Aus dem Institut für Systemische Neurowissenschaften
der Heinrich-Heine-Universität Düsseldorf

Movies in the Magnet

Dissertation

zur Erlangung des Grades eines „Doctor of Philosophy" (PhD) in Medical Sciences
der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Lisa Noreen Mochalski
(2023)

Als Inauguraldissertation gedruckt mit der Genehmigung der
Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf
gez.:
Dekanin/Dekan: Prof. Dr. med. Nikolaj Klöcker
Gutachter/innen: PD Dr. Susanne Weis, Prof. Dr. Julian Caspers

Zusammenfassung

Naturalistic viewing (NV) ist ein vielversprechender Ansatz zur Untersuchung des Gehirns in einer ökologisch valideren Umgebung als Experimente oder Ruhemessungen (RS). In NV schauen ProbandInnen Filme während einer funktionellen Magnetresonanztomographie (fMRT). Filme sind komplexe, kontinuierliche und dynamische Stimuli, die dem echten Leben in vielerlei Hinsicht ähneln (Bottenhorn et al., 2018; Sonkusare et al., 2019). NV ist weniger kontrolliert als Experimente, aber bietet im Gegensatz zu RS inhaltlich bedeutungsvolle Stimulation (Finn et al., 2017) und praktische Vorteile, wie erhöhte Compliance (Vanderwal et al., 2019; Eickhoff et al., 2020). Filme wurden verwendet, um Beziehungen zwischen Gehirn und Verhalten und individuelle Unterschiede zu untersuchen und die sozio-affektive Forschung voranzutreiben (Sonkusare et al., 2019; Saarimäki, 2021).

Die Ziele dieser Arbeit umfassen eine Beurteilung des Einsatzes von NV in sozio-affektiver und Forschung zu individuellen Unterschieden, und die Vorstellung des *Movies*-Datensatzes, der zur Erforschung dieser Themen kreiert wurde.

Zuerst wird ein Projekt vorgestellt, in dem der Einfluss eines ganzen Filmes auf die funktionelle Konnektivität in 14 funktionellen Gehirn-Netzwerken (NFC) untersucht wird. Inter- und intra-individuelle Variabilität in NFC wurden erhoben, und es wurde untersucht, wie Emotionen, die im Film dargestellt werden, diese beeinflussen.

Der *Movies*-Datensatz umfasst zwei RS, eine anatomische und acht NV (f)MRI-Aufnahmen, sowie umfassende Testungen der ProbandInnen hinsichtlich sozio-affektiver Eigenschaften als *traits* und *states*. Erweitert wurde der Datensatz durch eine Annotation der von den Filmen ausgelösten Emotionen in einer eigenen Stichprobe und eine Annotation der Eigenschaften der Filme durch zwei Beurteilende.

Daten des *Movies*-Datensatzes wurden genutzt, um zu untersuchen, wie *trait* und *state* Ängstlichkeit mit NFC in 14 Netzwerken ko-variieren. Die Ergebnisse dieser Studie zeigen, dass signifikante Korrelationen zwischen *state* Ängstlichkeit und NFC des Spiegelneuronennetzwerks durch drei bestimmte Filme hervorgerufen werden.

Sowohl die Literatur als auch meine eigene Arbeit unterstreichen, dass Film-Stimuli und ProbandInnen genau charakterisiert werden müssen, um zu verstehen, wie NV am besten in sozio-affektiver und Forschung zu individuellen Unterschieden eingesetzt werden kann. Auswahl von Stimulus, Verhalten und Gehirnmessung müssen genau abgewogen und abgestimmt werden (Eickhoff et al., 2020). Der *Movies*-Datensatz stellt einen großen fMRT-Datensatz dar zum Zweck der Untersuchung, wie diese Faktoren miteinander kombiniert werden müssen, um effektive NV-Forschung zu betreiben.

Summary

Naturalistic viewing (NV) is a promising approach to studying the brain in more ecologically valid environments than traditional task or resting state (RS) approaches allow. In NV, subjects commonly watch excerpts from or full movies while undergoing functional magnetic resonance imaging (fMRI). Movies are complex, continuous and dynamic stimuli that approximate real life in many ways (Bottenhorn et al., 2018; Sonkusare et al., 2019). NV is less constrained than traditional tasks, but still offers meaningful stimulation, which is missing in RS paradigms (Finn et al., 2017), and offers practical advantages like higher subject compliance and engagement (Vanderwal et al., 2019; Eickhoff et al., 2020). Movies have been employed to study brain-behavior relationships, individual differences, and advance socio-affective research (Sonkusare et al., 2019, Saarimäki, 2021).

The aims of the current work are an assessment of the use of NV in individual differences and socio-affective research and the presentation of the Movies dataset, which was created to study these research topics in particular.

First, a study is presented that investigates the influence of watching a full narrative movie on network functional connectivity (NFC) in 14 functional brain networks. Inter- and intra-subject variability in NFC are assessed, as well as how they are affected by emotions portrayed in the movie stimulus.

Afterwards, I present the Movies dataset, encompassing two RS, one anatomical and 8 movie (f)MRI acquisitions and extensive socio-affective phenotyping of the subjects, including robust assessment of traits and states. This dataset was extended with an annotation of the emotions induced by the movie stimuli in a separate sample and an annotation of the features of the movie stimuli by two raters.

Last, data from the Movies dataset was used to investigate the covariation in trait and state anxiety with NFC in 14 functional brain networks. Results indicate that significant correlations between state anxiety and NFC in the mirror neuron system network are not intrinsic, but elicited by three specific movie stimuli.

Both literature and my own work emphasize the conclusion that movie stimuli and subjects need to be characterized broadly and robustly to understand how NV can be best leveraged in individual differences and socio-affective research. Stimulus choice, phenotype of interest and brain measure need to be considered carefully and matched to uncover brain-behavior relationships (Eickhoff et al., 2020). The Movies dataset offers a large-scale neuroimaging dataset suitable for advancing the understanding of how these factors need to be combined for effective NV research.

Abkürzungsverzeichnis

ANOVA	analysis of variance
(a)/(t) CompCor	(anatomical)/(temporal) component-based noise correction
AS	adaptive strategies
BIC	Bayesian information criterion
BOLD	blood oxygenation level dependent
Cam-CAN	Cambridge Center for Ageing and Neuroscience
CSF	cerebrospinal fluid
FD	framewise displacement
(f)MRI	(functional) magnetic resonance imaging
FoV	field of view
FWE	family-wise error
FWHM	full-width half-maximum
FZJ	Forschungszentrum Jülich
GM	gray matter
HBN	Healthy Brain Network
HCP	Human Connectome Project
ICA	independent component analysis
IOA	inter-observer agreement
ISC	inter-subject correlation
(IS-)RSA	(inter subject-) representational similarity analysis
LMM	linear mixed model
MP-RAGE	Magnetization Prepared Rapid Acquisition with Gradient Echoes
MS	maladaptive strategies
(N)FC	(network) functional connectivity
NV	naturalistic viewing
ReMoTa	Real-time Movie Tagging

RDM	representational dissimilarity matrix
RS	resting state
SENSE	sensitivity encoding
SD	standard deviation
SDC	susceptibility distortion correction
T	tesla
TE	echo time
TI	inversion time
ToM	theory of mind
TR	repetition time
WM	white matter

movie stimuli:

DD	Dirty Dancing
DPS	Dead Poets Society
DMW	Dead Man Walking
FG	Forrest Gump
LIB	Life is Beautiful
S	Scream
SS	Short Sequences
TGTBTU	The Good, The Bad, The Ugly

questionnaires:

ADHS-E	Aufmerksamkeitsdefizit-/Hyperaktivitätsstörungsscreening für Erwachsene
ANPS	Affective Neuroscience Personality Scales
BDI	Beck's Depression Inventory
DSSQ	Dundee Stress State Questionnaire
EHI	Edinburgh Handedness Inventory
EKF-S	Emotionale-Kompetenz-Fragebogen Selbstauskunft
FEEL-E	Fragebogen zur Erhebung der Emotionsregulation bei Erwachsenen

ISK-K	Inventar sozialer Kompetenzen Kurzform
NEO-FFI	NEO Five Factor Inventory
PANAS	Positive and Negative Affect Scales
STAI	State Trait Anxiety Inventory
SAM	Self-Assessment Manikin
STAXI	State-Trait-Ärgerausdrucks-Inventar
TAS-26	Toronto Alexithymia Scale
ToM	Theory of Mind

networks:

AM	autobiographical memory network
CogAC	cognitive attention control network
eMDN	extended multiple demand network
EmoSF	emotional scene and face processing network
ER	emotion regulation network
eSAD	extended socio-affective default mode network
MNS	mirror neuron system network
Rew	reward network
SM	semantic memory network
ToM	theory of mind network
VigAtt	vigilant attention network
WM	working memory network

1. Introduction

Understanding how individual differences in brain architecture shape personality, cognitive abilities and socio-affective traits is a constant endeavor in cognitive neuroscience. The growing interest in individual differences research has led to the development of new paradigms that may allow for novel insights into individual brain architecture. Naturalistic viewing (NV) is a promising tool for designing more ecologically valid fMRI studies, thus providing the opportunity to measure individual differences under beneficial circumstances (Vanderwal et al., 2017).

NV uses stimuli that are complex, dynamic, continuous and present an approximation to the experiences we encounter in everyday life, such as movies or virtual realities (Saarimäki, 2021; Sonkusare et al., 2019). Originating in new approaches to vision research (Hasson, 2004; Bartels & Zeki, 2004), “natural vision” paradigms were introduced to better study how brains work under conditions more similar to the world outside the laboratory. Here, perception is usually multimodal, embedded in context and spans longer time periods during which changes occur. This raises questions like how similar different brains operate under such complex circumstances compared to the tightly controlled experiments previously used to carefully isolate specific brain-behavior relationships (Hasson, 2004; Bartels & Zeki, 2004). Questions which can be asked far beyond the domains of vision.

To uncover the brain’s functional architecture and study brain-behavior relationships, past functional magnetic resonance imaging (fMRI) research mainly followed one of two approaches, each of which comes with their own strengths and constraints. Traditionally, task-based fMRI has been essential for studying the neural correlates of cognition based on paradigms designed to target specific cognitive processes in a highly controlled environment (Dosenbach et al., 2006). However, the highly controlled nature of task fMRI leads to rather artificial participant experiences and thus sacrifices ecological validity for precision (Kingstone et al., 2008; Delgado-Herrera et al., 2021; Reggente et al., 2018; van Atteveldt et al., 2018). Alternatively, the intrinsic functional architecture of the brain has been studied using resting state (RS) fMRI, in which participants are scanned without any task or external stimulus. However, in this paradigm, subjects are left to follow their own thoughts, which impacts brain states in a way that is difficult to control (Gonzalez-Castillo et al., 2021).

While both task and RS-fMRI possess their strengths, their constraints have led to the development and integration of naturalistic stimuli into study designs. In NV, subjects are presented with movies or movie segments while undergoing MRI scanning, without any additional task instructions beyond attending to the movie. Movies are assumed to induce brain states similar to those evoked in real life situations, because movies are likewise complex, dynamic and continuous. Consequently, naturalistic stimuli require continuous, real-time processing of dynamic streams of information (Bottenhorn et al., 2018). The brain might even be more “tuned” to these stimuli, for example through similarities between naturalistic stimuli properties like complex temporal structures and processing in the perceptual system of the brain (Sonkusare et al., 2019). The complexity in naturalistic stimuli also allows for investigating multiple functional components simultaneously, both on a lower perceptual and higher cognitive level, and with a focus on individual variation (Shamay-Tsoory & Mendelssohn, 2019; Saarimäki, 2021). NV thus aims to deliver ecologically valid stimulation to the brain as far as this is possible within the confines of a MR scanner. Compared to task-based fMRI, participants’ brain states are less constrained while experiencing a less artificial situation during NV (Finn et al., 2017). Simultaneously, movies provide stimulation that is time-locked across participants, therefore creating a more constrained experimental setting than RS.

NV offers practical advantages too. Regarding the acquired data, NV can improve the reliability of the measured signal (O’Conner et al., 2017; Wang et al., 2017), avoids confounds tied to stimulus repetition in conventional tasks (Hasson et al., 2010), and captures signal variability that might be missed in task or RS paradigms (Cantlon & Li, 2013; Vanderwal et al., 2019). Regarding the subjects, advantages include improved engagement and compliance, which in turn allow for longer MRI measurements and thus more acquired data (Eickhoff et al., 2020). Head motion might be reduced too, and these benefits might become most apparent and useful in populations struggling the most with prolonged stillness and participation, such as children and psychiatric patients (Vanderwal et al., 2019; Eickhoff et al., 2020). NV might provide a valuable addition in efforts to study variability in typical and atypical brain function (Dziobek et al., 2006; Hasson et al., 2010; Salmi et al., 2013; Bryge et al., 2015; Simony & Chang, 2020).

Variability can also be found in the analytical approaches that can be applied to NV data. Although some traditional analytical methods designed for parametric data are limited in their application to the complex and dynamic NV data, they can sometimes still be used in conjunction with annotations of the naturalistic stimuli (Sonkusare et al., 2019). Moreso, analyses commonly applied to RS data, especially functional connectivity (FC) analyses, lend themselves to study the functional architecture of the brain under naturalistic stimulation (Vanderwal et al., 2019). Functional connectivity describes the correlation between the activity time courses of two distinct regions in the brain, describing their co-activation which is commonly interpreted as joint involvement in processing and functioning. Distributed but interacting brain regions are often viewed as the foundation of cognitive functions (Eickhoff & Grefkes, 2011) and constituting functional networks supporting these functions (e.g. Fox et al., 2005; Power et al., 2011; Smith et al., 2009; Yeo et al., 2011; Schaefer et al., 2018, Pervaiz et al., 2020). Given that movies are complex and multimodal stimuli that elicit widely distributed brain activity, a network perspective might help to untangle the effect of movies on brain functional architecture, specific cognitive functions and behavior. Moreover, movies can be used to dissociate between cognitive processes that are shared by all people, by specific sub-groups only, and those that are unique to an individual (Hasson et al., 2009). Here, inter-subject correlation analyses might be especially useful to disentangle factors shaping the brain responses to movies that are consistent across subjects or idiosyncratic to individuals (Nastase et al., 2019).

While the complexity and multidimensionality of movie stimuli might complicate the differentiation of different variables driving these brain responses, various approaches can help investigate these variables, such as manipulating stimuli (Hasson et al., 2008), modeling stimuli (Bartels & Zeki, 2004), reverse correlation (Hasson et al., 2004) and combination with behavioral protocols (Hasson et al., 2008). Movies are inherently variable stimuli that can be varied in length (Nguyen et al., 2016; Kauttonen et al., 2018), presented outside the MRI to build up the storyline in a more comfortable environment (Jääskeläinen et al., 2008; Kauppi et al., 2010) or presented only in relevant outtakes (Viinikainen et al., 2012; Ren et al., 2018). The content and features of the movies can be chosen, or even designed, for best studying the respective research interest (Vanderwal et al., 2015). Additionally, movies can be characterized and annotated in a plethora of ways (Moon & Lee, 2017, Saarimäki et al., 2021). However, the precise and robust testing of movies as naturalistic stimuli to study functional brain architecture, individual variability and brain-behavior relationships requires large-scale datasets and investigations of various types of datasets (Eickhoff et al., 2020). In light of the effort and cost of such studies regarding finances, time and expertise, calls for collaboration and data sharing have been increasing (Liu et al., 2014; Moon & Lee, 2017; Vanderwal et al., 2019; Eickhoff et al., 2020). In this thesis, I will present a dataset created in answer to this call, extending existing datasets with a unique combination of movie fMRI, extensive socio-affective phenotyping of the participants, and annotations of the movie stimuli.

1.1 Naturalistic viewing and individual differences research

Research on NV started with questions about the differences between subjects' responses to more complex, ecologically valid stimuli: "Do we all see the world in the same way?" was asked in Hasson's pioneering work (Hasson, 2004). Exploring individual differences is fundamental to NV, and has been a focus of many studies.

At the forefront have been studies utilizing inter-subject correlation (ISC) analysis, which correlates the brain activity time-courses induced by naturalistic stimuli across subjects (Hasson, 2004). If the activity time-course of a brain region is highly consistent across subjects, these regions are presumably strongly involved in processing of the stimulus. If, on the other hand, a brain region's activity time-course is weakly correlated across subjects, it may be more involved in processing specific to each subject or encode little information about the stimulus (Nastase et al., 2019). Critically, ISC relies on the time-locked character of naturalistic stimuli - while the stimulus is complex, all subjects still receive the same stimulus at the same time, making comparisons across subjects possible. Altogether, ISC analyses enable the differentiation of variability in the brain's activity into idiosyncratic, i.e. subject-specific, components, and components that are shared across subjects (Nastase et al., 2019). In addition, intra-subject correlation analyses allow for investigating the effects different movies or parts of movies can have on each individual.

In combination with movie viewing, ISC analyses could show widespread synchronization of brain activity across subjects (Hasson et al., 2004), that depends in extent on the particular movie, was shown to be time-locked to the stimulus and induced by a specific series of events in the movie (Hasson et al., 2008; Lerner et al., 2011). Specifically, a conventional Hollywood movie induced higher similarity between subjects than unedited movies (Hasson et al., 2010). Similarity between and within subjects was also influenced by the coherence of a movie, as backward presentation of a movie reduces these correlations (Hasson et al., 2010). ISC has been related brain activity under NV to traits and behaviors such as different thinking styles (Bacha-Trams et al., 2018), perspective taking (Lahnakoski et al., 2014; Bacha-Trams et al., 2017; Bacha-Trams et al., 2020), humor (Jääskeläinen et al., 2016) and paranoia (Finn et al., 2018). Each of these studies employed movie stimuli specifically geared to the trait or behavior of interest, for example, a movie allowing the viewer to take different perspectives on a moral issue, comedy clips to induce amusement, or a suspicious social situation. This amplifies that variability induced between and within subjects is dependent on the naturalistic stimulus, or rather: studying naturalistic stimuli can help identify the factors that drive inter-individual differences.

Another approach to studying individual differences using naturalistic stimuli is that of (inter-subject) representational similarity analysis ((IS-)RSA). RSA allows the investigation of brain-behavior relationships by relating variability in brain measures to variability in behavioral measures (Kriegeskorte et al., 2008). In NV, this may help establish relationships between variability in stimulus properties, brain responses and states or traits of the subjects (Finn et al., 2020). Using IS-RSA, Finn et al., 2020, could provide evidence that brain responses during naturalistic viewing might depend on item-level responses to personality instruments more than on global scores. This result was mirrored by a study using IS-RSA to explore the relationships between variability in activity in brain networks when watching erotic movies and variability in personal preferences in sociosexual desire and self-control (Chen et al., 2020). Here again multivariate representations of the behavioral variables, but not summary scores, showed relationships with network activity. Moreover, only the erotic movie stimulus, not a neutral one, could reveal these relationships, highlighting the importance of stimulus choice. In a different study, IS-RSA could reveal an association between individual brain state dynamics during movie watching and subjective ratings of the stimulus (van der Meer et al., 2020).

When exploring brain-behavior relationships, FC during NV yielded better prediction of both cognitive and emotional traits than RS (Finn & Bandettini, 2021). Relatively little data (from under three minutes of movie watching) was needed for accurate predictions, and movies with social content gave most accurate predictions (Finn & Bandettini, 2021). Taken together, movies might enhance individual differences in behaviourally relevant brain networks, which is in line with theoretical considerations arguing that movie watching might pose an optimal condition for individual differences research based on its beneficial effects on inter- and intra-subject variability (Finn et al., 2017).

1.2 Naturalistic viewing and functional networks

The functional interaction between brain areas has been recognized as the fundament for cognition and behavior (Fox et al., 2005; Sporns, 2014; Mišić and Sporns, 2016). In light of the complexity and multimodality of naturalistic stimuli, a network perspective allows for studying the widespread effects movies have on the brain while still accounting for the knowledge gathered about brain functional architecture a priori.

Effects of movies on functional connectivity in brain networks have been studied intensely in recent years. Studies have shown that watching movies leads to functional reorganization of the brain (Wolf et al., 2010; Simony et al., 2016; Kim et al., 2018; Gilson et al., 2018; Demirtaş et al.,, 2019), and modulation of FC in frontoparietal, default and dorsal attention networks previously established in RS (Betti et al., 2013; Gao and Lin, 2012; Vanderwal et al., 2015).

To study individual differences, these differences must be measured reliably, prompting the study of effects of movies on FC over multiple sessions or within sessions across time. Andric et al., 2016, showed that brain activity reorganizes when subjects watch a movie for a second time, and that these changes to network-level connectivity occur at multiple spatial scales. Specifically, more local connections involved in sensory processing maintained similar activation profiles, while networks reconfigured on the whole-brain level during a second viewing. Gao et al., 2020, likewise used a test-retest approach to study how reliably movies, RS, and tasks could capture inter-individual differences. They found that movies outperformed RS and performed comparable to established questionnaires. Additionally, Gao et al. could show that the brain regions most reliably capturing inter-individual differences under naturalistic stimulation retained their reliability across different movies with socio-affective content, highlighting the relevance of emotional content for detecting individual differences. When studying intra-session reliability of FC during movie watching, Hlinka et al., 2022, noted that the changing content of movies results in lower split half reliability than RS. To account for the progression of the movie, analyses on different temporal scales might be beneficial. For example, time-varying analysis of networks under naturalistic stimulation could link fluctuations in network configuration over time to features of the movie stimuli and inter-subject similarity (Betzel et al., 2019). Likewise, dynamic FC can be used to study how individuality is manifested in dynamic reconfiguration of large-scale brain networks in response to NV, and how arousal modulates inter–subject variability (Jang et al., 2017).

Network-based analyses have advanced the understanding of coordinated patterns of brain activity (Hasson, 2004), and yielded research on brain states, which describe patterns of brain activity related to distinct cognitive processes or behavioral states. Van der Meer et al., 2020, investigated brain states and brain state transitions induced by movie watching. They found that brain states differed between movie watching and RS, and that movie watching was associated with more transitions between more brain states. Movie watching was associated with greater consistency in brain states and between sessions than RS. Co-occurrence of brain states across subjects was highest during specific movie events, but generally low in RS. Brain states could be associated with distinct functional signatures, physiological signals (heart rate, pupil diameter) and movie-induced arousal, and changes in brain states were linked to the progression of the movie. Taken together, brain states can be influenced by and reflect the content and progression of a movie, physiological changes induced by the movie, and the subject's perception of the movie. This highlights how movies create meaningful links between brain dynamics and cognition, emotion, and behavior.

