1016/j.neuroimage.2015.01.063

Kennel, M. B., Brown, R., and Abarbanel, H. D. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* 45, 3403–3411. doi: 10.1103/PhysRevA.45.3403

Koskinen, M., Kurimo, M., Gross, J., Hyvärinen, A., and Hari, R. (2020). Brain activity reflects the predictability of word sequences in listened continuous speech. *NeuroImage* 219:116936. doi: 10.1016/j.neuroimage.2020.116936

Koskinen, M., and Seppä, M. (2014). Uncovering cortical MEG responses to listened audiobook stories. *NeuroImage* 100, 263–270. doi: 10.1016/j.neuroimage.2014.06.018

Kragel, P. A., and Labar, K. S. (2013). Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions. *Emotion* 13, 681–690. doi: 10.1037/a0031820

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis – Connecting the branches of systems neuroscience. *Front. Syst. Neurosci.*:2. doi: 10.3389/neuro.06.004.2008

Lähteenmäki, M., Kauramäki, J., Sauter, D. A., and Nummenmaa, L. (2019). Auditory affective processing requires awareness. *Emotion* 19, 53–69. doi: 10.1037/emo0000388

Lankinen, K., Saari, J., Hlushchuk, Y., Tikka, P., Parkkonen, L., Hari, R., et al. (2018). Consistency and similarity of MEG- and fMRI-signal time courses during movie viewing. *NeuroImage* 173, 361–369. doi: 10.1016/j.neuroimage.2018.02.045

Lumsden, J., Miles, L. K., Richardson, M. J., Smith, C. A., and Macrae, C. N. (2012). Who syncs? Social motives and interpersonal coordination. *J. Exp. Soc. Psychol.* 48, 746–751. doi: 10.1016/j.jesp.2011.12.007

March, T., Chapman, S., and Dendy, R. (2005). Recurrence plot statistics and the effect of embedding. *Physica D: Nonlinear Phenomena* 200, 171–184. doi: 10.1016/j.physd.2004.11.002

Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Phys. Rep.* 438, 237–329. doi: 10.1016/j.physrep.2006.11.001

Miura, N., and Noguchi, S. (2022). The presence of adjacent others facilitates interpersonal neural synchronization in the left prefrontal cortex during a simple addition task. *Sci. Rep.* 12:12662. doi: 10.1038/s41598-022-16936-3

Nakano, T. (2015). Blink-related dynamic switching between internal and external orienting networks while viewing videos. *Neurosci. Res.* 96, 54–58. doi: 10.1016/j.neures.2015.02.010

Nastase, S. A., Gazzola, V., Hasson, U., and Keyser, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* 14, 667–685. doi: 10.1093/scan/nsz037

Neufeld, J., Ioannou, C., Korb, S., Schilbach, L., and Chakrabarti, B. (2016). Spontaneous facial mimicry is modulated by joint attention and autistic traits. *Autism Res.* 9, 781–789. doi: 10.1002/aur.1573

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Comput. Biol.* 10:e1003553. doi: 10.1371/journal.pcbi.1003553

Pikovsky, A, Rosenblum, M, and Kurths, J (2001). *Synchronization. A universal concept in nonlinear sciences.* Cambridge University Press: Cambridge

Piskun, O., and Piskun, S. (2011). *Recurrence quantification analysis of financial market crashes and crises. Ar xiv preprint ar xiv: 1107.5420.* arXiv: Chaotic Dynamics.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., and Gulin, A. (2018). Cat boost: unbiased boosting with categorical features. In NIPS'18: Proceedings of the 32nd international conference on neural information processing systems, 6639–6649

Regev, M., Honey, C. J., Simony, E., and Hasson, U. (2013). Selective and invariant neural responses to spoken and written narratives. *J. Neurosci. Off. J. Soc. Neurosci.* 33, 15978–15988. doi: 10.1523/JNEUROSCI.1580-13.2013

Ruvolo, P., Messinger, D., and Movellan, J. (2015). Infants time their smiles to make their moms smile. *PLoS One* 10:e0136492. doi: 10.1371/journal.pone.0136492

Saarimäki, H. (2021). Naturalistic stimuli in affective neuroimaging: A review. *Front. Hum. Neurosci.* 15:675068. doi: 10.3389/fnhum.2021.675068

Sham, A. H., Tikka, P., Lamas, D. R., and Anbarjafari, G. (2022). Automatic reaction emotion estimation in a human-human dyadic setting using deep neural networks. *Signal Image Video Process* 17, 527–534. doi: 10.1007/s11760-022-02257-5