At the same time, this study is another reminder for the relevance of the stimulus. Tian et al., 2021, investigated the consistency of FC across different movies. They could show that the intra-class correlation of FC under naturalistic stimulation was comparable to RS FC, that identification of subjects based on FC from different movies was possible with high accuracies, and that models predicting FC trained on one movie could generalize well to other movies. These three approaches indicate that FC during movie-watching is consistent across different stimuli. However, there might be more nuance to the reliability of FC under naturalistic stimulation. A different study showed that move-watching could improve identification of subjects based on FC in some of the investigated networks, but not all (Kröll et al., 2023). The same study indicates that content and features of the stimulus, such as the presence of a narrative, influence differences between movie-watching and RS FC, such as higher inter-subject variability under naturalistic stimulation. Overall, results were linked to the specific network-stimulus combination, further emphasizing the importance of precise coordination of stimulus, brain measure and research question.

1.3 Naturalistic viewing and socio-affective research

Choosing the movie stimulus is an important step in creating naturalistic viewing paradigms, and inevitably leads to the question whether movies are inherently well suited to study particular research topics. While it was shown before that naturalistic viewing can be and has been applied to a plethora of objectives, one might lend itself especially well to movie watching: socio-affective research.

Emotions are a key ingredient in narratives, driving the involvement and attention of the viewers (Aldama, 2015). Conversely, movies are useful to induce emotions (Westermann et al., 1996). Movies have been used to elicit basic positive and negative emotions (Gross & Levenson, 1995; Schaefer et al., 2010), to investigate commonly co-occurring emotions (Gilman et al., 2017), and mixed emotions (Kreibig et al., 2013; Kreibig et al., 2015; Samson et al., 2016; Kreibig & Gross, 2017; Aaron et al., 2018), to study emotion induction across different languages (Hagemann et al., 1999) and age groups (Fernández-Aguilar et al., 2018). The advantage of movies is that emotions here are presented embedded in context and progressing over time, which allows for investigating multiple emotion components simultaneously, how they develop and interact dynamically (Goldin et al., 2005; Sonkusare et al., 2019; Saarimäki, 2021; Lettieri et al., 2022). Emotion profiles and affective states induced by movies can be studied in relation to the brain states they are associated with (Lettieri et al., 2019), as well as the effects of movie stimuli on emotion-related brain networks across time (Nummenmaa et al., 2012). Movies show emotions as they occur, develop and dwindle in real-life situations, giving viewers the chance to empathize with the characters experiencing the emotions.

Emotions are an important factor in social relationships and interactions, which are a main focus of most conventional movies. As the viewers follow the stories of the characters, they have a perfect situation to apply and train their social and theory of mind (ToM) skills, that is, skills that allow us to understand what other people think, feel, what their intentions are and why they behave the way they do. Movies are particularly useful as "trainings" of ToM, because they let the viewer make predictions about the characters' thoughts, feelings, and intentions, with a much quicker resolution than typically occurring in real life (Cutting, 2016; Cutting & Armstrong, 2018). Movies might help develop ToM skills in children (Mar et al., 2010), and might be designed and edited in ways that encourage engagement of ToM at particular points in the movies (Cutting & Armstrong, 2018). Investigating ToM using movies indicates that a broader brain network is involved than previously assumed, as the multimodality and continuous dynamics of movies offer different contexts than more artificial stimuli (Wolf et al., 2010). The benefits of movies might be particularly useful when studying atypical social processing, for example in the autism spectrum (Bryge et al., 2015; Redcay & Moraczewski, 2019). However, movies are limited in their direct involvement of the subject: while more captivating and nuanced than most artificial stimuli, they mostly still pose the subject as the observing viewer, not an active participant of the presented social situation. Generalizing results from movie watching studies needs to be done with caution, and more advancement in social neurosciences is still needed (Nummenmaa et al., 2018; Redcay & Schilbach, 2019).

1.4 Naturalistic movie stimuli and NV paradigms

Although the term "naturalistic" has been firmly established to describe stimuli such as movies for a while (Hasson et al., 2010), it can be argued that most movies used in NV paradigms are not completely natural, i.e. situations occurring exactly as they would in real life (Vanderwal et al., 2019). Conventional movies are carefully designed to tell a story: the writing, directing, acting and editing are all orchestrated to convey the filmmaker's vision to the viewers. Across many studies, Cutting and colleagues investigated features of conventional movies, such as shot durations (Cutting et al., 2010; Cutting et al., 2011b); shot transitions (Cutting et al.,, 2011b; Cutting et al., 2012); motion (Cutting et al., 2012; Cutting et al., 2011a); luminance and music (Cutting et al., 2012); shot types (Cutting & Candan, 2015); and narrative shifts from one scene to the next

(Cutting, 2014; Cutting & Iricinschi, 2015). Gathering evidence from these studies and more, Cutting formulated a theory of the narrative structures of popular movies (Cutting, 2016). Based on movies from the past 70 years, consistency in the sequential structure of the narration can be found, realized through a four act structure and physical cues from five dimensions of filmmaking (editing, framing, motion, sound, and lighting). Cutting draws multiple insights highly relevant to NV from this: first, that the narrative and the narration of a movie are intrinsically linked - there is relation between form and meaning, and therefore, variations in the features and physical properties of a movie can be related to its content and vice versa. This offers insights into which features of a movie need to be characterized to make meaningful assumptions about its progression and content. Second, the familiarity of the relationship between patterns in the narration and content of the narrative enables rapid processing of the stimulus in the viewers. As popular movies overwhelmingly follow these narrative structures, viewers implicitly know what to expect from which cues and can easily understand what is happening. Based on this established formula, changes in physical features of a movie can drive emotion, attention and physiological responses reliably across viewers.

The effects of the filmmaker's choices and their effects on the viewer have been discussed and investigated frequently. A director may use tools such as setting, costumes, lighting and the staging of action ("mise-en-scène", the construction of the image) to guide the viewers' attention and elicit emotions (Bordwell & Thompson, 2010; Baranowski & Hecht, 2017). This synchronization of viewers' attention was shown through eye gaze analysis in many movies (Goldstein et al., 2007; Hasson et al., 2008; Sawahata et al., 2008; Hart et al., 2009; Marchant et al., 2009; Nyström & Holmqvist, 2010; Dorr et al., 2010; Smith & Mital, 2013). Relative motion (e.g. a small area of the screen moving with respect to a static background) and faces (especially close medium shots from the chest upwards have been good predictor the the viewers' gaze direction (Berg et al., 2009; Vig et al., 2009; Mital et al., 2011; Dayan et al., 2018). Synchronization between brain activity across viewers also gives indication to processes guiding their attention (Hasson et al., 2008) and features and moments in a movie that are particularly good at synchronizing subjects, as, for example, emotionally evocative narrative events (Chang et al., 2021). Studying these stimulus features is important, so that they can be modeled accurately when combining them with brain imaging data (Saarimäki, 2021) and that all variables driving brain responses can be accounted for (Hasson et al., 2010).

Complex naturalistic stimuli allow studying perception, cognition and emotion in more ecologically valid environments, but this comes at the cost of distinctness in the paradigm. After all, many processes interleave and interact when engaging with the natural world. For example, it has been argued that naturalistic paradigms designed to target ToM and social skills arguably involve executive functioning as well (Baron-Cohen et al., 1997; Heavey et al., 2000; Dziobek et al., 2006). Therefore it might be necessary to characterize the (brain, cognitive and emotional) states induced by naturalistic paradigms and study if and how a movie can be regarded as the sum of its task-parts, or more than that (Eickhoff, Milham & Vanderwal, 2020). In regards to clinical applications of movie watching, the “golden triangle” approach was proposed by Eickhoff, Milham & Vanderwal, 2020, where brain measure “A” during movie “B” could identify the risk for disorder or symptom “C”. Complementing, the ideal choice of stimulus for detecting individual differences may depend on the specific behavior of interest (Finn et al., 2017; Finn et al., 2020). For example, negative movie clips were associated with greater differences between control subjects and subjects with melancholic depression than positive movie clips (Guo et al., 2015), and ambiguous movie stimuli might be beneficial for studying emotional experiences in alexithymia (Aaron et al., 2018). The key role the NV stimulus and its makeup plays in the NV paradigm warrants a review on what stimuli have been used in NV research so far.

As movies have been used for various purposes, the specific stimuli used changed accordingly. Movies used for emotion induction are usually short clips to induce basic emotions (Gross & Levenson, 1995; Schaefer et al., 2010; Jenkins & Andrews, 2012), dimensional emotion responses (Soleymani et al., 2009; Soleymani et al., 2012; Carvalho et al., 2012; Gilman et al., 2017) or mixed emotional states (Schaefer et al., 2010; Samson et al., 2015). Most of these clips are under one minute in duration, with only some lasting a few minutes. Similarly short are movies used in affective video content analysis and human-computer interaction studies where the aims are to summarize, segment or index videos or create personalized recommendations for viewers based on the affective content of the movies (Douglas-Cowie et al., 2007; Koelstra et al., 2012; Baveye et al, 2015; Moon & Lee, 2017; Kim et al., 2018). With the advent of naturalistic viewing for individual differences and socio-affective research, studies started employing longer stimuli or multiple sessions (Jääskeläinen et al., 2008; Vanderwal et al., 2015; Alexander et al., 2017; Nguyen et al., 2017; Vanderwal et al., 2017; Visconti di Oleggio Castellano et al., 2020; Kröll et al., 2023), and even showing complete or minimally shortened conventional movies (Hanke et al., 2014; Kauttonen et al., 2018). The duration of movie stimuli is a relevant aspect, as some cognitive and emotional processing only evolves over longer time frames (Hasson et al., 2010). Context is essential when understanding social situations and identifying or emphasizing with different characters. Arguably, more complex socio-affective processes can only be studied if the stimulus represents the full complexity of social relationships and emotions. However, stimulus duration is just one aspect out of many that can characterize a movie, and over time more efforts have been made to better profile and understand the stimuli used.

1.5 Annotations of movie stimuli & phenotyping participants

As described above, Cutting and colleagues made a concerted effort to study various features making up conventional movies in general. Some researchers have likewise invested into studying the specific movie stimuli used in NV paradigms. In that venture, annotations are an indispensable tool for characterizing movie stimuli. An annotation here means any organized description or documentation of the expression of certain features in a movie. The *studyforrest* project (Hanke et al., 2014), in which participants watched the movie Forrest Gump, later extended their dataset with annotation of cuts, depicted location and temporal progression (Häusler & Hanke, 2016), speech (Häusler & Hanke, 2021), portrayed emotions (Labs et al., 2015), semantic conflict (Hanke & Ibe, 2016), body contact, eye movement, music, and low-level perceptual confounds. Lahnakoski et al., 2012, assessed selected movie clips on social (faces, bodies, biological motion, goal-oriented action, emotion, social interaction, pain, and speech) and non-social (houses, objects, rigid motion, non-goal-oriented action, humans not participating in social interaction, and non-human sounds) features.

Machine-based and human annotations have been used to extract different semantic and physical features of movies (McNamara et al., 2017; Saarimäki, 2021), but the main focus has been on assessing emotional content in the movie or emotional responses from the movie viewers (for review and discussion of annotation methods used for affective research, see Saarimäki, 2021). Emotions in regards to movies can be seen from different perspectives: emotions the audience perceives or experiences, emotions expressed by the actors, and emotions intended by the director to be induced (Tian et al., 2017). Because most of the current application of NV are centered on the effects the stimuli have on the subjects, it might be most meaningful to focus on the audience's perceptions of and responses to the movie stimuli (Wang & Cheong, 2006).

Here it is relevant to distinguish between the emotions portrayed in the movie and those induced in the subjects, which may not correspond completely to one another (Bordwell & Thompson, 2008; Labs et al., 2015; Knautz & Stock, 2011; Tian et al., 2017). The emotions that are portrayed in the movie are often projected by low-level features like the soundtrack (Tarvainen et al., 2014), projection into the image (e.g. rain symbolizing sadness), particular camera settings and movement (e.g. a high camera angle to indicate helplessness) and emotion expression of the actors (Baranowski & Hecht, 2017). These depicted emotions usually elicit empathy and sympathy, but subjects might also feel emotions different from those expressed by the characters. For example, a character might enjoy cruelty while the viewers feel disgusted by it (Aldama, 2015). Similarly, amusement might be induced in viewers by a character's stoic, emotionless appearance in the face of chaos (Aldama, 2015). Viewers might enjoy horror movies (Hoffner & Levine, 2005) and tragedies (Hick & Derksen, 2017), or experience schadenfreude regarding a character’s misfortune (Van Dijk and Ouwerkerk, 2014). Therefore, characterizing the stimulus might not be not sufficient for inferring which emotions are induced in the viewer (Saarimäki, 2021).

There are two options for quantifying the emotions elicited in viewers: implicitly, through measuring their responses, and explicitly, by asking them about their experience (Moon & Lee, 2017). Physiological changes are inherent parts of emotional experiences (Chritchley & Harrison, 2013), and movies have been shown to elicit strong subjective and physiological changes (Frazier et al., 2004; Gross, 1998; Palomba et al., 2000). Studies have measured various peripheral physiological responses during movie watching, such as heart rate, respiration, pupil diameter, skin conductance, eye movements and more (Soleymani et al., 2009; Oliveira et al., 2013; Moon & Lee, 2017; van der Meer et al., 2020). The combination of these measures and brain responses (measured for example with electroencephalography (EEG), magnetoencephalography (EEG), fMRI or functional near-infrared spectroscopy (fNIRS)) during NV allows for studying brain-body interactions and interoception in a ecologically valid framework (Sonkusare, 2019).

Asking participants about the experienced emotions might seem like the most direct method to assess these emotions, but it comes with certain difficulties (for a review on behavioral experience sampling methods in naturalistic neuroimaging studies, see Jääskeläinen et al., 2022). Post-hoc assessment of induced emotions might not be an accurate representation of the viewers’ experiences across the movie’s progression, as they are not simple averages of the emotional fluctuations experienced throughout movie watching (Metallinou & Narayanan, 2013). But self-reports during movie watching might alter the experience and disrupt immersion, especially in an MRI experiment. To assess the emotions induced by movie stimuli more robustly, crowdsourcing emotion annotations outside of the neuroimaging measurement is a good alternative (Knautz & Stock, 2011; Tian et al., 2017). Online implementations of emotion annotations might make it easier to achieve big samples of viewers, but copyright concerns might limit this option when stimuli are created from conventional movies. However, consistent emotion annotations could previously be assessed with less than 100 subjects as well (Knautz & Stock, 2011).

Assessing the emotions induced by a movie stimulus is on the verge between characterizing the stimulus and characterizing the subject. While crowdsourcing annotations might give good estimations of the stimulus capabilities to induce certain emotions and emotion profiles, it has little regard for the individuality that might introduce differences to subjects' appraisal of the stimulus. However, these differences are of pivotal interest in individual differences research. Therefore, characterizing the subjects is critical in this application of NV.

Some larger research ventures have combined phenotyping of subjects and NV, such as the Human Connectome Project (HCP, Van Essen et al., 2013), the Healthy Brain Network biobank (HBN, Alexander et al., 2017) and the Cambridge Centre for Ageing and Neuroscience study (Cam-CAN, Shafto et al., 2014). Phenotyping might include cognitive, emotional, social, demographic and other assessments of the subjects. Cognitive and emotional traits as well as personality seem to modulate brain activity during NV (Finn et al., 2018; Finn & Bandettini, 2020). Altogether, extensive phenotyping of subjects is necessary to fully understand the relationships between movies, brain and behavior.

1.6 Aims of the current work

Above I have outlined the inherent aptitude of NV for individual differences and socio-affective research, as well as the need for large-scale datasets with extensive characterizations of movie stimuli and subjects. In this thesis, I will present one project focusing on the influence of movie watching on emotions and inter-individual differences in network functional connectivity (NFC) in a publicly available dataset. Then, I will present a comprehensive NV dataset which was acquired to be suitable to explore individual differences and brain-behavior relationships between neuroimaging data under naturalistic stimulation and socio-affective phenotypes. This dataset has been extended in two studies, one assessing annotation of emotions induced by the movie stimuli and one assessing features of the movie stimuli. Lastly, I will give an example of how this dataset can be used to study the covariation of socioaffective phenotypes with network functional connectivity derived from imaging data under naturalistic stimulation.

1.6.1 Inter- and intra-subject variability in NFC across a full narrative movie

Inter- and intra-subject variability are important factors in detecting individual differences (Finn et al., 2017). With respect to movie fMRI, Vanderwal and colleagues (2017) showed that the effect of movies is differentially distributed across the brain, with lower inter- than intra-subject variability in unimodal regions and higher inter- than intra-subject variability in heteromodal regions. However, a concrete comparison of these variabilities on the level of NFC has rarely been done (but see Kröll et al., 2023). Furthermore, it is yet unclear which features of a movie stimulus influence inter- and intra-subject variability. Emotions might play an important role in affecting similarity within and between subjects and might be best studied in a full narrative movie, which combines a long, overarching narrative and numerous distinct, individual scenes. The changing content and wide range of emotional scenes present within this narrative warrants a more detailed characterisation of the stimulus and its segments.

To fill this gap, the first project of this thesis investigates inter- and intra-subject variability in NFC over the course of a full narrative movie. In a first step, different segments of the movie stimulus were compared with regard to their portrayed valence and arousal, evincing differences in emotional content. Using linear mixed models (LMM), we analyzed how different factors, such as the narrative of the movie and portrayed valence and arousal, affect inter- and intra-subject variability in NFC across 14 meta-analytically defined networks.

1.6.2 Movies Dataset

The second project sought to create an extensive dataset using NV for socio-affective and individual differences research. The main sample underwent fMRI measurements including two RS scans, an anatomical scan, and 8 different movie watching scans. Subjects were characterized regarding various socio-affective traits, personality, depression, anxiety, verbal fluency, and more using questionnaires and tasks. Subject phenotyping was complemented by demographic questionnaires and hormone assessment. Mood, affective states and stress of the subjects were assessed directly before and after NV fMRI. Seven of the movies were scenes taken from conventional movies, intended to induce various emotions. The last movie was a compilation of shorter clips from conventional movies. In two additional studies, annotations of the movie stimuli were acquired. The first annotation assessed the emotions induced by the movies in a separate sample. In the second annotation, two raters assessed various social and non-social features of each movie stimulus.

The data presented here are a subset of the original dataset, as acquisition is still ongoing at this point in time to create a truly large-scale dataset. At a later point, the dataset will be made publicly available so that its richness can be used by many researchers and applied to many research questions.

1.6.3 Covariation of state & trait anxiety with NFC under naturalistic stimulation

The third project explores the covariation of state and trait anxiety with NFC in a subset of the previously acquired Movies dataset using representational similarity analysis (RSA). It is an example of how the dataset can be used to investigate individual differences, and gives insight into how variability in socio-affective phenotypes on the state and trait level might covary with variability in NFC induced by different movie stimuli.

2. Materials and Methods

2.1 Inter- and intra-subject variability in NFC across a full narrative movie

2.1.1 Sample

The sample consisted of 15 native German-speaking participants (6 females, range 21-39 years) (Hanke et al., 2016). One subject, which we excluded, was an outlier in the intra-subject correlation analysis, leading to a sample size of 14 (6 females, age range 21-39 years. Please note that mean age cannot be reported, because only age ranges were reported for each participant). The Ethics committee of Otto-Von-Guericke University, Germany approved acquisition of the data in the *studyforrest* project. For a more detailed description of the sample, see Sengupta et al., 2016. The full dataset can be found under: https://github.com/psychoinformatics-de/studyforrest-data-phase2. A list of subjects can be found in the supplementary table S1.

2.1.2 fMRI data acquisition

fMRI data acquisition took place in a single session which included a short break in the middle. To keep the stimulus at a length of two hours, some scenes were cut. The movie stimulus represents a full narrative movie, as only scenes that were less relevant to the plot were cut, thus preserving the overarching story. For the purpose of data acquisition, the movie stimulus was separated into 8 segments of approximately 15 minutes each, taking scene boundaries into consideration. This lead to an unequal number of volumes acquired per segment, which were 451, 441, 438, 488, 462, 439, 542, and 338 for segments 1 – 8 respectively (see Hanke et al., 2014 for details and code on movie segment creation). The movie segments were shown in chronological order. For each segment, T2*-weighted echo-planar images (gradient-echo, 2 s repetition time (TR), 30 ms echo time (TE), 90° flip angle, 1943 Hz/Px bandwidth, parallel acquisition with sensitivity encoding (SENSE) reduction factor 2, 35 axial slices, 3.0mm slice thickness, 80 x 80 voxels (3.0 x 3.0mm) in-plane resolution, 240 x 240 mm field-of-view, anterior-to-posterior phase encoding direction in ascending order, 10% inter-slice gap, whole-brain coverage) were acquired using a whole-body 3 Tesla Philips Achieva dStream MRI scanner equipped with a 32 channel head coil.

All downloaded data were minimally preprocessed as described in Hanke et al. (2016). In short, preprocessing steps included defacing, motion correction, reslicing and data interpolation using in-house codes that utilize the FSL toolkit. All codes are openly available under: https://github.com/psychoinformatics-de/studyforrest-data-aligned/tree/master/code. For this study, the native fMRI data were brought into MNI space using FSL's applywarp function for subsequent NFC extraction.

2.1.3 Portrayed valence and arousal

To characterize the movie segments with regards to the portrayed valence and arousal, we used the openly available data from Labs et al. (2019). This dataset contains annotations of portrayed emotions in the "Forrest Gump" movie stimulus used in the *studyforrest* dataset.

A group of observers (n=9, German female university students) were asked to evaluate scenes of the movie in terms of valence ("positive" or "negative") and arousal ("high" or "low") portrayed by the movie characters. All scenes were presented in random order to allow observers to focus on current indicators of portrayed emotions without being influenced by, for example, the conveyed mood of the movie plot. To evaluate the consistency of evaluations between observers, Labs and colleagues calculated the inter-observer agreement (IOA). The IOA is expressed as a value between 1 and -1, with "1" indicating perfect observer agreement regarding positive valence (or high arousal, respectively) and "-1" indicating perfect observer agreement regarding negative valence (or low arousal, respectively). IOAs were reported as a time series of the movie in correct order downsampled to 2 seconds, corresponding to the sampling rate of the fMRI data. We used code published by Lettieri et al. 2019 (https://osf.io/tzpdf/) to divide the IOA time series according to 8 movie segments for subsequent analyses.

2.1.4 Inter- and intra-subject variability in NFC

To investigate effects on a network level, we used 14 networks defined as sets of peak coordinates in different meta-analyses. These included the autobiographical memory (AM) network (Spreng et al., 2008), cognitive attention control (CogAC) network (Cieslik et al., 2015), extended multiple demand network (eMDN) (Camilleri et al., 2018), emotional scene and face processing (EmoSF) network (Sabatinelli et al., 2011), empathy network (Bzdok et al., 2012), theory of mind (ToM) network (Bzdok et al., 2012), emotion regulation (ER) network (Buhle et al., 2014), extended socio-affective default network (eSAD) (Amft et al., 2015), mirror neuron system (MNS) network (Caspers et al., 2010), motor network (Witt, Meyerand, & Laird, 2008), reward (Rew) network (Liu et al., 2011), semantic memory (SM) network (Binder et al., 2009), vigilant attention (VigAtt) network (Langner & Eickhoff, 2013), and the working memory (WM) network (Rottschy et al., 2012). A more detailed description of these networks are reported in the supplements (supplementary material S2). For each meta-analytical network, nodes were created by placing 6mm spheres around the peak coordinates (see supplementary material S3 for an overview of the peak coordinates and S4 for a figure of all nodes of all networks). The functional connectome of a given network was created using in-house MATLAB R2017a (The Mathworks Inc., 2017) code which computes the pairwise Pearson correlation between all nodes for each segment and each participant. This resulted in 1680 functional network connectomes (15 participants x 8 segments x 14 networks) saved as N-by-N matrices with N being the number of nodes.

To keep in line with previous studies (Vanderwal et al., 2015; Vanderwal et al., 2017; Finn et al., 2017; Nastase et al., 2019), we operationalized the inter- and intra-subject variability by assessing its inverse, i.e. the similarity of functional connectomes within and between subjects, quantified by Pearson correlation coefficients. Inter- and intra-subject similarity were computed per network, segment and subject as depicted in Figure 1. All computations are based on the unique connections between nodes (i.e. the lower triangle of the NFC matrix) and exclude all auto-correlations. For inter-subject similarity, we first computed the correlations between the NFC of one subject and all other subjects. After Fisher Z-transformation of the correlation coefficient, they were averaged and re-transformed, resulting in one value representing inter-subject similarity for the respective subject in the given segment and network. For calculating intra-subject similarity of a given subject and segment, we computed the correlations between NFCs of this segment and every other segment of the subject. After Fisher Z-transformation, we averaged these correlation values and re-transformed them, resulting in one value representing intra-subject similarity for the respective subject in the given segment and network. Both inter- and intra-subject variability were calculated based on Pearson correlation coefficients.

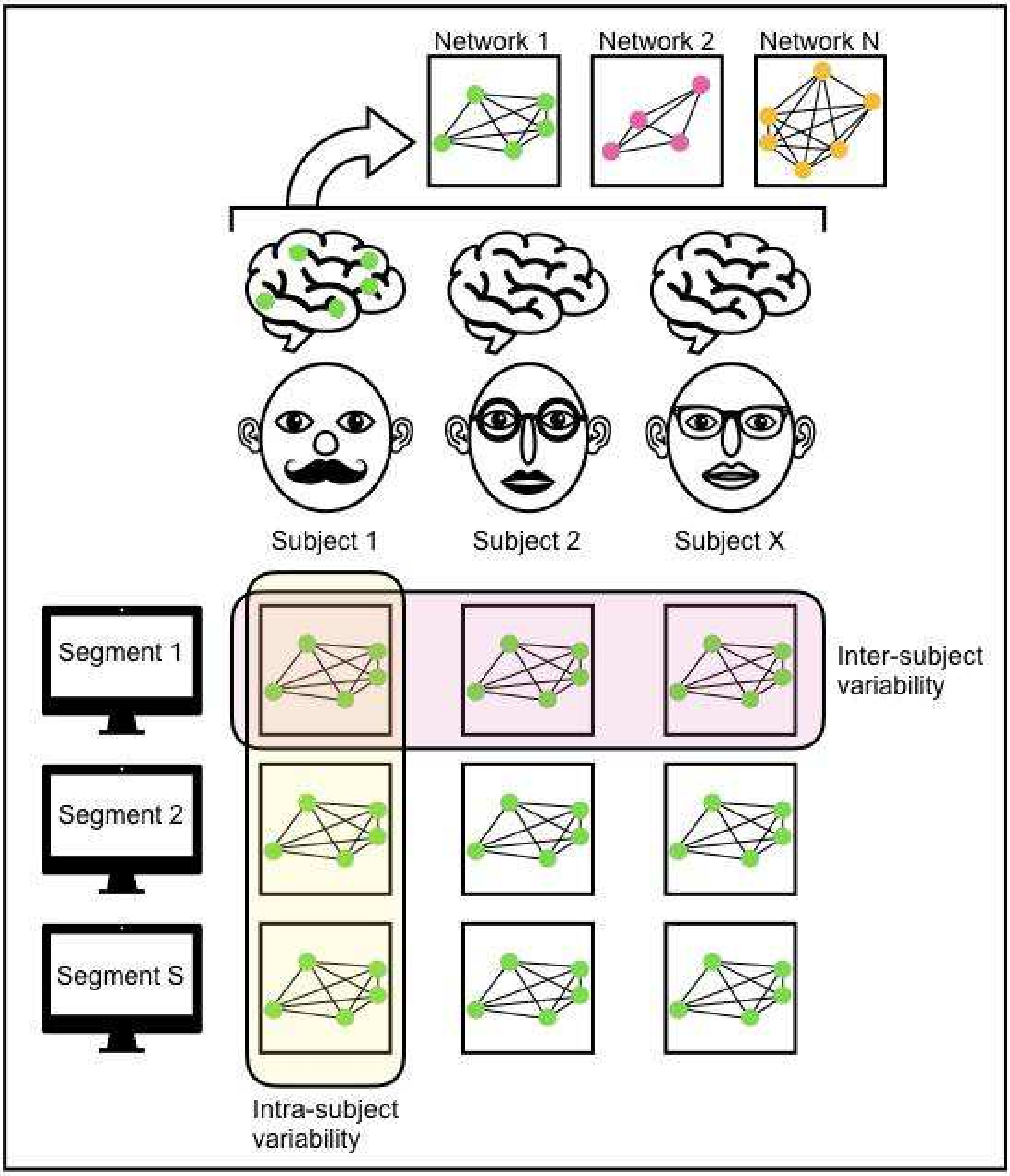


Figure 1: **Calculation of inter- and intra-subject variability.** For each subject, functional connectomes were computed for all 14 networks in each of the 8 movie segments. Inter-subject variability is assessed by calculating the average correlations between subjects within the same movie segments. Intra-subject variability is assessed by calculating the average correlation between movie segments within the same subject.

2.1.5 Statistical analyses

To investigate whether portrayed emotions are different across movie segments, we conducted a one-way ANOVA for each measure. Here, IOA values were used as independent variables with the movie segments as fixed factors. Post-hoc t-tests are reported with Bonferroni-adjusted p-values.

To test whether inter- or intra-subject variability differ across networks depending on movie segments and portrayed emotions, we applied linear mixed models (LMM,) using the statsmodels python package (https://www.statsmodels.org/stable/mixed_linear.html). Specifically, we created different random intercept models by choosing network, movie segment, arousal and valence as possible fixed effects, subject identity as a random effect, and inter- or intra-subject variability as the dependent variable. We chose subject identity as a random effect, because participants are the sampling unit of interest and contribute repeatedly to the NFC measures across all movie segments. We model individual differences by assuming different random intercepts for each subject, but no individual random slopes, because a simpler model structure was warranted by our data. Network was chosen as a fixed effect to test which networks are associated with changes in inter- or intra-subject variability induced by NV. It was included as a categorical factor with 14 levels. Movie segment was chosen as a fixed effect to test for an effect of the length and complexity of a full-narrative movie. Portrayed valence and arousal were chosen as fixed effects to represent the emotional content of the full-narrative movie, testing if emotions portrayed in a movie affect inter- or intra-subject variability in NFC. Models were generated using maximum likelihood to include all possible models, that is, each unique combination of one to four fixed effects and their respective interactions, resulting in 2128 models that were compared each for inter- and intra-subject variability. The model best fitting our data was selected using Bayesian information criterion (BIC, Schwarz, 1978) and used to calculate the parameter estimates for each effect. To test whether a specific network had a significant influence on inter- or intra-subject variability, we created a “mean network” representing the mean inter- or intra-subject variability values across all networks that we used as a reference category to compare all other networks against. P-values were obtained using Wald tests of the best models.

2.2 Movies Dataset

2.2.1 Sample

As of July 2023, we acquired data of 101 healthy native speakers of German, but will only report here data of the 84 subjects who completed the full study protocol (43 male, 41 female, age 18 - 35 years, mean age 24.33 years, SD 3.93 years). Exclusion criteria were any psychiatric or neurological disease, impaired vision (that cannot be corrected for using glasses/contact lenses) or hearing, alcohol, drug or medication addiction, usage of medication affecting the central nervous system within the last 3 months, and any MRI contraindication. All participants gave written informed consent. The study was conducted in accordance with the declaration of Helsinki and approved by the ethics committee of the Heinrich-Heine-Universität Düsseldorf (grant numbers "2018-100-Drittmittel", "2018-100_1-andere Forschung erstvotierend", "2018-100_2-andere Forschung erstvotierend" and "2018-100_3-andere Forschung erstvotierend").

2.2.2 Study protocol

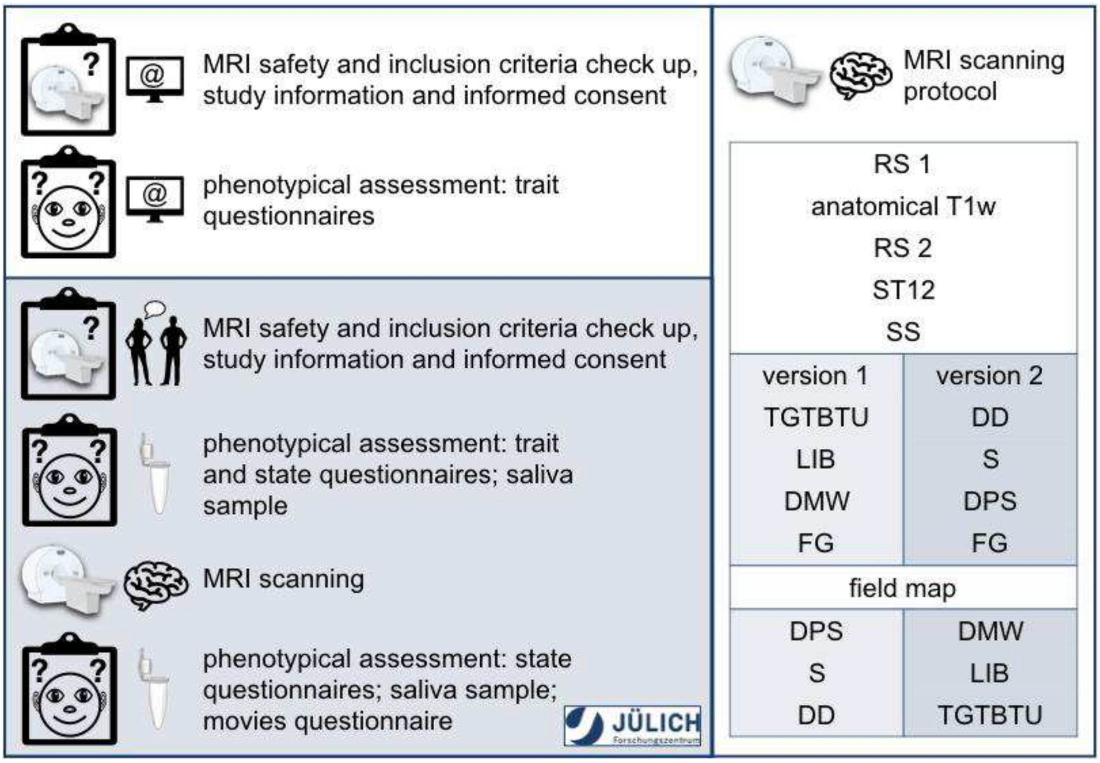


Figure 2. **Overview of study and MRI protocols.** The left side depicts the overall study protocol. Participants were first screened for inclusion criteria, if study participation was possible, participants next filled out online questionnaires assessing a range of traits. Then participants partook in the on site appointment including trait and state phenotypical assessment and MRI scanning. The right side depicts the MRI scanning protocol, including two versions of the movie watching protocol. RS = resting state, ST12 = short term 12, SS = short sequences, TGTBTU =

The Good, The Bad, The Ugly, LIB = Life is Beautiful, DMW = Dead Man Walking, FG = Forrest Gump, DPS = Dead Poets Society, S = Scream, DD = Dirty Dancing.

Participants were recruited through advertisements on social media, wanted ad websites, mailing lists, flyers distributed at public places in the surrounding area, and at public events representing the institute. The protocol of study participation and specifically the MRI scanning can be seen in Figure 2. Participants received an email with information about the study and inclusion criteria and were asked to fill out an online questionnaire screening for MRI safety inclusion criteria. If MRI safety was guaranteed and no other exclusion criteria reported back, participants were informed about their inclusion in their study and received consent forms and access to the first 13 online questionnaires assessing phenotypic traits. These questionnaires were the BDI, EHI, NEO-FFI, STAI, TAS-26, FEEL-E, SEE, EKF-S, ANPS, PANAS, STAXI, ISK-K, ADHD. All questionnaires are described in detail below in section 2.2.3. Participants were instructed to fill out the questionnaires at a time that works best for them and in any order, and report back once all questionnaires are completed. Afterwards, participants were invited to their on site measurement. Because of the hormone sampling using saliva, participants were asked not to eat or drink anything except water for an hour before their appointment. Upon arrival, participants were once again given study information and consent forms, including information, exclusion screening and consent forms to MRI scanning. Participants then concluded trait assessment by partaking in the computer-based tasks they could not complete online, that is, a theory of mind task and verbal fluency task. They also filled out questionnaires assessing demographic information, affective state, mood and task-related stress using state versions of the BDI, STAI, STAXI, PANAS, DSSQ and SAM questionnaires. Simultaneously, participants handed in their first saliva sample by filling a specific saliva tube. Participants were then prepared for MRI scanning. Details regarding the MRI procedures can be found in section 2.2.5. After MRI scanning, participants repeated some of the state questionnaires (PANAS, DSSQ), a questionnaire assessing their experience with the movie stimuli and general movie watching preferences and habits, donated their second saliva sample and completed their participation. Participants were reimbursed with 50 Euro for their time.

2.2.3 Phenotypic data acquisition

The goal in acquiring this dataset was to achieve a comprehensive assessment of socio-affective traits of the participants along naturalistic viewing fMRI data. We here focused on traits related to emotional processing and behavior, and social processing and behavior. To assess the expression of these traits, we used established questionnaires that were German or had German versions. For copyright reasons, the original items and instructions of the questionnaires are not included for all questionnaires. Precise psychometric properties, evaluation studies, and normative data can be found in the respective manuals referenced here. We assessed personality traits using the NEO-FFI and ANPS questionnaires. Social skills were measured using the ISK-K questionnaire and a theory of mind task. Emotion-related traits, such as emotion regulation, the experience of emotions and emotion skills, were assessed using the FEEL-E, SEE, EKF-S and STAXI questionnaires. In a more clinical context, depression, anxiety, ADHD and alexithymia were assessed using the BDI, STAI, ADHS-E and TAS-26, respectively. We assessed handedness using the EHI, and verbal fluency using a specific task for comparability to another study of our institute focusing on speech and executive functions (Camilleri et al., under revision). Affect and mood were measured using the PANAS and SAM, while study-related affect was measured using the DSSQ.

To complement the phenotypical characterization of the participants, we included state versions of some questionnaires to assess the current mental and emotional state of the participants (BDI, STAI, STAXI, PANAS, DSSQ). These tests were applied directly before and, in case of the PANAS and DSSQ, after MRI scanning. They give the option to probe the affective state of participants before and after the movie watching paradigm, potentially include these questionnaires as confounding factors, and study the relationship between naturalistic viewing data and state- or trait-level characteristics. An example of such a study is presented in project 3 which investigates the covariation in state and trait anxiety in relation to the covariation in functional network connectivity in naturalistic viewing fMRI data. The SAM questionnaire was additionally used to probe the participants' affective state after each movie stimulus during the MRI session. All questionnaires and tasks are introduced below.

2.2.3.1 Demographics

We created a questionnaire assessing a wide range of demographic information about our participants. This questionnaire was filled out on site before the MRI measurement. The demographics questionnaire asked for date of birth, sex and gender identity, age, sexual orientation, marital status, highest level of education, current profession, need for a visual aid (glasses or contact lenses), impaired vision, previous and current psychological, psychotherapeutic or neurological treatment including timeframe of the treatment and reason, current medication, languages spoken other than German and frequency of their use, hours slept last night and how usual that is, current tiredness, consumption of caffeinated drinks on that day and how usual that is, smoking, smoking on that day and how usual that is, alcohol consumption the previous night, alcohol consumption on that day and how usual that is, general well-being at that moment, current ailments, habitual use of computer screens, hormonal contraception, current day or quarter of the menstrual cycle, and pregnancy.

2.2.3.2 Beck's Depression Inventory (BDI)

The BDI (Beck et al., 1961) and later revised BDI-II (Beck et al., 1996) is a self-report questionnaire assessing the severity of depressive symptoms. We here used the German translation (Hautzinger et al., 2009). The revised BDI entails 21 questions querying the severity of symptoms that characterize depression according to the DSM IV. Items are rated on a 4-point Likert scale from 0 to 3 with higher numbers representing greater symptom severity. The revision of the questionnaire included an exchange of 4 items, an adaptation of 2 items and reformulation of most items, and a change of assessment period to two weeks compared to the original version. The English version of the BDI II is one of the most used assessments of depression since its publication and shows high internal consistency, good discrimination between depressed and non-depressed subjects, and good content and structural validity (Wang & Gorenstein, 2013). The German version was introduced 10 years later after thorough investigation of its psychometric properties in clinical and non-clinical samples (Hautzinger et al., 2009). The 21 items assess the following symptoms: sadness, pessimism, feelings of failure, loss of joy, guilt, feelings of being punished, self rejection, self criticism, suicidal thoughts, crying, restlessness, loss of interest, inability to make decisions, worthlessness, loss of energy, change in sleep, irritability, change of appetite, difficulty concentrating, tiredness, loss of sexual interest. For each item, the subject should evaluate how severely they experienced the respective symptom in the past two weeks. The BDI-II is evaluated by summing the values of the item responses, resulting in a score ranging from 0 to 63. The BDI is not intended to diagnose depression, but can be interpreted using cut-off values. The creation of these cut-off values by Beck et al., 1996, as a screening for depression focuses more on sensitivity than specificity, representing a lower threshold for depression. The cut-off values for the BDI-II scores are: 0 - 8 = no depression, 9 - 13 = minimal depression, 14 - 19: light depression, 20 - 28: moderate depression, 29 - 63: severe depression.

We used this questionnaire as a trait version, assessing the severity of symptoms generally (with no reference timeframe), and as a state version directly before MRI scanning, assessing the severity of symptoms on that day.

2.2.3.3 Edinburgh handedness inventory (EHI)

The EHI (Oldfield, 1971) is a self-report questionnaire assessing handedness. 12 items assess which hand is preferred to fulfill a task or use an object, with the items being:

writing, drawing, throwing, scissors, toothbrush, knife (without fork), spoon, broom (upper hand), striking match (match), opening box (lid), kicking (which foot), eye preference. The items are answered using a 5-point Likert scale with ranks “only left”, “left”, “indifferent”, “right” and “only right”. These ranks translate to the scores -2, -1, 0, 1, 2 when evaluating the test, which are then summed. A negative test score implies left-handedness while a positive test score implies right-handedness.

2.2.3.4 NEO Five-Factor Inventory (NEO-FFI)

The NEO-FFI was developed on the basis of the five factor model of personality, which organizes personality traits on five dimensions: extraversion, agreeableness, conscientiousness, neuroticism and openness to experience (Tupes & Christal, 1992; Norman, 1963). This theory of personality is historically based on lexical approaches and factorization studies investigating the underlying structure of personality traits, and has received a lot of empirical support (Goldberg, 1990; McCrae & John, 1992). The NEO-FFI is a 60-item self-report questionnaire featuring a list of statements describing a person. Each statement is ranked on a 5-point Likert scale (“strong disagreement”, “disagreement”, “neutral”, “agreement”, “strong agreement”) in regards to how well the respective statement describes the subject. The 60 items are divided into 5 subscales, each comprising 12 items and evaluating one of the 5 dimensions of personality. Results of the NEO-FFI are interpreted for each subscale separately. A revision of the NEO-FFI has been contemplated that includes a slightly different choice of items from the larger item pool of the more extensive NEO-PI-R (Costa & McCrae, 1992), but this version showed only modest improvements in reliability and factor structure, and equivalent validity, resulting in a recommendation to keep the older version in use (McCrae & Costa, 2004) and no changes made to the German version to keep comparability to previous work on psychometric properties of the NEO-FFI (Borkenau & Ostendorf, 2008).

2.2.3.5 State Trait Anxiety Inventory (STAI)

We used the German version (Laux et al., 1981) of the State-Trait-Anxiety Inventory (Spielberger et al., 1970). Fundamental to the STAI is the distinction between state and trait anxiety (Spielberger, 1972). Spielberger (1972) defines state anxiety as an emotional state marked by specific feelings (tension, worry, nervousness, inner restlessness, fear of future events) and increased activity of the autonomic nervous system. It varies in intensity across time and situations. Spielberger (1972) further defines trait anxiety as stable individual differences in the predisposition to assess situations as threatening and react with increasing state anxiety. The trait STAI consists of 20 items that are rated on a 4-point Likert scale assessing frequency (“almost never”, “sometimes”, “often”, “almost

always”). Participants are asked to respond how they are feeling generally. The state STAI also consists of 20 items that are rated on a 4-point Likert scale assessing intensity (“not at all”, “somewhat”, “moderately so”, “very much so”). Here, participants are asked to respond how they are feeling now, at this moment. Results from both scales are calculated as separate sum scores, respecting that some items need to be reversed before scoring. Therefore, participants can score between 20 and 80 on each scale. In the state scale, a score of 20 corresponds to the absence of feelings of anxiety and a score of 80 corresponds to the highest intensity of that feeling. The score of the trait scale reflects stable individual differences in anxiety disposition.
We used the trait scale of the STAI prior to the on site measurements where participants then responded to the state scale directly before MRI scanning.