Shapiro, K. L., Hanslmayr, S., Enns, J. T., and Lleras, A. (2017). Alpha, beta: the rhythm of the attentional blink. *Psychon. Bull. Rev.* 24, 1862–1869. doi: 10.3758/s13423-017-1257-0

Shaw, P. V., Wilson, G. A., and Antony, M. M. (2021). Examination of emotional contagion and social anxiety using novel video stimuli. *Anxiety Stress Coping* 34, 215–227. doi: 10.1080/10615806.2020.1839729

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., et al. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* 7:12141. doi: 10.1038/ncomms12141

Smallwood, J., and Schooler, J. W. (2015). The science of mind wandering: empirically navigating the stream of consciousness. *Annu. Rev. Psychol.* 66, 487–518. doi: 10.1146/annurev-psych-010814-015331

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* 3, 42–55. doi: 10.1109/T-AFFC.2011.25

Soloviev, V., Serdiuk, O., Semerikov, S., and Kiv, A. (2020). Recurrence plot-based analysis of financial-economic crashes. CEUR Workshop Proceedings

Stark, E. A., Cabral, J., Riem, M. M. E., Van IJzendoorn, M. H., Stein, A., and Kringelbach, M. L. (2020). The power of smiling: the adult brain networks underlying learned infant emotionality. *Cereb. Cortex* 30, 2019–2029. doi: 10.1093/cercor/bhz219

Stephens, C. L., Christie, I. C., and Friedman, B. H. (2010). Autonomic specificity of basic emotions: evidence from pattern classification and cluster analysis. *Biol. Psychol.* 84, 463–473. doi: 10.1016/j.biopsycho.2010.03.014

Stephenson, L. J., Edwards, S. G., and Bayliss, A. P. (2021). From gaze perception to social cognition: the shared-attention system. *Perspect. Psychol. Sci.* 16, 553–576. doi: 10.1177/1745691620953773

Stevanovic, M., Tuhkanen, S., Järvensivu, M., Koskinen, E., Savander, E., and Valkia, K. (2021). Physiological responses to proposals during dyadic decision-making conversations. *PLoS One* 16:e0244929. doi: 10.1371/journal.pone.0244929

Stuldreher, I. V., Thammasan, N., van Erp, J. B. F., and Brouwer, A. M. (2020). Physiological synchrony in EEG, Electrodermal activity and heart rate detects Attentionally relevant events in time. *Front. Neurosci.* 14:575521. doi: 10.3389/fnins.2020.575521

Takens, F. (2006). "Detecting strange attractors in turbulence" in *Dynamical systems and turbulence, Warwick 1980: Proceedings of a symposium held at the University of Warwick 1979/80* (Berlin, Heidelberg: Springer Berlin Heidelberg), 366–381.

Thiel, M., Romano, M. C., Kurths, J., Meucci, R., Allaria, E., and Arecchi, F. T. (2002). Influence of observational noise on the recurrence quantification analysis. *Physica D Nonlinear Phenomena* 171, 138–152. doi: 10.1016/S0167-2789(02)00586-9

Tikka, P., Kaipainen, M., and Salmi, J. (2023). Narrative simulation of social experiences in naturalistic context – a neurocinematic approach. *Neuropsychologia* 188:108654. doi: 10.1016/j.neuropsychologia.2023.108654

Tikka, P., Kauttonen, J., and Hlushchuk, Y. (2018). Narrative comprehension beyond language: common brain networks activated by a movie and its script. *PLoS One* 13:e0200134. doi: 10.1371/journal.pone.0200134

Tricoche, L., Ferrand-Verdejo, J., Pélisson, D., and Meunier, M. (2020). Peer presence effects on eye movements and attentional performance. *Front. Behav. Neurosci.* 13:280. doi: 10.3389/fnbeh.2019.00280

Trujillo, J. P., and Holler, J. (2023). Interactionally embedded gestalt principles of multimodal human communication. *Perspect. Psychol. Sci.* 18, 1136–1159. doi: 10.1177/17456916221141422

Turkheimer, F. E., Rosas, F. E., Dipasquale, O., Martins, D., Fagerholm, E. D., Expert, P., et al. (2022). A complex systems perspective on neuroimaging studies of behavior and its disorders. *Neuroscientist* 28, 382–399. doi: 10.1177/1073858421994784

van den Hoorn, W., Hodges, P. W., van Dieën, J. H., and Kerr, G. K. (2020). Reliability of recurrence quantification analysis of postural sway data. A comparison of two methods to determine recurrence thresholds. *J. Biomech.* 107:109793. doi: 10.1016/j.jbiomech.2020.109793

Varela, F. J., Thompson, E., and Rosch, E. (1991). *Embodied mind: cognitive science and human experience*. Cambridge, Massachusetts: MIT Press.