2.2.3.6 Toronto Alexithymia Scale (TAS-26)

The TAS-26 is a self-report questionnaire assessing alexithymia, a disorder of affect regulation marked by the difficulty to identify and describe feelings and reduced creativity and fantasy (Taylor et al., 1985). It includes 26 items comprising 4 scales (difficulties identifying emotions (orig. “Schwierigkeiten bei der Identifikation von Gefühlen”), difficulties describing emotions (orig. “Schwierigkeiten bei der Beschreibung von Gefühlen”), externally oriented thinking (orig. “Extern orientierter Denkstil”), reduced daydreaming (orig. “Reduzierte Tagträume”)). This version has been revised into the TAS-20 (Bagby et al., 1994), but the respective German translation showed worse internal consistency and bad representation of the construct “externally oriented thinking” in the third scale (Bach et al., 1996), resulting in the continued use of the German version of the origins TAS-26. Nevertheless, the work on the revised version yielded evidence that the fourth scale, “reduced daydreaming”, was incompatible with the construct of alexithymia (Bagby et al., 1994), and is therefore excluded from scoring of the questionnaire in the German version, too (Kupfer et al., 2001). This results in 18 items comprising 3 scales for scoring. Each item is answered on a 5-point Likert scale (“strongly disagree”, “moderately disagree”, “neither agree nor disagree”, “moderately agree”, “strongly agree”). Some items need to be inverted before scoring. Each scale is scored separately by calculating a sumscore, additionally, a composite score of “alexithymia" is calculated by adding the scores of the subscales.

2.2.3.7 Fragebogen zur Erhebung der Emotionsregulation bei Erwachsenen (FEEL-E)

The FEEL-E is a self-report questionnaire assessing emotion regulation strategies of anger, fear and sadness in adults (Grob & Horowitz, 2014). For each emotion, 12 different strategies to regulate the emotions are assessed using 24 items, respectively, resulting in 72 items in total. The strategies are separated into 6 adaptive strategies (problem-oriented behavior (orig. “Problemorientiertes Handeln”), acceptance (orig. “Akzeptieren”), cognitive problem solving (orig. “Kognitives Problemlösen”), re-evaluation (orig. “Umbewerten”), improving mood (orig. “Stimmung anheben”), forgetting (orig. “Vergessen”)) and 6 maladaptive strategies (withdrawal (orig. “Rückzug”), self-devaluation (orig. “Selbstabwertung”), giving up (orig. “Aufgeben”), perseveration (orig. “Perseveration”), catastrophizing “orig. “Katastrophisieren”), blaming others (orig. “Anderen die Schuld zuweisen”)). This leads to 20 subscales: adaptive strategies across all emotions and for each emotion separately, maladaptive strategies across all emotions and for each emotion separately, plus one subscale for each strategy. For each subscale, scores are summed of all items included in the subscale. The emotion-specific subscales for each strategy consist of 2 items, respectively. The emotion-specific subscales for all adaptive and maladaptive strategies consist of 12 items each. Finally, the emotion-independent subscales for adaptive and maladaptive strategies consist of 36 items each. Each item is scored on a 5-point Likert scale (“almost never”, “seldom”, “sometimes”, “often”, “almost always”). Each subscale is scored by summing the responses to the respective items of the subscale.

2.2.3.8 Skalen zum Erleben von Emotionen (SEE)

The SEE is a self-report questionnaire assessing the perception, evaluation and coping with emotions in adults based on concepts of emotional intelligence (Behr & Becker, 2004). The questionnaire consists of 42 items divided into seven subscales using factor analysis: accepting own emotions (orig. “Akzeptanz eigener Emotionen”, 6 items), experiencing emotional flooding (orig. “Erleben von Emotionsüberflutung”, 7 items), experiencing lack of emotions (orig. “Erleben von Emotionsmangel”, 5 items), body-focused symbolizing of emotions (orig. “Körperbezogene Symbolisierung von Emotionen”, 8 items), imaginative symbolizing of emotions (orig. “Imaginative Symbolisierung von Emotionen”, 6 items), experiencing emotion regulation (orig. “Erleben von Emotionsregulation”, 4 items), and experiencing self-control (orig. “Erleben von Selbstkontrolle”, 6 items). Each item is rated on a 5-point Likert scale (“strongly disagree”,

“disagree”, “neither agree nor disagree”, “agree”, “agree strongly”). Each subscale is scored by calculating a sumscore across the respective items of the scale.

2.2.3.9 Emotionale-Kompetenz-Fragebogen Selbstauskunft (EKF-S)

The EKF-S is a self-report questionnaire assessing emotional skills and attitudes towards emotions (Rindermann, 2009). 62 items are constructing four subscales: perceiving, recognising and understanding one's own emotions (orig. “Erkennen und Verstehen eigener Emotionen”, 15 items), recognising other’s emotions based on their behavior, verbal statements, facial expressions and body language in relation to the current situation (orig. “Erkennen von Emotionen bei anderen”, 17 items), regulation and control of one’s own emotions (orig. “Regulation und Kontrolle eigener Emotionen”, 13 items), and emotional expressivity, which describes the ability and willingness to express one’s own emotions (orig. “Emotionale Expressivität”, 17 items). Each item is rated on a 5-point Likert scale (“strongly disagree”, “disagree”, “neither agree nor disagree”, “agree”, “agree strongly”). To calculate the score of each subscale, the mean of the items of the subscale is standardized using values defined in the user manual.

2.2.3.10 Affective Neuroscience Personality Scales (ANPS)

The ANPS (Davis et al., 2003) is a self-report questionnaire assessing personality operationalised as the dimensions SEEKING, FEAR, CARE, ANGER, PLAY, SADNESS and spirituality. The ANPS is grounded in the assumption that emotions are an important foundation of personality and conceptualized as a tool to capture variability in personality related to subcortical affective systems (Davis et al., 2003). Thereby, the ANPS is based on the neuroanatomical, neurochemical and physiological affective systems underlying the Five Factor Model (Davis & Panskepp, 2011).

We employed the German version (Reuter et al., 2017). The questionnaire consists of 110 questions. 14 items each belong to the SEEKING, FEAR, CARE, ANGER, PLAY and SADNESS subscales, 12 items comprise the spirituality subscale, 5 items make up a “lie” subscale to indicate social desirability or dishonest responding, and 9 items are filler items not corresponding to any scale. Each item is ranked on a 4-point Likert scale (“strongly disagree”, “disagree”, “agree”, “strongly agree”). Each subscale is rated separately by calculating a sumscore, taking into account that some items need to be reversed before scoring.

2.2.3.11 Positive and Negative Affect Scales (PANAS)

The PANAS (positive and negative affect schedule; Watson et al., 1988) is a self-report questionnaire assessing two dimensions of affect. Positive affect represents feelings of enthusiasm, activity and alertness, whereas negative affect represents feelings of distress and unpleasantness such as anger, contempt, disgust, guilt, fear and nervousness (Watson et al., 1988). We employed the German version of the PANAS (Krohne et al., 1996) that can be used as a trait measure (asking for habitual affect, “How are you feeling in general?”) and as a state measure (asking for current affect, “How are you feeling at this moment?”). The PANAS consists of 20 items that are rated on a 5-point Likert scale (“very slightly or not at all”, “a little”, “moderately”, “quite a bit”, “very much”). Each subscale (positive and negative affect) comprise 10 items, respectively, that are scores in a sumscore for each subscale.

We used the trait version of the PANAS prior to the on site measurements, and the state version directly before and after MRI scanning.

2.2.3.12 State-Trait-Ärgerausdrucks-Inventar (STAXI)

The STAXI is a self-report questionnaire assessing different aspects of anger (Rohrmann et al., 2013). There is a state version and a trait version of the STAXI. The state version assesses situational anger using 15 items constituting 3 subscales: feelings of anger, verbal anger impulse and physical anger impulse. Each subscale is scored by calculating the sum of all constituting items. Each item is rated on a 4-point Likert scale (“not at all”, “a bit”, “a lot”, “very much”). The trait version assesses various dispositions regarding anger using 36 items in 8 subscales. In the first part of the questionnaire, trait anger is assessed using 10 items. These items also constitute the subscales anger temperament and anger reaction. The second part of the questionnaire focuses on anger expression and control, using 26 items constituting the subscales anger expression outwards, anger expression inwards, and anger control, which includes the subscales anger control outwards and anger control inwards. The score of each subscale is calculated by summing the responses of the respective items. Each item is rated on a 4-point Likert scale (“almost never”, “sometimes”, “often”, “almost always”).

We used the trait version of the STAXI during trait phenotypic assessment and the state version once before MRI measurement.

2.2.3.13 Inventar sozialer Kompetenzen Kurzform (ISK-K)

The ISK-K is the short version of the ISK, a self-report questionnaire assessing general social skills (Kanning, 2009). Social skills have been investigated in many subfields of psychology, resulting in no unifying theory behind the construct. Underlying the ISK is the author's definition stating that social skills are the entirety of knowledge, capabilities and skills of a person which promote the quality of their social behavior in the sense of socially skillful behavior, which itself is defined as behavior that contributes to achieving one's goals while simultaneously preserving the social acceptance of that behavior in a specific situation (Kanning, 2002a). The ISK assesses 17 social skills which can be condensed to 4 more abstract skills via factor analysis. These four "secondary" factors (social orientation (orig. "Soziale Orientierung"), offensiveness (orig. "Offensivität"), self-regulation (orig. "Selbststeuerung") and reflexibility (orig. "Reflexibilität")) are assessed in the shorter ISK-K using 10, 8, 8, and 7 items for each scale, respectively, resulting in 33 items in total. The items are answered on a 4-point Likert scale ("strongly disagree", "disagree", "agree", "strongly agree"). Some items need to be inverted before scoring. For each subscale, a sumscore is calculated individually.

2.2.3.14 ADHS-Screening für Erwachsene (ADHS-E)

The ADHS-E (Schmidt & Petermann, 2013) is a self-report questionnaire assessing symptoms of ADHD in adults. The symptoms are based on diagnostic guidelines by the DGPPN (Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde; Ebert et al., 2003), the Wender-Utah criteria (Wender, 1995) and diagnostic criteria described in the DSM IV and ICD-10. The authors developed two versions, one "core screening" (ADHS-E) and a long version (ADHS-LE). We here used the ADHS-E which includes five subscales assessing emotion/affect (orig. "Emotion und Affekt"), control of attention (orig. "Aufmerksamkeitssteuerung"), restlessness/hyperactivity (orig. "Unruhe und Überaktivität"), impulse control/disinhibition (orig. "Impulskontrolle und Disinhibition") and stress intolerance (orig. "Stressintoleranz"). The ADHS-E comprises 25 items that are answered on a 4-point Likert scale ("does not apply", "applies a bit", "applies mostly", "applies always"). For each subscale, a sumscore is calculated. Additionally, a general score subsuming all subsores can be calculated.

2.2.3.15 Schuhfried's Wiener Testsystem: Theory of Mind

The ToM test of the Wiener Testsystem (SCHUHFRIED GmbH, Mödling, Österreich) is a computer-based assessment of theory of mind, the ability to take on other people's perspective and make assumptions about their feelings, thoughts, intentions, desires and motivations. In the task, participants are presented four pictures in random order. Participants need to bring the pictures into the correct order, so that the pictures tell a cohesive story. If participants fail to put the pictures in the correct order, they receive feedback and are asked to try again. Upon success, participants are asked questions about the assumed intentions and thoughts of characters shown in the pictures. In this way, the test assesses understanding of social interactions (ordering of the pictures) and understanding of others' thoughts and intentions (questions). We employed the short version of the test.

2.2.3.16 Verbal Fluency Task

The verbal fluency task is a computer-based task used to assess different linguistic information. Here, we recorded the participant's voice while completing the task to assess information from the fluent speech of participants. We employed three different verbal fluency tasks. In the lexical verbal fluency task, participants are presented with a letter and asked to produce as many words starting with that letter as possible within 2 minutes. Some rules restricted which words were valid responses: words should not be repeated, words should not have the same stem and proper names should not be used. Adjectives, verbs and nouns are all valid. This task was employed twice, once using "M" as the initial letter and once using "K" as the initial letter. Then, the task was modulated to include a switching component. Participants were now asked to apply the same rules, but switch back and forth between words starting with the letter "G" and words starting with the letter "R". In the semantic verbal fluency task, participants were asked to name as many words belonging to a certain category as possible within 2 minutes. No word should be repeated. The categories were "animals", "jobs", and during the switching task, participants were asked to alternate between "sports" and "fruit". Lastly, we employed an interference task. Participants were presented with pictures of objects and asked to name each object as quickly as possible. Simultaneously, participants listened to a spoken word with each picture that was presented. The spoked word should be ignored by the participants and not named. The spoken word was either semantically related or unrelated to the name of the presented object, therefore either supporting or impeding the naming of the object.
This cognitive task was included in our study to bridge this dataset to a dataset that was previously acquired in our institute.

2.2.3.17 Dundee Stress State Questionnaire short version (DSSQ-short)

The DSSQ (Matthews et al., 1999) is a self-report questionnaire assessing task pr performance stress. The DSSQ is a state measure based on the trilogy of mind (Hilgard, 1980) which distinguishes between affective, motivational and cognitive domains underlying mental activity and states. The DSSQ assesses three factors: task engagement, distress and worry (Matthews at al., 2013). We here used the German version of the short version of the DSSQ. The DSSQ can quantify the effect of task stressors on the state of participants by using two assessments, one before the task and one after the task. Each questionnaire of the short version consists of 30 items that can be rated on a 5-point Likert scale. The first questionnaire assesses the state of the participant before the task, including feelings, thoughts and expectations regarding the task. The second questionnaire assesses the same items, but in regards to how participants felt during the task. We applied the questionnaires before and after MRI scanning. Although the participants did not perform a classic task in our study, we wanted to check for stress state and how it is impacted by movie watching. “Task performance” can here be interpreted as the success in continually paying attention to the movie stimuli, but because there was no measurement or feedback on this performance, some items might not apply perfectly to our situation.

2.2.3.18 Self-Assessment Manikin (SAM)

The SAM is a pictorial self-report assessment used to measure affective responses (Lang, 1985). It is based on an approach that situates emotions on three dimensions: valence (pleasure-displeasure), arousal and dominance (Morris, 1995). Through its visual modality, the assessment is independent from language and quick to administer. Each dimension is assessed through one pictogram: a graphic character is depicted in five versions with a nine-point Likert scale underneath, each point corresponding to one character or the space between characters. To assess valence, the character is shown with a frown and strongly downwards turned mouth on one end of the scale, the next picture shows the character with a less intense, but still negative expression, the middle picture shows a neutral expression, then the character is depicted with a slight smile and with a big smile and positive expression at the other end of the scale. To assess arousal, the character is shown with eyes closed, a neutral expression and a small circle in its middle on one end of the scale, then the small circle is replaced by an irregular shape with strokes indicating motion surrounding it. The middle picture shows the character with open eyes and an even larger “commotion” inside of it, which growths further until it covers most of the character in the picture at the other end of the scale. To assess dominance, the character is depicted very small at one end of the scale and then grows continuously until it outgrows the borders of the box it is depicted in in the picture at the other end of the scale. Accordingly, low response values indicate negative valence/displeasure, low dominance/high submissiveness, and low arousal, respectively, while high response values indicate positive valence/pleasure, high dominance and high arousal. The SAM has been used previously in studies investigating affective analysis and emotion annotation of movies (Arifin & Cheung, 2008; Soleymani et al., 2009; Teixeira et al., 2012; Carvalho et al., 2012; Tian et al., 2017, Fernández-Aguilar et al., 2018).

The SAM was filled out once before MRI measurement, and then after each movie stimulus as a quick assessment of the participant's emotional state after each stimulus.

2.2.3.19 Movies-Questionnaire

Lastly, we created a questionnaire to assess the participants' habits regarding movie and TV watching, related preferences and impressions and feelings about the chosen movie stimuli.

Specifically, we first asked questions regarding the movie stimuli that were included in the study. We asked if the longer movie stimuli and the shorter stimuli comprising the SS were known to the participants beforehand. This was answered on a scale representing different levels of familiarity with the stimuli ("I did not know the movie" (orig. "Der Film war mir nicht bekannt."), "I know the movie, but have never seen it" (orig. "Der Film war mir bekannt, ich habe ihn aber noch nicht gesehen."), "I have seen the movie, but do not remember the complete story" (orig. "Ich habe den Film gesehen, kann mich aber nicht an die gesamte Handlung erinnern."), "I have seen the movie" (orig. "Ich habe den Film gesehen")). We then asked, if the movies had been seen, when they had been seen. Afterwards, we asked questions pertaining to all stimuli as shown, including the SS as a whole. We asked how participants liked the stimuli ("I did not like the movie excerpt at all", "I did not like the movie excerpt", "I liked the movie excerpt", "I like the movie excerpt a lot"), how well they could understand the movies acoustically ("not at all", "badly", "moderately", "well", "very well"), if and during which scenes participants closed their eyes or looked away, during which MRI scans participants fell asleep or drowsed (including the resting state and anatomical scans), how well participants could put themselves in the position of the characters in the movie stimuli ("not at all", "a bit", "well", "very well"), in which character's position the participants could put themselves best and worst, and which of the movie stimuli the participants associated most with different feelings (exciting, joyful, sad, positive, negative, inspiring, disgusting, entertaining, surprising, most action, scary, angering, boring).

Then, we posed questions about the participants' habits and preference regarding TV and movie watching. Here, we asked for preferences in movie genre (horror, comedy, thriller/crime, romance, drama, fantasy/sci-fi, not watching movies, other (could be specified)), what participants paid most attention to when watching movies (technical aspects (cuts, camera settings, lighting, etc), musical score, social interactions and relationships between characters), places or landscapes, closeness to reality or continuity ("plotholes", errors in logic or presentation, etc), content (plot or story), actors/actresses (fame, which character is portrayed by who, which other movies did the actors/actresses act in), other (could be specified)), how many hours participants watch movies or TV series in a week, and in which language participants preferentially watch movies or TV series (in the original language, if it is English, in the original language, even if it is not English, partially German, partially other languages, other languages than German with German subtitles, only in German).

2.2.3.20 Hormone sampling

We acquired two saliva samples on site at the day of the MRI measurement. One sample was taken before MRI scanning and one sample was taken after MRI scanning. These samples were not meant or evaluated as two separate timepoints to check for changes in hormone levels over the course of the study. Instead, two samples were taken to provide a stronger estimate of the hormone levels by averaging results of both samples. Should one sample provide too little saliva or inconclusive results, the second sample still provided an estimate of the hormone levels. Hormone sampling was done in collaboration with a specialized company (IBL International GmbH) who provided saliva tubes and evaluated the samples. The participants were asked to fill each tube at least halfway with saliva. We stored the samples in a freezer at -80°C until they were sent in for evaluation. Each sample was tested for levels of cortisol (µg/dl), estradiol (pg/ml), progesterone (pg/ml), and testosterone (pg/ml).

2.2.4 Movie stimuli

Movie stimuli were chosen based on multiple criteria. Against the backdrop of the socio-affective focus of this study and dataset, movie stimuli were intended to depict social interactions or situations, and portray emotions of characters and induce emotions in the viewers. Both portrayed and induced emotions were intended to be complex and multifaceted, with multiple different emotions represented in each movie stimulus. In real life, emotions rarely occur separately and individually, especially in social context and over longer periods of time, therefore, a naturalistic stimulus should likewise convey the complex nature of the emotional experience. There have been previous efforts to create a film library for the induction of mixed emotional states, but here the stimuli were significantly shorter than what we aimed for (Samson et al., 2016). To capture the multiplicity that often only unfolds over the progression of situations and stories, the movie stimuli should cover multiple minutes to capture a specific situation. The viewers lack the context of the entire movie, but each movie stimulus should be long enough for the viewers to understand the situation, empathize with characters and follow depicted social interactions.

Although each movie contains this multiplicity of emotions, we still sought to choose distinct movies with either more positive or more negative emotional profiles overall or even a focus on specific emotions if they could be well represented. The reasoning behind choosing each stimulus is explained in the respective section on the stimulus. Generally, movies were chosen either based on previous research using these movies in emotion induction or naturalistic viewing studies or from personal judgment of the investigators that a movie excerpt would fit into the criteria in a way that no previously studied movie would. Previous studies include a study by Schaefer et al. (2010) on emotion elicitation where 70 movie excerpts corresponding to specific emotion states were rated by participants on emotional arousal, positive and negative affect, discrete emotions felt and corresponding positive and negative composite scores. The ranking was reported in a descending order, meaning that the highest ranking movies were rated highest in the respective category. The movie stimuli used in this study had durations of 1 to 7 minutes, so while many excerpts were shorter than our intended stimulus durations, we checked which excerpts could be prolonged without changing the content and emotional profile of the excerpt profoundly. We also evaluated how much each excerpt did not only correspond to a single emotion, but offered some complexity and multifaceted emotional experience. To do so, we did two additional studies, one focussing on the emotions felt by viewers of our movie stimuli, and one annotation of features of our movie stimuli. These studies are described in sections 2.1 and 2.2.

We created two orders in which movie stimuli were presented during the MRI session. The order of the stimuli presented in a study can generally have an impact on the participants as processing of earlier stimuli can affect processing of later stimuli or RS acquisitions (Gordon et al., 2014; Gaviria et al., 2020), which is why experiments usually use counterbalanced or randomized designs. Here, randomisation might have introduced more uncontrolled confounds if the features of the movie stimuli became very unbalanced for some participants, for example by watching first all more negative and then the more positive movie stimuli. Therefore, we created an order of stimuli that sought to balance the stimuli in regards to general valence and arousal. To still control for the effect of this order, we created a second, inverted order of movie stimuli. Altogether, the naturalistic viewing part of the MRI session started with a short trial movie stimulus that allowed participants to familiarize themselves with the movie watching, adjust the audio volume for best intelligibility and try out the response boxes used to answer the SAM pictorial questionnaire after each movie stimulus. Then, the SS were presented before one of the two versions of the movie stimulus ordering was applied: version one presented TGTBTU, LIB, DMW, FG, a field map scan, DPS, S, and DD, while version two presented DD, S, DPS, FG, a field map scan, DMN, LIB, TGTBTU. The two versions were alternated for participants of the same sex, so that both versions were equally presented to both sexes. The following movie scenes were chosen as movie stimuli:

2.2.4.1 Dirty Dancing (DD)

Dirty Dancing (Artisan Pictures, 1987, directed by Emile Ardolino) is a 1987 American romantic drama dance movie following the story of a young woman falling in love with a dance instructor at a holiday resort. The movie stimulus features the last scene of the movie, where the protagonists perform a dance together in front of the guests of the holiday resort, including an iconic lift. The pair dances to the popular theme song of the movie, "(I've had) The Time of My Life" by Bill Medley and Jennifer Warnes. After their dance, the couple encounter the young woman's father who apologizes for a past error of judgment, compliments their dancing and thus approves of the relationship. All resort guests dance, the couple joins them again and kisses.

This movie scene was chosen because it is a long scene where positive emotions are portrayed continually. Often in movies, positive emotions are closely tied to specific events and reactions to them (for example, someone is happy about receiving positive news). Once the reaction has been shown, the story moves on, which means that the depiction of positive emotions does not last very long on the screen. In DD, a longer period is used to portray the happy ending of the movie including the dance to upbeat music. While this movie stimulus might arguably not be as equivocal as others, we felt that the representation of strong positive emotions over a longer period of time is a valuable addition. The stimulus lasts 7 minutes 35 seconds and corresponds to an excerpt from playtime 01:25:54:12 to 01:33:29:11 or frames 128862 to 140236 of the DVD version of the movie.

2.2.4.2 Scream (S)

Scream (Woods Entertainment, 1996, directed by Wes Craven) is a 1996 American slasher film telling the story of a highschool student and her group of friends who become the target of a masked murderer. The movie stimulus features the beginning of the movie, in which a high school student receives anonymous phone calls while preparing for a movie night with her boyfriend, who is supposed to join her later. While the first phone call appears to be someone calling the wrong number, the caller repeatedly rings again, insisting on talking to the teenage girl about horror movies and her relationship. When the caller mentions that he is watching the girl, she becomes scared and ends the conversation. However, the caller calls again, soon threatening her to stay on the phone and play a game with him. Eventually, the caller reveals that the girl's boyfriend is tied up and hurt on the terrace, again threatening to kill him if the girl does not play a game with him. Within the game, the girl has to answer questions about horror movies. When she answers incorrectly, her boyfriend gets murdered off-screen.

This movie stimulus was chosen, because the beginning scene of this movie creates a lot of tension and suspense. In the previously mentioned study on emotion elicitation using movie excerpts by Schaefer et al. (2010), the beginning of S was one of the excerpts included in the study, although the excerpt differs from the one we used: it is 6 minutes 33 seconds long, starts with a later phone call and ends with the girl's murder. This excerpt was ranked 6 out of 70 on negative affect and 3 out of 70 on fear. In our study, we chose to start the movie stimulus earlier to capture the building suspense as the phone calls get increasingly threatening, and end the excerpt before any explicit violence is shown. Here, our focus was more on threat and tension instead of violence and brutality. Therefore, results from the Schaefer et al. (2010) study are not directly comparable to the stimulus used in our study, but still provide evidence of emotions induced by our stimulus. The stimulus lasts 7 minutes 17 seconds and corresponds to an excerpt from playtime 00:00:16:20 to 00:07:34:01 or frames 420 to 11351 of the DVD version of the movie.

2.2.4.3 Dead Poets Society (DPS)

Dead Poets Society (Touchstone Pictures, 2003, directed by Peter Weir) is a 1989 American drama film following the story of the teacher John Keating at an elite boarding school who inspires his students to think critically and enjoy life to the fullest, beyond their strict and traditional upbringing. As a result of his unorthodox teaching methods and the suicide of a student, Keating gets fired. The movie stimulus features the ending of the movie. The student Todd Anderson is brought into the principal's office and coerced to sign a letter of wrong allegations against his teacher, allowing the school to evict Keating. He is shown watching some students on the school grounds as he is preparing to leave the school. While his class is covered by the headmaster, Keating enters the classroom to gather his belongings. Before Keating leaves the classroom, Anderson tries to explain how he was coerced into signing the letter of allegations and that Keating is innocent. Despite the headmaster's threats of expulsion, Anderson climbs on his desk, exclaiming "O Captain! My Captain!" in reference to the poem by Walt Whitman, introduced to the class by Keating. Soon other students join Anderson and likewise climb on their desks in an emotional farewell gesture to Keating. The headmaster is furious and unable to control his students, while Keating is very touched, thanks them and departs.

This stimulus was chosen because it featured different emotions surrounding complex social topics such as justice, integrity, solidarity and power. The last 2 minutes 40 seconds of our stimulus were also included as an excerpt in the previously mentioned Schaefer et al. (2010) study on emotion elicitation. This excerpt showed the highest ranking on positive affect compared to 70 other movie excerpts and scored rank 6 out of 70 on tenderness and rank 6 out of 70 on the composite positive score. While these rankings indicate mainly positive emotions to be associated with this movie excerpt, the stimulus used in our study includes scenes preceding this excerpt that introduce more contextual information to the excerpt. These scenes include different kinds of conflict, which likely induced more mixed than just positive emotions. The stimulus lasts 7 minutes 50 seconds and corresponds to an excerpt from playtime 01:52:29:03 to 02:00:19:09 or frames 168728 to 180484 of the DVD version of the movie.

2.2.4.4 Forrest Gump (FG)

Forrest Gump (Paramount Pictures, 1994, directed by Robert Zemeckis) is a 1994 American romantic comedy-drama film depicting the life of Forrest Gump, a friendly man of below average intelligence who becomes involved in many important historical events of the times he lives in. The movie stimulus features the beginning of the movie. Forrest Gump, as a grown man, finds a feather and puts it in his suitcase as he is shown sitting on a bench at a bus stop. A young woman comes and sits next to him, taking out a magazine to read, as Forrest Gump starts a conversation by introducing himself. He offers the woman pralines, comments on her shoes and starts reminiscing about his life, capturing the woman's attention after all. A throwback shows Forrest Gump as a child getting leg braces at the doctor's office, how he was named after the founder of the Ku Klux Klan, and about his life with his mother who runs a boarding house in Alabama. His mother encourages him to live a normal life and not let people's gossip get him down, although Forrest's headmaster explains to her his below average intelligence, intending to send Forrest to special needs school. But his mother wants him to receive the best school education and have the same chances in life as other children, so that she has sex with the headmaster in exchange for him letting Forrest go to his school. Later, Elvis Presley stays at the Gump's boarding house where Forrest inspires him to his iconic dance movements. Forrest and his mother later see Elvis perform the dance on a TV when his mother is taken aback by the sexual innuendo of the dance moves. The throwback sequence ends with the young woman intently listening to Forrest's story on the park bench.

This movie stimulus was chosen following the *studyforrest* project. In collaboration with researchers involved in the original project, we here used the beginning of the first movie segment of the *studyforrest* stimulus. The original segment was approximately 15 minutes long, so to keep the duration similar to our other stimuli, we cut the segment at an appropriate scene. No other changes were made to the stimulus to keep high comparability between both datasets. The stimulus lasts 9 minutes 40 seconds and corresponds to an excerpt from playtime 00:02:07:10 to 00:11:47:10 or frames 3186 to 17685 of the DVD version of the movie.

2.2.4.5 Dead Man Walking (DMW)

Dead Man Walking (Orion Pictures Corporation, 1995, directed by Tim Robbins) is a 1995 American crime drama film following the developing relationship of spiritual guidance between Sister Helen Prejean and death row convict Matthew Poncelet. The movie stimulus corresponds to the ending of the movie. Poncelet, hands and feet cuffed, is slowly escorted by prison guards along a corridor to the execution room. He is accompanied by Sister Helen, who puts her hand on his shoulder and reads from the bible to him for solace. She is prohibited from entering the execution room, and sits in an adjacent room where the families of the Poncelet's victims are watching the execution. A clock shows that it is 5 minutes to twelve. Poncelet is strapped to an examination table that has his arms stretched out and a needle is put into his arm. The clock draws closer to twelve. From the visitor's room, a curtain is pulled and Poncelet is presented to the victims' families and Sister Helen. The table is elevated, so that Poncelet is in an upright position, facing the visitors. He is shaking visibly and makes use of his right to say last words. He apologizes to the father of his male victim for killing his son, and tells the parents of his female victim that he hopes that his death will give them some peace. He adds that he thinks killing is wrong, no matter who does it. He is put back into a horizontal position and turned sideways, so that he and Sister Helen can look at each other. She extends her arm, reaching out to him. At twelve o'clock, two guards turn on the machine executing the lethal injection. A flashback shows Poncelet and his accomplice in a forest with their victims. Interspersed with scenes showing the injection process, the flashbacks show him and his accomplice raping and stabbing the female victim, then shooting the male victim and her. As Poncelet dies, his victims and him are shown from a bird's perspective, all in the same position and dead.

This movie stimulus corresponds to the extended version of an excerpt used in the Schaefer et al. (2010) study. The unextended version starts when Poncelet is being strapped to the examination table in the execution room. This excerpt ranked 3 out of 70 on arousal, 6 out of 70 in sadness and 5 out of 70 on a composite negative score. The movie stimulus shows explicit scenes of sexual assault, murder and execution by lethal injection, making it the most violent stimulus. The stimulus lasts 8 minutes 33 seconds and corresponds to an excerpt from playtime 01:38:38:16 to 01:47:11:20 or frames 147966 to 160795 of the DVD version of the movie.

2.2.4.6 Life is Beautiful (LIB)

Life is Beautiful (Miramax, 1997, directed by Roberto Benigni) is a 1997 Italian comedy-drama film following the story of an Italian Jewish family who get deported to a concentration camp while under German occupation. The father, Guido Orefice, uses his imagination to present life in the concentration camp as a game to his young son, Giosuè, promising him that he will get a tank as a reward for completing tasks. The movie stimulus covers the ending of the movie. Guido, the father, pretends to be a woman to search for his wife, Dora, as the camp gets shut down because of the Allied forces' approach. However, he gets found out and is captured by German soldiers. As he is being marched through the desolated camp, he passes a sweat box, where his son is hiding. He winks to his son, keeping up his facade and the game, before he is led to an alleyway and shot. The next morning, the last Nazis leave the deserted camp, followed by the few prisoners remaining, and after all is quiet, young Giosuè leaves his hiding place. Soon after, the Allied forces arrive, led by a tank which Giosuè believes to be his reward. An American soldier emerges from the tank, greets the boy and lets him ride on the tank. The tank follows a procession of army forces and prisoners leaving the camp, and Giosuè spots his mother at the roadside. He is lifted off the tank, runs to her and is finally reunited with his mother, celebrating the end of the war and their liberation.

In the study by Schaefer et al. (2010), four different excerpts from Life is Beautiful were used. Two of those are included in our movie stimulus: the scene in which Guido, the father, is killed, and the liberation of the camp and the succeeding reunion of mother and son. To make the movie stimulus of similar duration to all other stimuli, we chose to use an excerpt of the movie spanning both scenes. The reunion scene of mother and son was rated 10 out of 70 on positive affect, 3 out of 70 on tenderness and 8 out of 70 on a composite positive score. The scene in which the father is killed ranked 10 out of 70 on sadness. This shows that our movie stimulus includes scenes of very disparate emotional effects. The stimulus lasts 7 minutes 26 seconds and corresponds to an excerpt from playtime 01:49:30:03 to 01:56:56:00 or frames 164253 to 175400 of the DVD version of the movie.

2.2.4.7 The Good, the Bad, the Ugly (TGTBTU)

The Good, the Bad and the Ugly (Alberto Grimaldi Productions, 1966, directed by Sergio Leone) is a 1966 Italian spaghetti Western following the story of three gunslingers trying to find a hidden gold treasure during the American Civil war. The movie stimulus takes part towards the end of the movie. One man, Tuco Ramirez ("The Ugly"), comes to a grave and uses a wooden plank he finds nearby to dig up the dirt on the grave. As he is uncovering a wooden casket, a shadow approaches and a shovel is thrown towards him. The person throwing the shadow is revealed to be Blondie (“The Good”), who wants Tuco to continue digging for an apparent treasure hidden in the grave. Another shovel is thrown to the men by Angel Eyes (“The Bad”) who joins them, pointing a pistol at them to force them to dig up the treasure. Blondie reveals that no treasure is hidden in the grave, kicking open a wooden casket to reveal only a skeleton. Blondie, who knows the true location of the treasure and has fooled the others, writes the name on the grave in which the treasure lies on a rock and places it in the middle of an empty yard. Slowly, the three men take up positions around the rock, entering a Mexican stand-off. They are observing each other suspiciously during the three-way duel. They slowly move their hands towards their guns, as their eyes are shown in accelerating progression. Finally, Blondie shoots Angel Eyes first, while Tuco finds his gun malfunctioning.

An excerpt of this movie was used in one of the first studies on naturalistic viewing (Hasson et al., 2004), and different excerpts have since been used in various studies with different research foci (Mantini et al., 2012; Betti et al., 2013; Tarvainen, Westmann & Oittinen, 2015; Lu et al., 2016; Gilson et al., 2018; Demirtaş et al., 2019; Hlinka et al., 2022). In line with our focus on socio-affective themes in movies, we chose to use the ending of the movie as it includes the climax of the conflict between the three main characters. Although tension is created in the final stand-off scene, the way the film is directed and edited may feel unusually drawn-out and slow in comparison to more modern movies, which usually show shorter scenes and more cuts (Cutting & Candan, 2015). The stimulus lasts 8 minutes 23 seconds and corresponds to an excerpt from playtime 02:32:57:20 to 02:41:20:11 or frames 229445 to 242011 of the DVD version of the movie.

2.2.4.8 Short Sequences (SS)

The “Short Sequences” were constructed as a complementing modality to the longer movie stimuli. The SS are 12 short movie clips taken from 10 movies presented as one stimulus with no breaks between clips. The stimulus lasts 10 minutes 39 seconds. Contrary to longer stimuli spanning a contiguous excerpt of one movie, these short clips resemble more strongly the stimuli used in emotion induction paradigms, where short scenes are commonly used to induce specific emotions. Here, too, we chose movie scenes that show some emotional and social content, although this content is not as unequivocal as content used in emotion induction paradigms. As with the longer movie stimuli, we sought scenes that could evoke different emotional and cognitive reactions in viewers. However, the much shorter duration limits the potential for emotional complexity of the stimuli when scenes are supposed to be taken in and understood in a timeframe of seconds. We included a big range of emotion by choosing very different scenes portraying anger, hurt, love, fear, aggression, embarrassment and more emotional states. In an effort to balance the stimulus as a whole, we alternated scenes with rather positive and scenes with rather negative emotional content. We further alternated scenes with higher and lower arousal.

The clips are the following (in order of appearance in the SS compilation):

1. Gone Girl (Twentieth Century Fox, 2014, directed by David Fincher): A man awakes from hearing car tires screeching and leaves his house to check for the noise. A car has crashed into his front garden and a woman, wearing only a white dress and covered in blood, leaves the car and goes to the astounded man. The street is lined with media cars and reporters who rush in to photograph the scene and interview the couple. The woman hugs the man, who whisperingly insults her and she collapses into his arms.

This scene was chosen as it starts calmly, takes an unexpected turn and intrigues without much resolve of the characters and their story. The stimulus lasts 54 seconds and corresponds to an excerpt from playtime 02:00:26:18 to 02:01:20:16 or frames 180668 to 182016 of the DVD version of the movie.

2. Seven (New Line Cinema, 1995, directed by David Fincher): In a desert, two men are arguing about a third man, wearing an orange prison suit. While one man throws his gun away at the beginning of the scene and tries to deescalate the fight, the other man keeps gesticulating with his gun, desperately discussing the contents of a box. The prisoner provokes the upset man who continues to threaten him with his gun in return. The prisoner insinuates the murder of a woman and her unborn child until the first man hits him to stop him from speaking further, leaving the upset man in shock.

A longer excerpt including this scene was used in Schaefer et al.'s (2010) study. It ranked 8 out of 70 on arousal and 9 out of 70 on anger. The stimulus lasts 53 seconds and corresponds to an excerpt from playtime 01:53:27:01 to 01:54:20:06 or frames 170176 to 171506 of the DVD version of the movie.

3. Sex of Angels (PRO-FUN MEDIA, 2011, directed by Xavier Villaverde): Two young men hug each other. One starts kissing the other, they smile and continue kissing passionately. A photo collage on a wall is shown, containing photos of one of the men with a woman.

This scene was included as it portrays a romantic homosexual scene between two men, which might induce various reactions in viewers depending on their feelings towards homosexuality. The stimulus lasts 34 seconds and corresponds to an excerpt from playtime 00:35:59:14 to 00:36:34:10 or frames 53989 to 54860 of the DVD version of the movie.

4. The Hobbit - Battle of the Five Armies (Warner Bros. Pictures, 2014, directed by Peter Jackson): A flying dragon dives from the air with fire growing in his throat to attack a

city surrounded by water. The dragon spits fire across the houses and people, many of whom are in boats on the waterways between the houses. People try to flee from the flames and a man in a prison cell tries to break open the bars in front of a window. The dragon keeps attacking and spitting fire, while one boat carrying large amounts of gold and treasure tries to flee the city. They ram another boat, but never stop to help other people.

This scene was included because it shows a fictional attack by a fantastic animal and the ensuing destruction and chaos. The fantasy element is unique to this stimulus. The stimulus lasts 59 seconds and corresponds to an excerpt from playtime 00:02:46:24 to 00:03:46:00 or frames 4174 to 5650 of the DVD version of the movie.

5. Lost in Translation (Constantin Film Verleih, 2003, directed by Sofia Coppola): a busy street in Tokyo is shown. A white man walks through the crowd, apparently following a blonde woman who is walking away. The man catches up to the woman and talks to her, so that she turns around. She smiles and he steps very closely in front of her. They hug and he pats the back of her head as she burrows her face into his shoulder. He whispers something incomprehensible into her ear, she agrees and they continue hugging.

This scene was included because it depicts a subtle social scene in everyday life where nothing dramatic happens. The stimulus lasts 1 minute 1 second and corresponds to an excerpt from playtime 01:28:51:07 to 01:29:52:08 or frames 133282 to 134808 of the DVD version of the movie.

6. Blue is the warmest color (Wild Bunch, 2013, directed by Abdellatif Kechiche): A woman is watching the crowd around her as loud music plays. She sees multiple lesbian couples kissing each other. A blue-haired woman gives her a kiss on the cheek. They hold hands and dance to the music, watching other people dance too, partially wearing elaborate make-up und costumes. They dance in a crowd waving many rainbow flags.

Again, this scene was chosen because of the LGBTQIA+ representation, this time in a more fun-centered and social setting. The stimulus lasts 58 seconds and corresponds to an excerpt from playtime 01:18:39:24 to 1:19:38:13 or frames 117999 to 119463 of the DVD version of the movie.

7. Star Wars - Revenge of the Sith (Twentieth Century Fox, 2005, directed by George Lucas): A man is flying on a droid across a lava stream as he is swirling a lightsaber. Another man, flying across the lava on a bigger metal platform, attacks him with his lightsaber. They fight viciously. In a short break, they argue with each other about who is evil. The first man jumps across the second man on his platform and they continue fighting with their lightsabers.

This scene was chosen because it shows strong conflict between two characters that is fought out through violence in a fictional setting. The stimulus lasts 1 minutes 2 seconds and corresponds to an excerpt from playtime 01:53:17:00 to 01:54:18:24 or frames 169925 to 171474 of the DVD version of the movie.

8. Short Term 12 (Short Term Holdings, 2013, directed by Destin Daniel Cretton): A man and a woman sit next to each other. The woman leans on the man. Next, they are shown as the woman receives an ultrasound of her pregnant belly and the doctor guides their attention to the images generated by the ultrasound showing their baby and his or her heartbeat. The man and the woman hug with great emotion and she is crying.

This scene was chosen as it depicts an intimate, tender moment between two expecting parents experiencing overwhelming emotions. The stimulus lasts 44 seconds and corresponds to an excerpt from playtime 01:24:27:11 to 01:25:11:18 or frames 126686 to 127793 of the DVD version of the movie.

9. Shining (Warner Bros. Entertainment, 1980, directed by Stanley Kubrick): A man repeatedly smashes an ax into a door as a woman cries out in terror. She is shown hiding behind the door, holding a knife. The door splinters as the ax comes through it. The woman begs the man to stop. When there is a bigger hole in the door, he presses his face through it, threatening the woman. The man reaches through the hole with his hand towards the door knob to open the door. The woman slashes his hand with the knife. The man screams and pulls back his hand.

This scene is included in an excerpt used in Schaefer at al.'s (2010) study, where the longer excerpt ranked 9 out of 70 on negative affect and 2 out of 70 on fear. The shorter scene we used here is the ending of the excerpt and depicts the eventual confrontation between the man and woman and the end of their chase. The stimulus lasts 53 seconds and corresponds to an excerpt from playtime 01:35:54:19 to 01:36:48:20 or frames 143875 to 145220 of the DVD version of the movie.

10. When Harry Met Sally (Castle Rock Entertainment, 1989, directed by Rob Reiner): A man is shown while a woman is moaning loudly in the background. The woman is shown and seems to be having an orgasm, while people around her including the man sitting opposite of her stare at her in astonishment. They are sitting in a diner with a lot of guests. The woman finishes her performance, smiles at the man and starts eating the food in front of her. Another guest is ordering whatever the woman is eating.

This scene is part of an excerpt used in both Gross & Levenson (1995) and Schaefer et al.'s (2010) study, where the longer excerpt was ranked 6 out of 70 on positive affect and 9 out of 70 on amusement. It was included because it might induce second-hand embarrassment and/or amusement in regards to the public behavior of the woman and the surprised man who is with her. The stimulus lasts 46 seconds and corresponds to an excerpt from playtime 00:44:46:00 to 00:45:32:13 or frames 67150 to 68313 of the DVD version of the movie.

11. Blue is the warmest color (Wild Bunch, 2013, directed by Abdellatif Kechiche): Two women are shown arguing. The woman with short hair tells the other woman to leave her apartment. The other woman, her hair in a bun, is crying. The first woman pulls a bag from a closet and starts throwing the other woman's clothing into it. The second woman tries to stop her and both argue intensely. The first woman slaps the second woman, repeatedly screaming at her to leave. The second woman still tries to argue and begs to talk, but the first woman shoves her out of the apartment and shuts the door forcefully. Glass shatters.

This scene was included because it depicts a heated argument between two women in a relationship in their home, which is again a more everyday life setting than other stimuli. The stimulus lasts 1 minute 4 seconds and corresponds to an excerpt from playtime 02:13:33:22 to 02:14:38:19 or frames 200347 to 201969 of the DVD version of the movie.

12. Short Term 12 (Short Term Holdings, 2013, directed by Destin Daniel Cretton): A woman is shown running in slow motion, followed by two men, trying to catch a body wearing only shorts and an American flag around his shoulders. The people run across the lawn in front of a flat building as the camera zooms out.

This scene is the ending of the movie and was included to end the compilation on a neutral or positive scene with lower arousal. The stimulus lasts 46 seconds and corresponds to an excerpt from playtime 01:28:33:05 to 1:29:19:24 or frames 132830 to 133999 of the DVD version of the movie.

2.2.4.9 Movie stimulus creation

Movie stimuli were created using the “melt” command line video editor (Yates & Dennedy, 2012). The stimulus was cut from the DVD version. All movie segments were brought into the same format and gray bars were added above and below the videos for uniform presentation. The frames-per-second were set to 25 for each stimulus and audio volume was adjusted individually.