Wallot, S., and Leonardi, G. (2018). Analyzing multivariate dynamics using cross-recurrence quantification analysis (CRQA), diagonal-cross-recurrence profiles (DCRP), and multidimensional recurrence quantification analysis (MdRQA)--a tutorial in R. *Front. Psychol.* 9:2232. doi: 10.3389/fpsyg.2018.02232

Webber, C. L. Jr., Marwan, N., Facchini, A., and Giuliani, A. (2009). Simpler methods do it better: success of recurrence quantification analysis as a general purpose data analysis tool. *Phys. Lett. A* 373, 3753–3756. doi: 10.1016/j.physleta.2009.08.052

Webber, C. L., and Marwan, N. Eds. (2015). "Recurrence quantification analysis" in *Theory and best practices* (Cham: Springer)

Webber, C. L., and Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In M. Riley and OrdenG. Van (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* 26–94. USA: National Science Foundation.

Xu, T. L., de Barbaro, K., Abney, D. H., and Cox, R. F. A. (2020). Finding structure in time: visualizing and analyzing behavioral time series. *Front. Psychol.* 11:1457. doi: 10.3389/fpsyg.2020.01457

Yang, W., Makita, K., Nakao, T., Kanayama, N., Machizawa, M. G., Sasaoka, T., et al. (2018). Affective auditory stimulus database: an expanded version of the international affective digitized sounds (IADS-E). *Behav. Res. Methods* 50, 1415–1429. doi: 10.3758/s13428-018-1027-6

Young, K. S., Parsons, C. E., Jegindoe Elmholdt, E. M., Woolrich, M. W., van Hartevelt, T. J., Stevner, A. B. A., et al. (2016). Evidence for a caregiving instinct: rapid differentiation of infant from adult vocalizations using magnetoencephalography. *Cereb. Cortex* 26, 1309–1321. doi: 10.1093/cercor/bhv306

Zajonc, R. B. (1965). Social facilitation. *Science* 149, 269–274. doi: 10.1126/science.149.3681.269

Zbilut, J. P., Thomasson, N., and Webber, C. L. (2002). Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals. *Med. Eng. Phys.* 24, 53–60. doi: 10.1016/s1350-4533(01)00112-6

Zbilut, J. P., and Webber, C. L. Jr. (1992). Embeddings and delays as derived from quantification of recurrence plots. *Phys. Lett. A* 171, 199–203. doi: 10.1016/0375-9601(92)90426-M

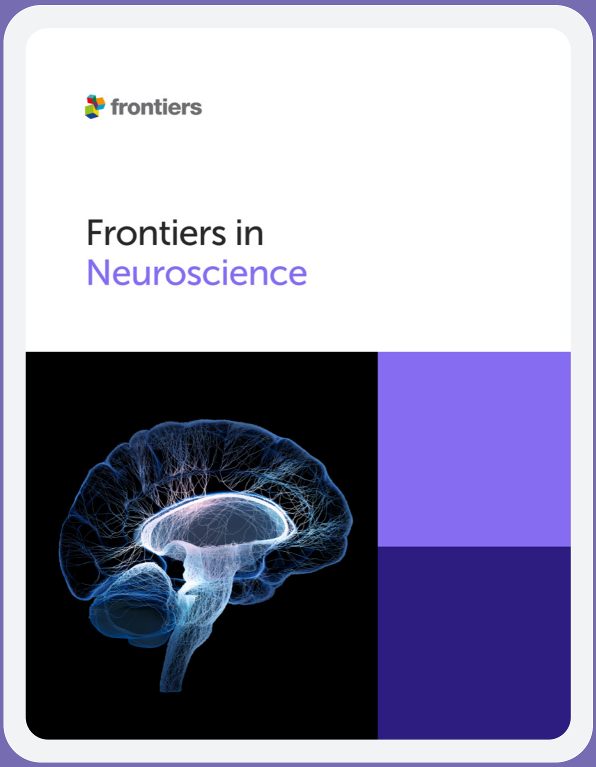
Frontiers in Neuroscience

Provides a holistic understanding of brain function from genes to behavior

Part of the most cited neuroscience journal series which explores the brain - from the new eras of causation and anatomical neurosciences to neuroeconomics and neuroenergetics.

Discover the latest Research Topics

See more →



frontiers

Frontiers in
Neuroscience

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

frontiers | Research Topics