2.2.5 fMRI data acquisition

MRI scanning was performed on a Siemens 3T Prisma scanning (Siemens, Erlangen, Germany) with a 64-channel head coil. A structural T1w image was acquired using an MP-RAGE sequence (TR=2000 ms, TE=2.45 ms, TI=900 ms, flip angle=8°, FoV: 256mm) yielding 1mm3 voxel size. Functional data were acquired using a T2w multiband echo planar imaging sequence (TR=980 ms, TE=30 ms, flip angle=70°, voxel size=2.2 x 2.2 x 2.0mm3, slice thickness=2 mm, multiband acceleration factor=4, phase encoding direction=AP, FoV=207mm) across 64 slices. A mirror was fixed on the head coil to allow participants to see a screen that was used to display the movie stimuli. In-ear headphones were used for ear protection and to deliver the sound of the movie stimuli. All participants completed two RS scans (8 minutes 21 seconds) with their eyes held open and the screen turned off and an anatomical T1w scan (5 minutes 6 seconds). Afterwards, a short movie was shown to assess audio quality, adjust the volume of the movie stimuli if needed to the subject's preference and test the response button used to respond to the SAM that was used to assess valence, arousal and dominance after each movie. Data of this "test movie" were not used in any analyses. Participants then watched the movie stimulus SS, which is a compilation of 12 short movie clips (with durations between 35 seconds to 66 seconds) taken from conventional movies with a total duration of 10 minutes 34 seconds. Afterwards, participants watched 7 more movie clips in one of two versions of stimulus order. All movies were presented in the German dubbed version. Participants were instructed to watch each movie and afterwards respond to the SAM. Foam wedges were fitted around participants' heads to increase comfort and reduce movement.

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 20.1.0 (Esteban, Markiewicz, et al. (2018); Esteban, Blair, et al. (2018); RRID:SCR_016216), which is based on *Nipype* 1.4.2 (Gorgolewski et al. (2011); Gorgolewski et al. (2018); RRID:SCR_002502).

2.2.6 Statistical analysis of phenotypic data

The phenotypic data that was collected in the Movies study is characterized based on the distribution of scores attained by the participants. To make the distributions more comparable between measures within this study, despite the varying scoring methods and possible score ranges, all scores were z-standardized. Unstandardized scores are still reported for comparability of the scores with the same measures in other samples. Because data of the verbal fluency test are still unprocessed in raw, audio format, no further statistical analysis is reported. Paired samples t-tests were used to test for significant differences in the full sample's means of measures acquired pre and post MRI measurement (PANAS, DSSQ).

2.3 Annotation of emotions induced by the movie stimuli

We conducted a study on the emotions perceived by viewers of our movie stimuli to better characterize the movie stimuli used in the main “Movies” study.

2.3.1 Sample

We acquired data of 44 German participants (23 male, age: 20-30 years, mean age 25.05 years, SD 3.23 years). Exclusion criteria were any psychiatric or neurological disease, impaired vision (that cannot be corrected for using glasses) or hearing, alcohol, drug or medication addiction, usage of medication affecting the central nervous system within the last 3 months, and any MRI contraindication. All participants gave written informed consent. The study was conducted in accordance with the declaration of Helsinki and approved by the ethics committee of the Heinrich-Heine-Universität Düsseldorf (grant number “2018-100_3-andere Forschung erstvotierend”).

2.3.2 Study Protocol

Participants were invited to an on-site assessment. After giving informed consent, participants filled out a demographics questionnaire and state versions of the BDI, STAI, PANAS and SAM questionnaires. Then, participants watched all movie stimuli used in our main study while simultaneously rating the emotions they experience during movie watching. Afterwards, participants filled out the questionnaires about movie watching created for the main study, and again the state version of the PANAS. Introductions to all measures and questionnaires can be found in section 2.2.3.

To measure the emotions experienced by participants while watching the movies, we used the ReMoTa toolbox (ReMoTa - Real-time Movie Tagging v0.0 by Luca Cecchetti, Giada Lettieri and Giacomo Handjaras; Lettieri et al., 2019). Here, participants were presented with the movie stimuli on a laptop (Fig 3). Below the video, six bars represented six basic emotions (happiness, fear, surprise, sadness, disgust and anger, Ekman & Friesen, 1971; Ekman et al., 1983). Each bar was filled with white color in relation to how strongly participants indicated to feel each emotion. To rate their feelings, participants used the keyboard of the laptop. Two pairs of keys were used to increase or decrease the reported intensity of each emotion. Keys used to decrease the perceived intensity of an emotion were situated below the keys used to increase the perceived intensity of an emotion. Happiness was increased by pressing the w key and decreased by pressing the a key, fear was increased by pressing the e key and decreased by pressing the s key, surprise was increased by pressing the r key and decreased by pressing the d key, sadness was increased by pressing the t key and decreased by pressing the f key, disgust was increased by pressing the z key and decreased by pressing the g key and anger was increased by pressing the u key and decreased by pressing the h key. All emotions were rated simultaneously. Participants were instructed to indicate every in- or decrease in the perceived intensity of the emotions via the respective keys, when no change in intensity was happening, ratings remained the same. The intensity of perceived emotions was rated on a scale from 0 to 100. Each keystroke in- or decreased the intensity by 20 percent, meaning that 5 keystrokes were needed to change the rating from 0 to 100. The sampling rate was set to 10 hz, which corresponds to a recording of the rating every 100ms. This resulted in one timeseries for each emotion, rated in each movie stimulus.

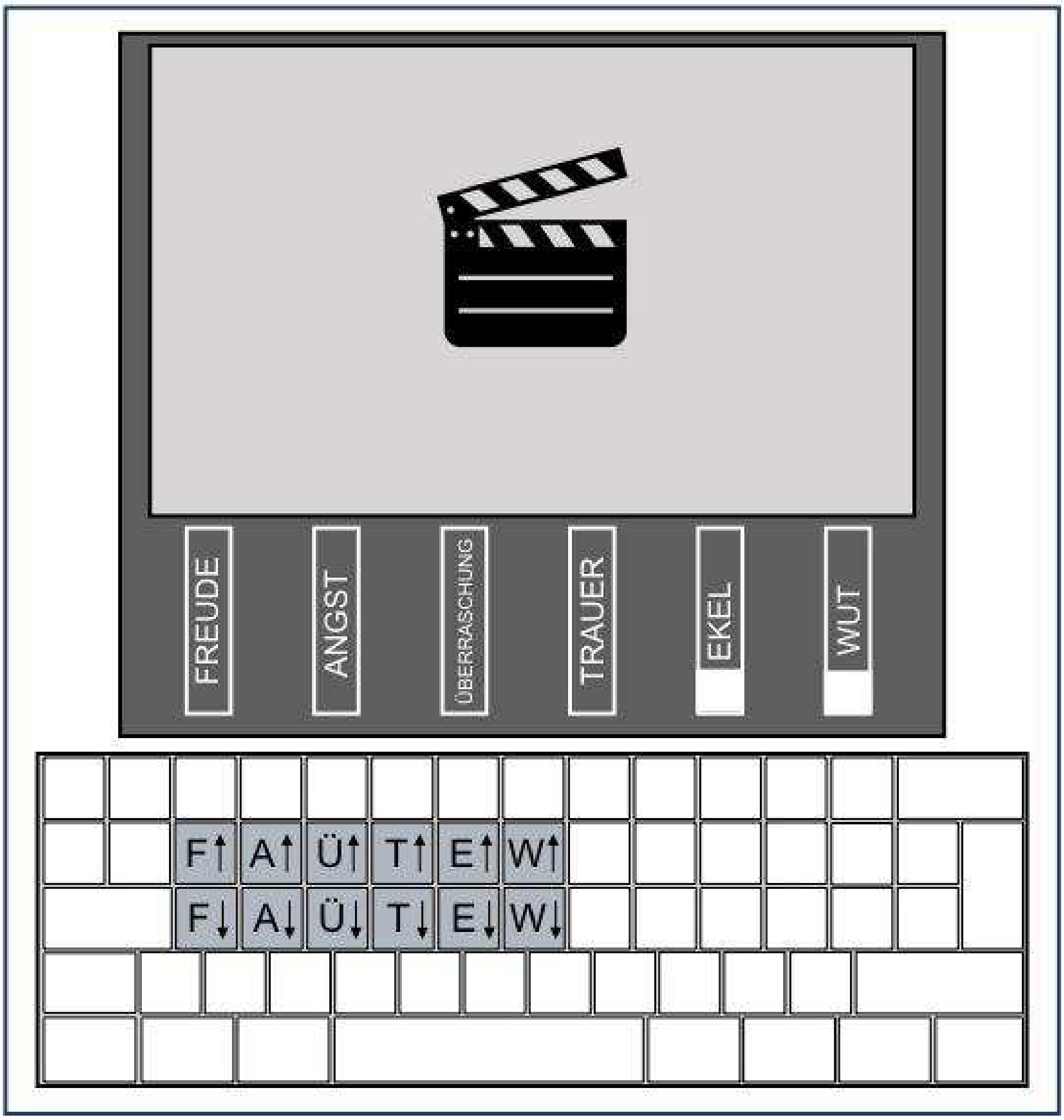


Fig. 3. **Sketch of the experimental set-up of the ReMoTa toolbox.** On the screen of the laptop the movies were displayed, while bars depicted the allocated intensities of perceived emotions below via the height of their white filling. Bars were marked with the German translation of the six basic emotions. The height of the white filling of the emotion bars was regulated using the keyboard of the laptop. From left to right, emotions were regulated in the same order as shown on the screen. Upper keys (W,E,R,T,Z,U) were used to increase the intensity of perceived emotion

and lower keys (A,S,D,F,G,H) were used to decrease the intensity of perceived emotions. Keys were marked with the first letter of the basic emotion they were used for and an arrow indicating the direction of rating.

2.4 Annotation of features of the movie stimuli

The annotation of emotions elicited by the movie stimuli of our main study is an important part of the characterization of our stimuli. To further expand this characterization, we carried out a second annotation study. Here, we focused on features of the movie stimuli themselves instead of the effects they are having on viewers. Specifically, this annotation set out to quantify the following features: faces, bodies, presence of a crowd, male presenting characters, female presenting characters, presence of hands, presence of movement, ethnicity of characters, children, adults, social interactions, place (inside or outside), place (urban vs. non-urban), time of day (day or night), presence of music, presence of buildings, presence of vehicles, presence of food, presence of landscapes, presence of animals, presence of plants, presence of camera movements, weather.

Two independent raters then used the video annotation software *Advene* (Annotate Digital Video Exchange on the NEt; Aubert & Prié, 2005) to quantify the presence or number of occurrences of each feature. The software enables users to make notes (annotations) to precise timepoints in the movie or spanning multiple timepoints. In this study, the raters rated each feature (for example, number of faces or presence of hands) for the entire duration of each movie stimulus. Ratings included the comments on the feature and beginning and end times of each feature comment both in total ms from beginning of the movie stimulus and hours:minutes:seconds format (for example: Annotation of number of adults, begin ms: 0, end ms: 7000, begin time 00:00:00, end time 00:00:07, feature rating = 1; begin ms: 7000 end ms: 8000, begin time 00:00:07, end time 00:00:08, feature rating = 8).

2.5. Covariation of state & trait anxiety with NFC under naturalistic stimulation

2.5.1 Sample

We used data from 58 healthy native speakers of German (32 male, 26 female, age 18 - 33 years, mean age 23.4 years, SD 3.5 years) acquired in the main “Movies” dataset (see Project 2). Exclusion criteria, study protocol and ethics approval are detailed in section 2.2.1 and 2.2.2.

2.5.2 Data preprocessing

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 20.1.0 (Esteban, Markiewicz, et al. (2018); Esteban, Blair, et al. (2018); RRID:SCR_016216), which is based on *Nipype* 1.4.2 (Gorgolewski et al. (2011); Gorgolewski et al. (2018); RRID:SCR_002502).

2.5.2.1 Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.2.0 (Avants et al. 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target 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, Brady, and Smith 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, RRID:SCR_001847, Dale, Fischl, and Sereno 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). Volume-based spatial normalization to standard space (MNI152NLin6Asym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following templates were selected for spatial normalization: *FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model* [Evans et al. (2012), RRID:SCR_002823; TemplateFlow ID: MNI152NLin6Asym], *ICBM 152 Nonlinear Asymmetrical template version 2009c* [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

2.5.2.2 Functional data preprocessing

First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were 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). Susceptibility distortion correction (SDC) was omitted. 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 six degrees of freedom. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. 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 into standard space, generating a *preprocessed BOLD run in MNI152NLin6Asym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Automatic removal of motion artifacts using independent component analysis (ICA-AROMA, Pruim et al. 2015) was performed on the *preprocessed BOLD on MNI space* time-series after removal of non-steady state volumes and spatial smoothing with an isotropic, Gaussian kernel of 6mm FWHM (full-width half-maximum). Corresponding "non-aggressively" denoised runs were produced after such smoothing. Additionally, the "aggressive" noise-regressors were collected and placed in the corresponding confounds file. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). 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 were 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 were estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask was obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components were 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). Components were also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components were dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. 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 output 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).

2.5.3 State-Trait-Anxiety Inventory

We here used data from the STAI (see section 2.2.3.5). We used data of both the trait and state version of the STAI.

2.5.4 Meta-analytical networks and NFC

The functional connectome of each of 14 meta-analytically defined networks was created by calculating the Pearson correlation between all nodes of the respective network. In further analysis, only the unique values of the symmetrical connectivity matrix excluding auto-correlation values were used, representing the strength of correlation between all possible pairs of nodes in the network. These correlation values constitute the NFC. NFC was computed for each network in each condition (8 movie conditions, 2 RS conditions).
The meta-analytical networks are described in section 2.1.4. All nodes, their respective MNI peak coordinates and abbreviations can be found in supplementary table 2. A detailed description of the networks can be found in the supplements.

2.5.5 Representational similarity analysis (RSA)

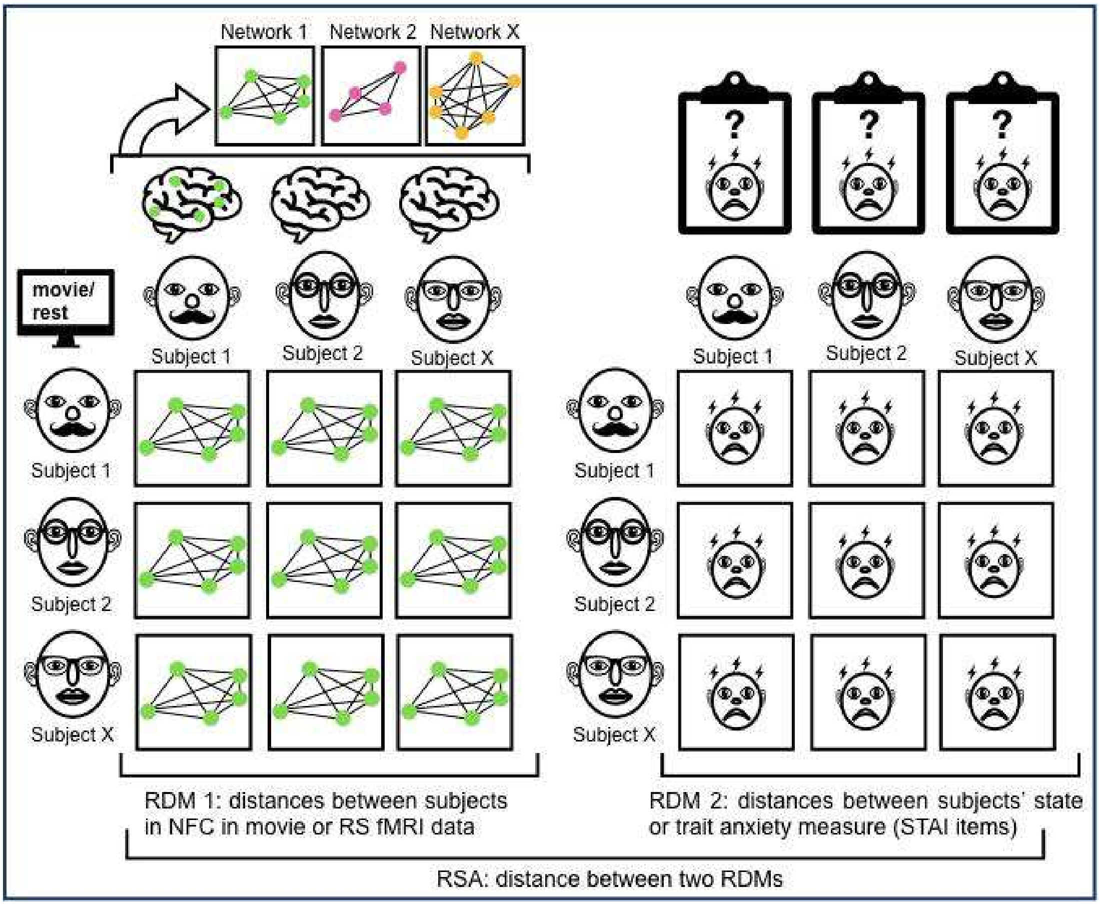


Figure 4. **Representational similarity analysis (RSA).** RSA between representational dissimilarity matrices based on distances between subjects in NFC (left) or based STAI items (right).

We employed a representational similarity analysis (RSA, Figure 4) to investigate the similarity of covariation between subjects in NFC and anxiety scores. To do so, we first computed one representational dissimilarity matrix (RDM) on the neuroimaging data and one on the behavioral data. The RDM consists of the distances between the measures of all subjects. For the neuroimaging data, the distances are calculated as 1 - the Pearson correlation between the NFC of each subject and all other subjects, in each network and condition. For the behavioral data, the distances are calculated as 1 - the Spearman correlation between the response to all items of the state or trait STAI of each subject and all other subjects. These distance measures represent the covariation between subjects in their NFC or anxiety measures. Within the RSA, we calculated the distance between a neuroimaging-based RDM and a behavioral RDM using 1 - Spearman correlation, which results in a measure of how similar the covariation between subjects is in both. Low distances represent high similarity in covariance, so that subjects whose NFC are very similar to one another also respond very similarly to items of the STAI. High distances represent low similarity in covariance, so that subjects whose NFC are very similar to one another do not respond similarly to items of the STAI. We calculated a RSA for each network in each condition (8 movies, 2 RS) and for each anxiety measure (state and trait anxiety). To correct for multiple comparisons, we performed family-wise error correction (FWE) per network.

2.5.6 Statistical analysis

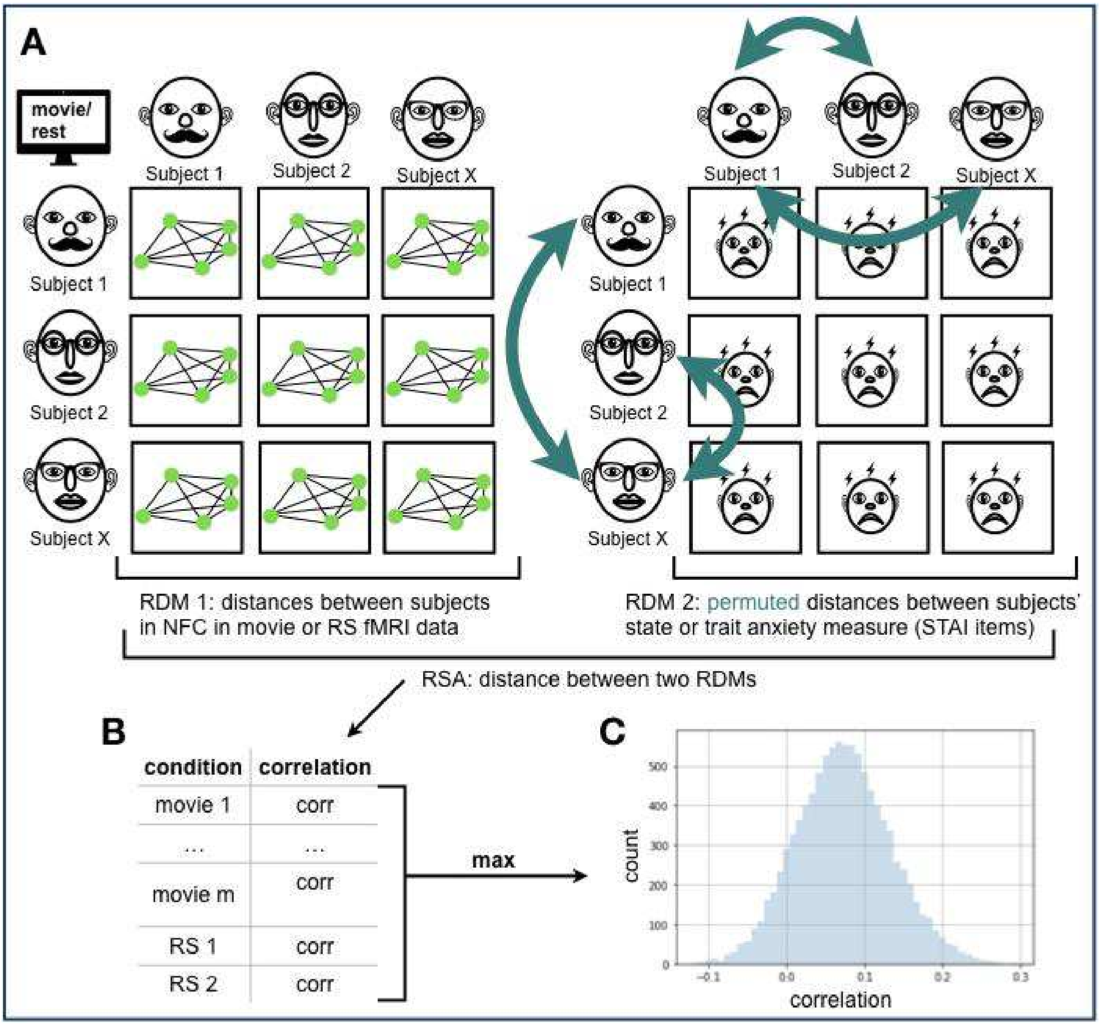


Figure 5. **Creation of null distributions for each combination of network and behavioral measure.** A) The subject labels of the behavioral RDM were reordered in 10,000 permutations. Then the RSA was calculated with the original NFC RDM. B) The results of the RSA in every condition were collected and only the maximum value is added to C), the distribution of surrogate correlation values. This process was repeated 10,000 times, thus, the null distribution consists of 10,000 correlation values.

To evaluate the significance of the RSAs we used bootstrapping to create a family-wise null distribution per network and behavioral measure (s. Figure 5). For each behavioral measure, the subject labels of the respective RDM (rows and columns) were randomly reordered, which means that the relationship between the item responses to the state or trait STAI and the identity of the subject these responses belong to were destroyed. This permutation was done 10,000 times to assure randomness of the results. Afterwards, we calculated the correlation between the permuted behavioral RDM and the original NFC RDM in 10,000 permutations. Each permutation was done across all conditions (8 movies, 2 RS) and from the resulting correlation coefficients, only the maximum was chosen to be included in the family-wise null distribution of the respective network. This bootstrapping approach yielded 28 null distributions, one for each of the 14 networks in combination with each one of the two permuted behavioral measures. The test for significance for the following RSA was done by comparing the true correlation coefficient of the non-permuted data to the family-wise null distribution of the corresponding behavioral measure and network. Here, we assessed the percentile rank of the true correlation coefficient relative to the null distribution and calculated the corresponding p-value with (100 - percentile)/100. We chose $p < .05$ as the threshold for significance.

3. Results

3.1 Inter- and intra-subject variability in NFC across a full narrative movie

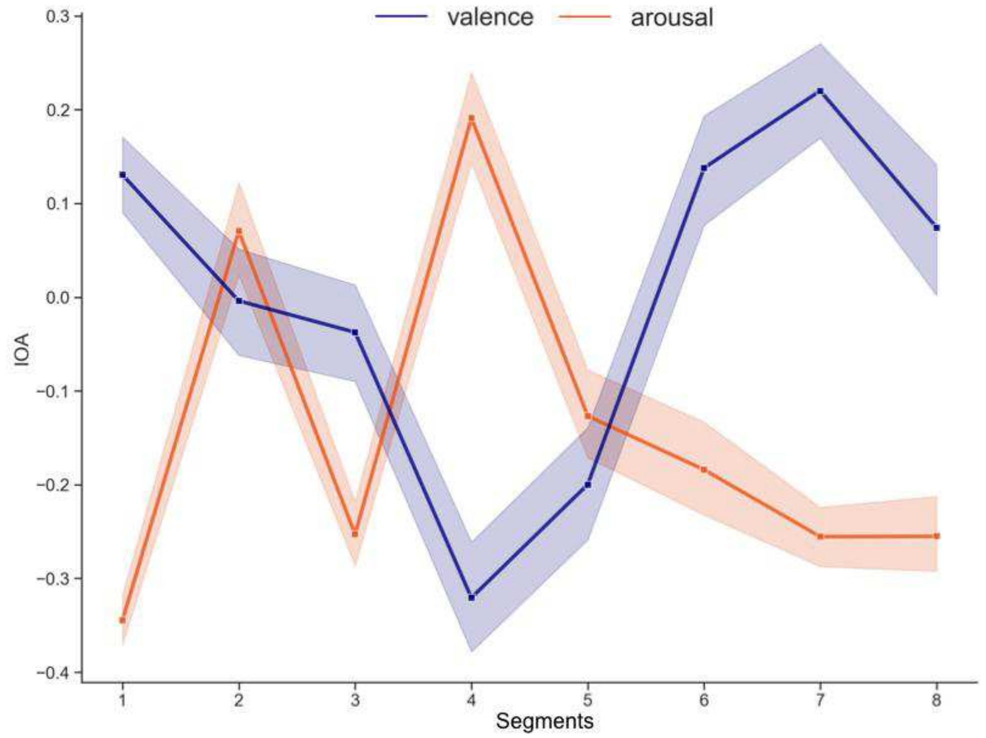
Emotions and an overarching narrative are hallmark features of conventional Hollywood movies, which are frequently employed in NV research because of their engaging and complex nature. However, most NV studies use only shorter clips from these movies, essentially excluding effects of the ongoing narrative. Therefore, is it not yet clear how these features might impact inter- and intra-subject variability in NFC in a full narrative movie. Here, we investigated portrayed valence and arousal across a full narrative movie and how these factors contribute to explaining inter- and intra-subject variability in NFC in 14 networks.

3.1.1 Movie segments and portrayed emotions

We used a previously reported description (Labs et al., 2019) of portrayed valence and arousal for comparisons between the emotional content of different movie segments. Our results showed that movie segments differed in the direction (i.e.: positive/negative valence ; high/low arousal) and the extent of agreement between observers concerning these measures (Figure 6). Figure 6 shows large differences in the evaluation of valence and arousal across movie segments. For segments 1, 6, 7, and 8 IOA values indicate consistency in portrayed positive valence, while the segments 4 and 5 portrayed negative valence. Segment 2 and 3 showed little consistency in the evaluation of portrayed valence, as IOA values are close to zero. Concordantly, the ANOVA on the valence IOA values resulted in a significant main effect of segment ($F(7,3534) = 45.879$, $P < .001$), and Bonferroni-corrected post-hoc testing revealed significant differences between the consecutive segments 1 and 2 ($t = 3.378$, $p = .021$), 3 and 4 ($t = 7.236$, $p < .001$), 4 and 5 ($t = -3.131$, $p = .049$), 5 and 6 ($t = -8.519$, $p < .001$) and 7 and 8 ($t = 3.552$, $p = .011$). Segment 4 had the strongest agreement on negative valence between observers, while segment 7 showed the strongest agreement on positive valence between observers. Figure 2 further shows that segments 2 and 4 portrayed high arousal, while the other segments portrayed low arousal. The ANOVA on arousal IOA values showed a significant main effect of segment as well ($F(7, 3534) = 15.479$, $p < .001$). Bonferroni-corrected post-hoc testing revealed significant differences between consecutive segments 1 and 2 ($t = -13.448$, $p < .001$), 2 and 3 ($t = 10.397$, p $.001$), 3 and 4 ($t = -14.628$, $p < .001$) and 4 and 5 ($t = 10.617$, $p < .001$).

Given how much portrayed emotions and the narrative are intertwined, our results are best interpreted in the light of the content of the movie segments. Segment 1 spans the

A



B

Valence	S1	S2	S3	S4	S5	S6	S7
S1							
S2	3.378						
S3	4.212	0.853					
S4	11.630	8.107	7.236				
S5	8.401	4.948	4.095	-3.131			
S6	-0.185	-3.535	-4.363	-11.722	-8.518		
S7	-2.371	-5.879	-6.745	-14.587	-11.162	-2.157	
S8	1.319	-1.822	-2.599	-9.410	-6.455	1.482	3.552

Arousal	S1	S2	S3	S4	S5	S6	S7
S1							
S2	-13.448						
S3	-2.957	10.397					
S4	-17.799	-3.976	-14.628				
S5	-7.139	6.424	-4.106	10.617			
S6	-5.196	8.180	-2.218	12.359	1.860		
S7	-3.034	11.026	0.081	15.526	4.410	2.416	
S8	-2.704	9.788	0.060	13.705	3.897	2.136	-0.012

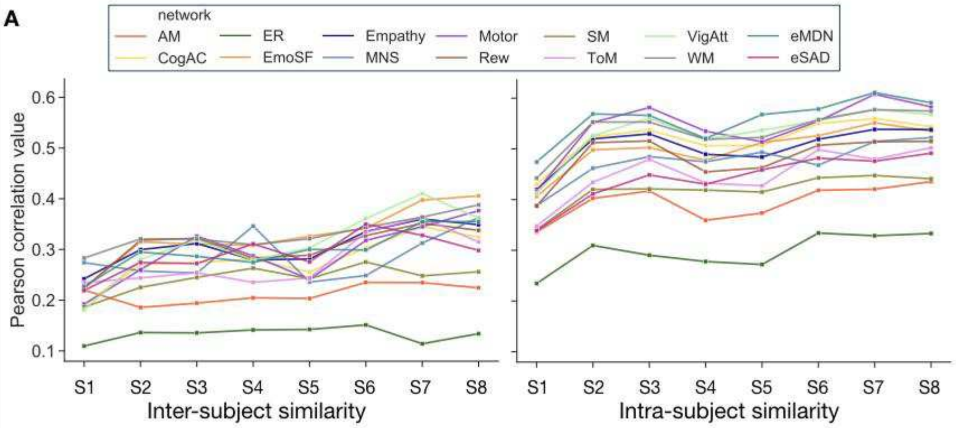
Figure 6. **Results of the ANOVA on valence and arousal inter-observer agreement (IOA) in each movie segment.** (A) Valence and arousal IOA across movie segments (portrayed valence: purple and arousal: orange). Positive IOA values indicate that observers agreed on the portrayal of positive valence and high arousal, while negative IOA values indicate that observers agreed on the portrayal of negative valence and low arousal. The amount of deviation from zero in IOA values corresponds to the strength of agreement between observers. (B) Post-hoc results of the ANOVA on valence (left) and arousal (right). Bonferroni-corrected significance levels are represented by colors: orange signifies p-values < .001, yellow marks p-values < .05 and white marks no significance. S1-S8 = segments 1-8. Direction of the T-tests are column minus row element.

introduction of Forrest Gump and scenes from his childhood, containing both positive (caring mother, close friendship with neighbor girl Jenny) and negative (walking impairments, bullying) elements. Segment 2 was marked by low IOA in both valence and arousal, showing less agreement between observers on the portrayed emotions in this segment. During this segment the movie shows Forrest's highschool and college time, addressing athletic successes and first dating experiences. Low IOA values continue in the valence dimension in segment 3, whereas observers agreed more strongly on low arousal being portrayed here. Here, the movie shows Forrest joining the army, reuniting with Jenny in a nightclub where she works as a dancer, and being deployed in the Vietnam war. Segment 4 prominently features a different pattern than any other segment: observers agreed that movie characters displayed high arousal and low valence during this segment. This can likely be attributed to the war scenes involving an ambush in Vietnam causing Forrest's best friend's death, and following scenes in a military hospital, although the segment also contains Forrest receiving the Medal of Honor and speaking at an anti-war rally in front of the Pentagon. Segment 5 is marked by lower IOA values indicating some negative valence and low arousal, featuring the Black Panther movement, Forrest's ping pong career and reunions with friends Jenny and Lt. Dan. The last three segments again display a pattern of higher agreement between observers on positive valence and low arousal, when the movie spans Forrest's successful shrimp fishing business, two episodes of living happily with Jenny, a three-year cross-country marathon, Forrest meeting his son, Jenny's death and the ending of the movie.

3.1.2 Inter- and intra-subject variability in NFC

Inter- and intra-subject correlations were calculated for every network on the level of single segments, i.e. the different segments of the movie. Figure 7 summarizes the results across all networks and segments based on Pearson correlation coefficients (Figure 7A), and shows results of the LMM analyses on inter- (Figure 7B) and intra-subject variability (Figure 7C). We found that intra- is overall higher than inter-subject variability and inter- and intra-subject variability fluctuate across time for all networks. We will further analyze the statistical significance in the following sections. Because all analyses were based on inter- and intra-subject correlation coefficients representing similarities between subjects, all interpretations for inter- and intra-subject variability are reversed (i.e.: lower coefficient = higher variability).

A



B

INTER	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	0.245	0.011	23.028	0.000	0.224	0.266
network[T._AM]	-0.066	0.008	-8.641	0.000	-0.081	-0.051
network[T._CogAC]	0.002	0.008	0.238	0.812	-0.013	0.017
network[T._ER]	-0.146	0.008	-19.073	0.000	-0.161	-0.131
network[T._EmoSF]	0.050	0.008	6.520	0.000	0.035	0.065
network[T._Empathy]	0.029	0.008	3.743	0.000	0.014	0.044
network[T._MNS]	0.008	0.008	1.002	0.316	-0.007	0.023
network[T._Motor]	0.015	0.008	1.911	0.056	0.000	0.030
network[T._Rew]	0.028	0.008	3.634	0.000	0.013	0.043
network[T._SM]	-0.036	0.008	-4.721	0.000	-0.051	-0.021
network[T._ToM]	-0.001	0.008	-0.102	0.919	-0.016	0.014
network[T._VigAtt]	0.033	0.008	4.346	0.000	0.018	0.048
network[T._WM]	0.053	0.008	6.880	0.000	0.038	0.068
network[T._eMDN]	0.020	0.008	2.655	0.008	0.005	0.035
network[T._eSAD]	0.012	0.008	1.609	0.108	-0.003	0.027
movie segment	0.006	0.001	6.220	0.000	0.004	0.008
valence	-0.311	0.038	-8.082	0.000	-0.386	-0.235
movie segment:valence	0.073	0.008	8.833	0.000	0.057	0.089
subject Var	0.001	0.007				

C

INTRA	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	0.473	0.029	16.279	0.000	0.416	0.529
network[T._AM]	-0.080	0.011	-7.091	0.000	-0.103	-0.058
network[T._CogAC]	0.044	0.011	3.874	0.000	0.022	0.066
network[T._ER]	-0.178	0.011	-15.707	0.000	-0.200	-0.156
network[T._EmoSF]	0.025	0.011	2.242	0.025	0.003	0.048
network[T._Empathy]	0.029	0.011	2.548	0.011	0.007	0.051
network[T._MNS]	0.000	0.011	0.012	0.990	-0.022	0.022
network[T._Motor]	0.067	0.011	5.927	0.000	0.045	0.090
network[T._Rew]	0.008	0.011	0.693	0.489	-0.014	0.030
network[T._SM]	-0.057	0.011	-5.056	0.000	-0.080	-0.035
network[T._ToM]	-0.026	0.011	-2.254	0.024	-0.048	-0.003
network[T._VigAtt]	0.056	0.011	4.958	0.000	0.034	0.079
network[T._WM]	0.062	0.011	5.427	0.000	0.039	0.084
network[T._eMDN]	0.084	0.011	7.397	0.000	0.062	0.106
network[T._eSAD]	-0.034	0.011	-2.971	0.003	-0.056	-0.011
movie segment	0.002	0.002	0.981	0.327	-0.002	0.005
arousal	0.015	0.023	0.630	0.529	-0.031	0.061
valence	-0.771	0.090	-8.601	0.000	-0.946	-0.595
movie segment:valence	0.180	0.018	10.151	0.000	0.145	0.215
movie segment:arousal:valence	0.172	0.021	8.253	0.000	0.131	0.213
subject Var	0.010	0.046				

not sign.
$p < .05$
$p < .001$

AM = autobiographical memory, CogAC = cognitive attention control, ER = emotion regulation, EmoSF = emotional scene and face processing, MNS = mirror neuron system, Rew = reward, SM = semantic memory, ToM = theory of mind, VigAtt = vigilant attention, WM = working memory eMDN = extended multiple demand network, eSAD = extended socio-affective default mode network, S1-S8 = segments 1-8

Figure 7. **Results of the LMM showing how the fixed effect network, movie segment, arousal and valence and the random effect subject identity contribute to inter- and intra-subject**

variability in NFC. (A) Inter- and intra-subject correlation of each movie segment averaged in each network, based on Pearson correlation coefficients. The x axis depicts different movie segments. The y axis represents the averaged variability of the functional connectivity matrix derived from one subject compared to all other subjects within each network. (B) Results of the LMM on inter-subject variability. (C) Results of the LMM on intra-subject variability.

3.1.2.1 Inter-subject variability

The best model that best fitted on inter-subject variability as selected using BIC consists of the random factor subject identity, the fixed factors network, movie segment and valence, and the interaction between the fixed factors movie segment and valence. All parameter estimates and p-values can be seen in Figure 7B. The intercept for inter-subject variability is 0.245, representing the average inter-subject correlation value. Of all 14 networks, the AM, ER, EmoSF, Empathy, Rew, SM, VigAtt, WM and eMDN networks differed significantly from the “mean network” reference category representing the mean inter-subject variability across all networks. The AM, ER and SM networks show negative coefficients, indicating that inter-subject variability is higher in these networks than on average. The EmoSF, Empath, Rew, VigAtt, WM and eMDN networks were associated with lower inter-subject variability than average. Movie segment, valence and their interaction effect reached significance as well. While movie segment and the movie segment-valence interaction were associated with lower inter-subject variability, valence was associated with higher inter-subject variability. The estimated coefficient for subject identity was 0.001, indicating a low effect of subject identity on inter-subject variability and small differences between subjects.

3.1.2.2 Intra-subject variability

The best model that best fitted on intra-subject variability as selected using BIC consists of the random factor subject identity, the fixed factors network, movie segment, valence and arousal, and the interactions between fixed factors movie segment and valence and between movie segment, arousal and valence. All parameter estimates and p-values can be seen in Figure 7C. The intercept for intra-subject variability is 0.473, representing the average intra-subject correlation value. The AM, CogAC, ER, EmoSF, Empathy, Motor, SM, ToM, VigAtt, WM, eMDN and eSAD network differed significantly from the reference category representing the mean intra-subject variability across all networks. The AM, ER, SM, ToM and eSAD networks were associated with higher intra-subject variability, whereas the CogAC, EmoSF, Empathy, Motor, VigAtt, WM and eMDN networks were associated with lower intra-subject variability than average. While movie segment and arousal did not reach significance, valence, the movie segment-valence interaction and the movie segment-arousal-valence interaction did. Valence was associated with higher intra-subject variability, while the movie segment-valence and movie segment-arousal-valence interaction were associated with lower intra-subject variability. The estimated coefficient for subject identity was 0.01, indicating a low effect of subject identity on intra-subject variability and small differences between subjects.

3.2 Movies in the Magnet: a comprehensive socio-affective movie watching dataset

3.2.1 Phenotypic characterization

		total	sex: male	sex: female
number of subjects		84	43	41
age (years)	range	18 - 35	18 - 35	18 - 30
	mean	24.33	24.35	24.32
	STD	3.93	4.36	3.6
education	no graduation	0	0	0
	Realschule	1	1	0
	vocational training	5	1	4
	Abitur	41	23	18
	university degreee	37	18	19
gender different from sex		1	1	0

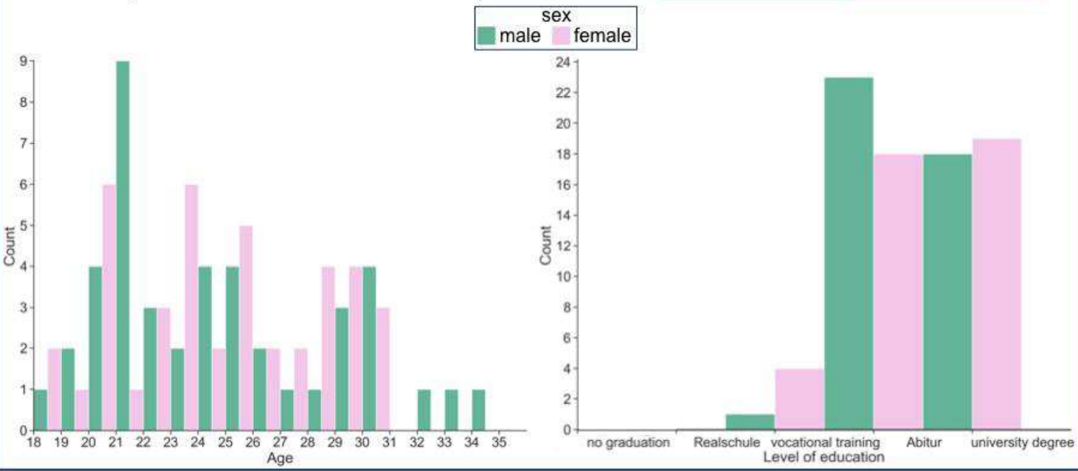


Figure 8. **Demographic information of the “Movies” dataset sample.** Information about male subjects are presented in green, information about female subjects are presented in pink.

The sample of the “Movies” dataset containing all subjects that had undergone MRI measurement and complete phenotypic characterization consists of 84 subjects. Distributions of age and gender can be seen in Figure 8. 43 are of male and 41 are of female sex. The age range of the sample is 18 - 35 years (mean: 24.33 years, SD: 3.93 years). Male and female participant showed similar distributions of age (male mean: 24.35 years, male SD: 4.36 years; female mean: 24.32 years, female SD: 3.6 years) and education (18 males and 19 females had a university degree, 23 males and 18 females had graduated highschool with “Abitur”, 1 male and 4 female had graduated from vocational training, 1 male and no females had graduated from “Realschule”, i.e. completed 10 years of schooling).

To compare participants' scores across all questionnaires and measurements, all scores were z-standardized. As far as possible, all results are presented using the same scale.

3.2.1.1 Beck's Depression Inventory (BDI)

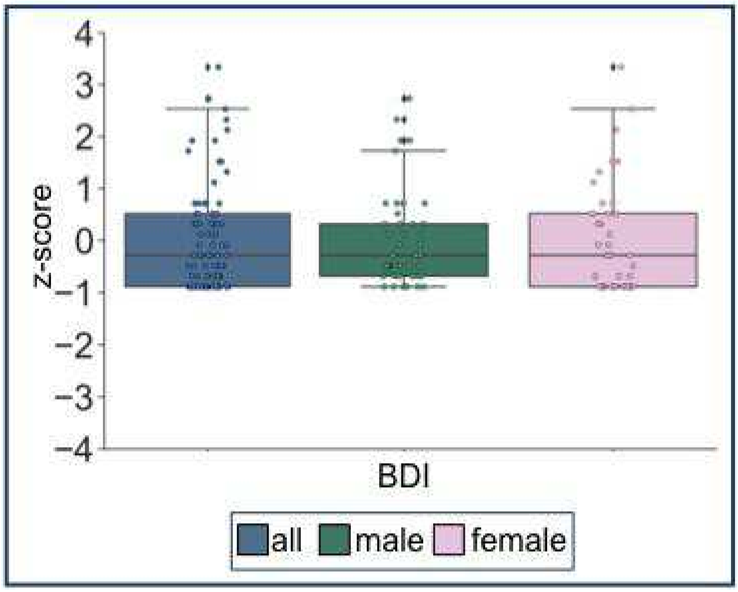


Figure 9. **Z-standardized scores of the BDI.**

Unstandardized scores of the BDI for the complete sample range from 0 to 21 (mean: 4.44, SD: 4.97). Unstandardized scores of the BDI for male participants range from 0 to 18 (mean: 4.56, SD: 4.63). Unstandardized scores of the BDI for female participants range from 0 to 21 (mean: 4.32, SD: 5.35).

3.2.1.2 Edinburgh handedness inventory (EHI)

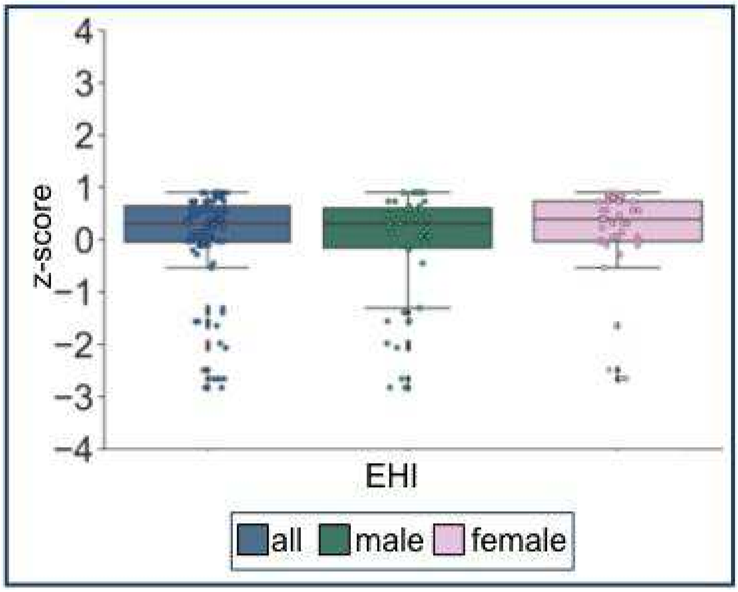


Figure 10. **Z-standardized scores the EHI.**

Unstandardized scores of the EHI for the complete sample range from -22 to 22 (mean: 11.04, SD: 11.82). Unstandardized scores of the EHI for male participants range from -22 to 22 (mean: 10.19, SD: 12.8). Unstandardized scores of the EHI for female participants range from -20 to 22 (mean: 12.68, SD: 10.71).

3.2.1.3 NEO Five-Factor Inventory (NEO-FFI)

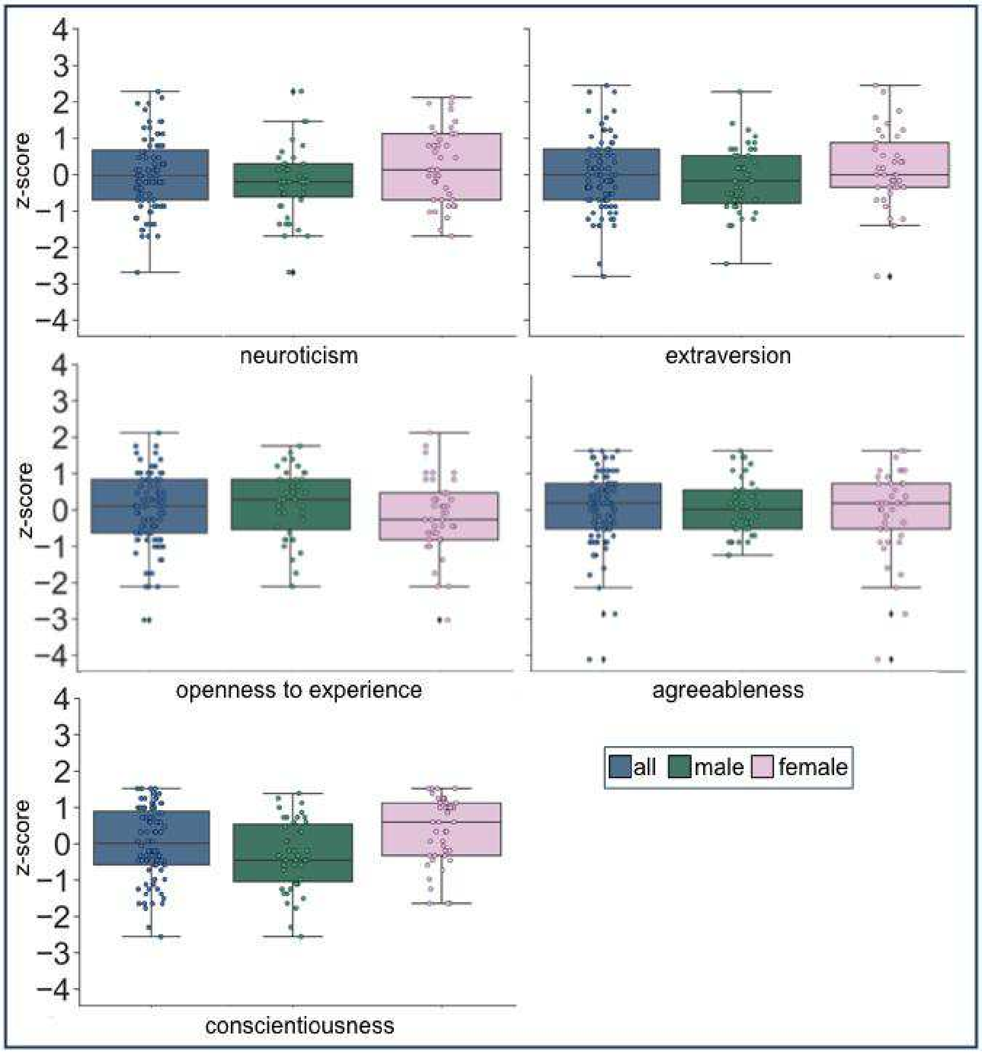


Figure 11. **Z-standardized scores of the NEO-FFI.**

Unstandardized scores of the NEO-FFI subscale neuroticism for the complete sample range from 0.2 to 6.2 (mean: 3.44, SD: 1.2). Unstandardized scores of the NEO-FFI subscale neuroticism for male participants range from 0.2 to 6.2 (mean: 3.18, SD: 1.12). Unstandardized scores of the NEO-FFI subscale neuroticism for female participants range from 1.4 to 6 (mean: 3.72, SD: 1.24).

Unstandardized scores of the NEO-FFI subscale extraversion for the complete sample range from 2.2 to 8.2 (mean: 5.4, SD: 1.42). Unstandardized scores of the NEO-FFI subscale extraversion for male participants range from 2.6 to 8 (mean: 5.23, SD: 1.05). Unstandardized scores of the NEO-FFI subscale extraversion for female participants range from 2.2 to 8.2 (mean: 5.58, SD: 1.22).

Unstandardized scores of the NEO-FFI subscale openness to experience for the complete sample range from 3 to 8.6 (mean: 6.29, SD: 1.09). Unstandardized scores of the NEO-FFI subscale openness to experience for male participants range from 4 to 8.2 (mean: 6.47, SD: 0.99). Unstandardized scores of the NEO-FFI subscale openness to experience for female participants range from 3 to 8.6 (mean: 6.1, SD: 1.52).

Unstandardized scores of the NEO-FFI subscale agreeableness for the complete sample range from 2.2 to 8.6 (mean: 6.79, SD: 1.12). Unstandardized scores of the NEO-FFI subscale agreeableness for male participants range from 5.4 to 8.6 (mean: 6.87, SD: 0.87). Unstandardized scores of the NEO-FFI subscale agreeableness for female participants range from 2.2 to 8.6 (mean: 6.7, SD: 1.33).

Unstandardized scores of the NEO-FFI subscale conscientiousness for the complete sample range from 3.2 to 9.4 (mean: 7.09, SD: 1.52). Unstandardized scores of the NEO-FFI subscale conscientiousness for male participants range from 3.2 to 9.2 (mean: 6.61, SD: 1.46). Unstandardized scores of the NEO-FFI subscale conscientiousness for female participants range from 4.6 to 9.4 (mean: 7.6, SD: 1.43).

3.2.1.4 State Trait Anxiety Inventory (STAI)

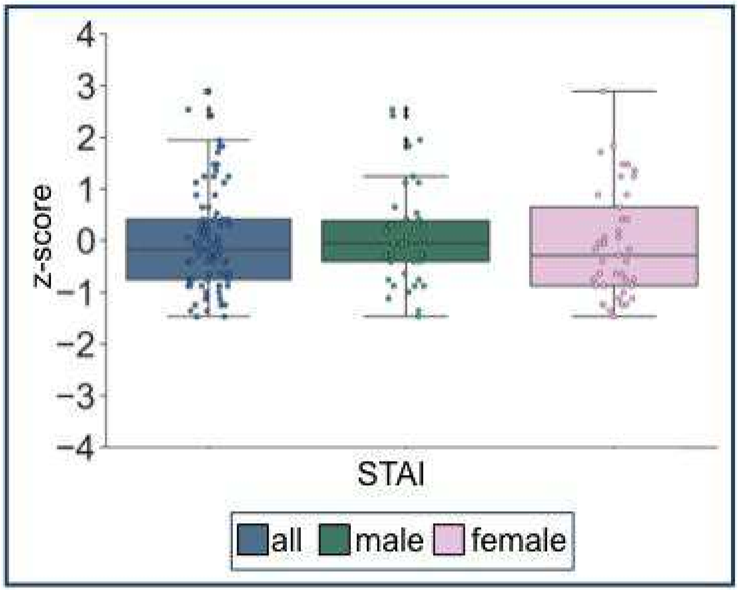


Figure 12. **Z-standardized scores of the STAI.**

Unstandardized scores of the STAI for the complete sample range from 21 to 58 (mean: 33.48, SD: 8.48). Unstandardized scores of the STAI for male participants range from 21 to 55 (mean: 34.02, SD: 7.96). Unstandardized scores of the STAI for female participants range from 21 to 58 (mean: 32.93, SD: 9.05).

3.2.1.5 Toronto Alexithymia Scale (TAS-26)

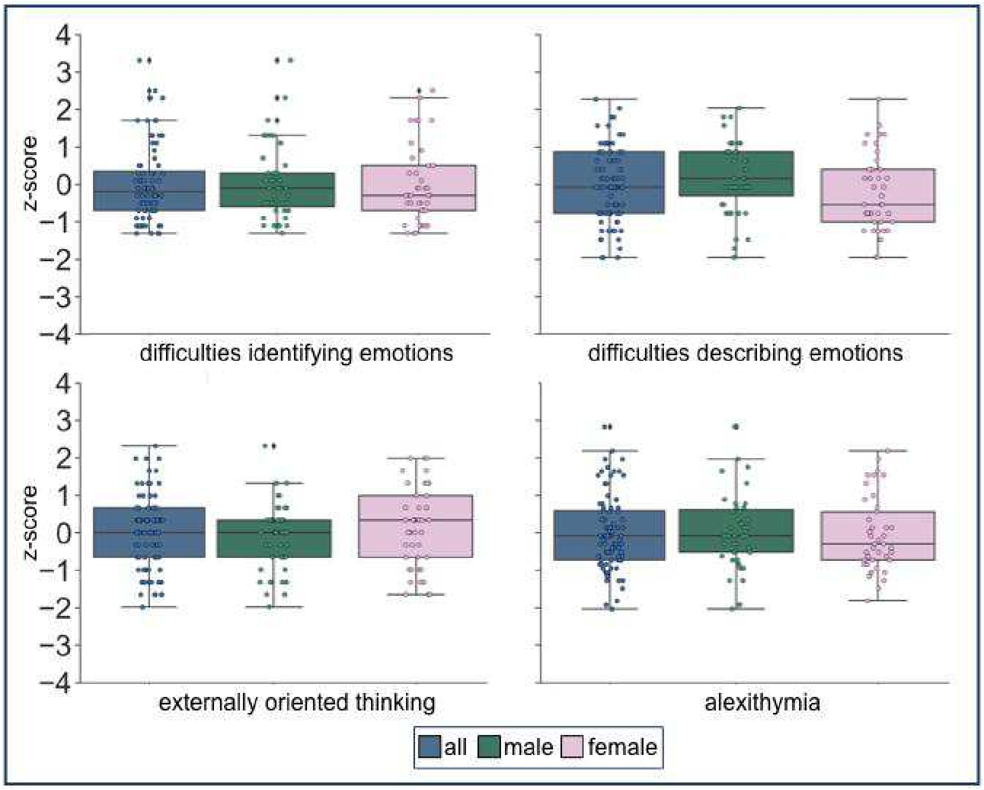


Figure 13. **Z-standardized scores of the TAS-26.**

Unstandardized scores of the TAS-26 subscale difficulties identifying emotions for the complete sample range from 7 to 30 (mean: 13.5, SD: 4.98). Unstandardized scores of the TAS-26 subscale difficulties identifying emotions for male participants range from 7 to 30 (mean: 13.6, SD: 4.93). Unstandardized scores of the TAS-26 subscale difficulties identifying emotions for female participants range from 7 to 26 (mean: 13.39, SD: 5.09).

Unstandardized scores of the TAS-26 subscale difficulties describing emotions for the complete sample range from 5 to 23 (mean: 13.31, SD: 4.27). Unstandardized scores of the TAS-26 subscale difficulties describing emotions for male participants range from 5 to 22 (mean: 14.16, SD: 4.09). Unstandardized scores of the TAS-26 subscale difficulties describing emotions for female participants range from 5 to 23 (mean: 12.42, SD: 4.32).

Unstandardized scores of the TAS-26 subscale externally oriented thinking for the complete sample range from 7 to 20 (mean: 12.99, SD: 3.03). Unstandardized scores of the TAS-26 subscale externally oriented thinking for male participants range from 7 to 20 (mean: 12.63, SD: 2.77). Unstandardized scores of the TAS-26 subscale externally oriented thinking for female participants range from 8 to 19 (mean: 13.37, SD: 3.27).

Unstandardized scores of the TAS-26 composite alexithymia scale for the complete sample range from 21 to 66 (mean: 39.8, SD: 9.25). Unstandardized scores of the TAS-26 composite alexithymia scale for male participants range from 21 to 66 (mean: 40.4, SD:

9.25). Unstandardized scores of the TAS-26 composite alexithymia scale for female participants range from 23 to 60 (mean: 39.17, SD: 9.33).

3.2.1.6 Fragebogen zur Erhebung der Emotionsregulation bei Erwachsenen (FEEL-E)

Figure 14. **Z-standardized scores of the FEEL-E.** AS = adaptive strategies. MS = maladaptive strategies.

Unstandardized scores of the FEELE subscale adaptive strategies (AS) across emotions for the complete sample range from 93 to 164 (mean: 128.6, SD: 16.58). Unstandardized scores of the FEELE subscale AS across emotions for male participants range from 93 to 164 (mean: 126.86, SD: 15.55). Unstandardized scores of the FEELE subscale AS across emotions for female participants range from 97 to 162 (mean: 130.41, SD: 17.6).

Unstandardized scores of the FEELE subscale AS anger for the complete sample range from 20 to 58 (mean: 41.75, SD: 6.5). Unstandardized scores of the FEELE subscale AS anger for male participants range from 29 to 54 (mean: 42.26, SD: 5.62). Unstandardized scores of the FEELE subscale AS anger for female participants range from 20 to 58 (mean: 41.23, SD: 7.35).

Unstandardized scores of the FEELE subscale AS fear for the complete sample range from 29 to 57 (mean: 44.1, SD: 6.29). Unstandardized scores of the FEELE subscale AS fear for male participants range from 29 to 57 (mean: 42.63, SD: 6.36). Unstandardized scores of the FEELE subscale AS fear for female participants range from 29 to 55 (mean: 45.63, SD: 5.92).

Unstandardized scores of the FEELE subscale AS sadness for the complete sample range from 25 to 57 (mean: 42.75, SD: 6.83). Unstandardized scores of the FEELE subscale AS sadness for male participants range from 25 to 57 (mean: 41.98, SD: 6.48). Unstandardized scores of the FEELE subscale AS sadness for female participants range from 30 to 55 (mean: 43.56, SD: 7.17).

Unstandardized scores of the FEELE subscale problem-oriented behavior for the complete sample range from 16 to 30 (mean: 23.6, SD: 3.91). Unstandardized scores of the FEELE subscale problem-oriented behavior for male participants range from 17 to 30 (mean: 23.14, SD: 3.62). Unstandardized scores of the FEELE subscale problem-oriented behavior for female participants range from 16 to 30 (mean: 24.07, SD: 4.17).

Unstandardized scores of the FEELE subscale acceptance for the complete sample range from 13 to 29 (mean: 21.3, SD: 3.8). Unstandardized scores of the FEELE subscale acceptance for male participants range from 14 to 28 (mean: 21.4, SD: 3.47). Unstandardized scores of the FEELE subscale acceptance for female participants range from 13 to 29 (mean: 21.2, SD: 4.15).

Unstandardized scores of the FEELE subscale cognitive problem solving for the complete sample range from 14 to 39 (mean: 24.27, SD: 3.5). Unstandardized scores of the FEELE subscale cognitive problem solving for male participants range from 17 to 30 (mean: 24.16 SD: 3.36). Unstandardized scores of the FEELE subscale cognitive problem solving for female participants range from 14 to 30 (mean: 24.39, SD: 3.65).

Unstandardized scores of the FEELE subscale re-evaluation for the complete sample range from 7 to 29 (mean: 17.94, SD: 4). Unstandardized scores of the FEELE subscale re-evaluation for male participants range from 12 to 29 (mean: 18.16, SD: 4.01). Unstandardized scores of the FEELE subscale re-evaluation for female participants range from 7 to 24 (mean: 17.7, SD: 4.03).

Unstandardized scores of the FEELE subscale improving mood for the complete sample range from 7 to 30 (mean: 20.23, SD: 5.19). Unstandardized scores of the FEELE subscale improving mood for male participants range from 7 to 27 (mean: 19.35, SD: 4.88). Unstandardized scores of the FEELE subscale improving mood for female participants range from 10 to 30 (mean: 21.15, SD: 5.4).

Unstandardized scores of the FEELE subscale forgetting for the complete sample range from 14 to 30 (mean: 21.26, SD: 3.68). Unstandardized scores of the FEELE subscale forgetting for male participants range from 15 to 30 (mean: 20.65, SD: 3.03). Unstandardized scores of the FEELE subscale forgetting for female participants range from 14 to 28 (mean: 21.9, SD: 4.19).

Unstandardized scores of the FEELE subscale maladaptive strategies (MS) across emotions for the complete sample range from 57 to 128 (mean: 89.14, SD: 15.76). Unstandardized scores of the FEELE subscale MS across emotions for male participants range from 63 to 128 (mean: 90, SD: 14.44). Unstandardized scores of the FEELE subscale MS across emotions for female participants range from 57 to 116 (mean: 88.24, SD: 17.17).

Unstandardized scores of the FEELE subscale MS anger for the complete sample range from 20 to 44 (mean: 31.06, SD: 5.57). Unstandardized scores of the FEELE subscale MS anger for male participants range from 20 to 40 (mean: 30.4, SD: 4.72). Unstandardized scores of the FEELE subscale MS anger for female participants range from 20 to 44 (mean: 31.76, SD: 6.32).

Unstandardized scores of the FEELE subscale MS fear for the complete sample range from 14 to 44 (mean: 27.37, SD: 5.9). Unstandardized scores of the FEELE subscale MS fear for male participants range from 16 to 44 (mean: 28.47, SD: 5.7). Unstandardized scores of the FEELE subscale MS fear for female participants range from 14 to 36 (mean: 26.22, SD: 5.96).

Unstandardized scores of the FEELE subscale MS sadness for the complete sample range from 12 to 5507 (mean: 30.71, SD: 6.93). Unstandardized scores of the FEELE subscale MS sadness for male participants range from 18 to 5507 (mean: 31.14, SD: 6.14). Unstandardized scores of the FEELE subscale MS sadness for female participants range from 12 to 42 (mean:30.27, SD: 7.72).

Unstandardized scores of the FEELE subscale withdrawal for the complete sample range from 6 to 24 (mean: 15.33, SD: 4.76). Unstandardized scores of the FEELE subscale withdrawal for male participants range from 8 to 23 (mean: 15.51, SD: 4.17. Unstandardized scores of the FEELE subscale withdrawal for female participants range from 6 to 24 (mean: 15.15, SD: 5.36).

Unstandardized scores of the FEELE subscale self-devaluation for the complete sample range from 10 to 26 (mean: 18.76, SD: 3.83). Unstandardized scores of the FEELE subscale self-devaluation for male participants range from 10 to 26 (mean: 19.49, SD: 3.63). Unstandardized scores of the FEELE subscale self-devaluation for female participants range from 10 to 26 (mean: 18, SD: 3.94).

Unstandardized scores of the FEELE subscale giving up for the complete sample range from 6 to 22 (mean: 13.83, SD: 4.06). Unstandardized scores of the FEELE subscale giving up for male participants range from 8 to 22 (mean: 14.02, SD: 3.68). Unstandardized scores of the FEELE subscale giving up for female participants range from 6 to 22 (mean: 13.63, SD: 4.47).

Unstandardized scores of the FEELE subscale perseveration for the complete sample range from 7 to 29 (mean: 19.46, SD: 4.6). Unstandardized scores of the FEELE subscale perseveration for male participants range from 9 to 29 (mean: 19.09, SD: 3.71). Unstandardized scores of the FEELE subscale perseveration for female participants range from 7 to 27 (mean: 19.85, SD: 5.39).

Unstandardized scores of the FEELE subscale catastrophizing for the complete sample range from 6 to 22 (mean: 8.5, SD: 3.53). Unstandardized scores of the FEELE subscale catastrophizing for male participants range from 6 to 20 (mean: 8.37, SD: 3.15). Unstandardized scores of the FEELE subscale catastrophizing for female participants range from 6 to 22 (mean: 8.63, SD: 3.93).

Unstandardized scores of the FEELE subscale blaming others for the complete sample range from 6 to 23 (mean: 13.25, SD: 3.55). Unstandardized scores of the FEELE subscale blaming others for male participants range from 7 to 23 (mean: 13.51, SD: 3.62). Unstandardized scores of the FEELE subscale blaming others for female participants range from 6 to 21 (mean: 12.98, SD: 3.5).

3.2.1.7 Skalen zum Erleben von Emotionen (SEE)

Unstandardized scores of the SEE subscale acceptance of own emotions for the complete sample range from 14 to 29 (mean: 23.54, SD: 3.31). Unstandardized scores of the SEE subscale acceptance of own emotions for male participants range from 14 to 29 (mean: 23, SD: 3.24). Unstandardized scores of the SEE subscale acceptance of own emotions for female participants range from 17 to 29 (mean: 24.1, SD: 3.32).

Unstandardized scores of the SEE subscale experiencing emotional flooding for the complete sample range from 7 to 30 (mean: 17.11, SD: 4.67). Unstandardized scores of the SEE subscale experiencing emotional flooding for male participants range from 7 to 30 (mean: 16.74, SD: 4.5). Unstandardized scores of the SEE subscale experiencing emotional flooding for female participants range from 9 to 28 (mean: 17.49, SD: 4.87).

Unstandardized scores of the SEE subscale experiencing lack of emotions for the complete sample range from 5 to 22 (mean: 11.23, SD: 3.72). Unstandardized scores of the SEE subscale experiencing lack of emotions for male participants range from 8 to 22 (mean: 12.51, SD: 3.47). Unstandardized scores of the SEE subscale experiencing lack of emotions for female participants range from 5 to 22 (mean: 9.88, SD: 3.52).

Unstandardized scores of the SEE subscale body-focused symbolizing of emotions for the complete sample range from 9 to 36 (mean: 24.1, SD: 5.94). Unstandardized scores of the SEE subscale body-focused symbolizing of emotions for male participants range from 9 to 33 (mean: 22.05, SD: 5.56). Unstandardized scores of the SEE subscale body-focused symbolizing of emotions for female participants range from 9 to 36 (mean: 26.24, SD: 5.61).

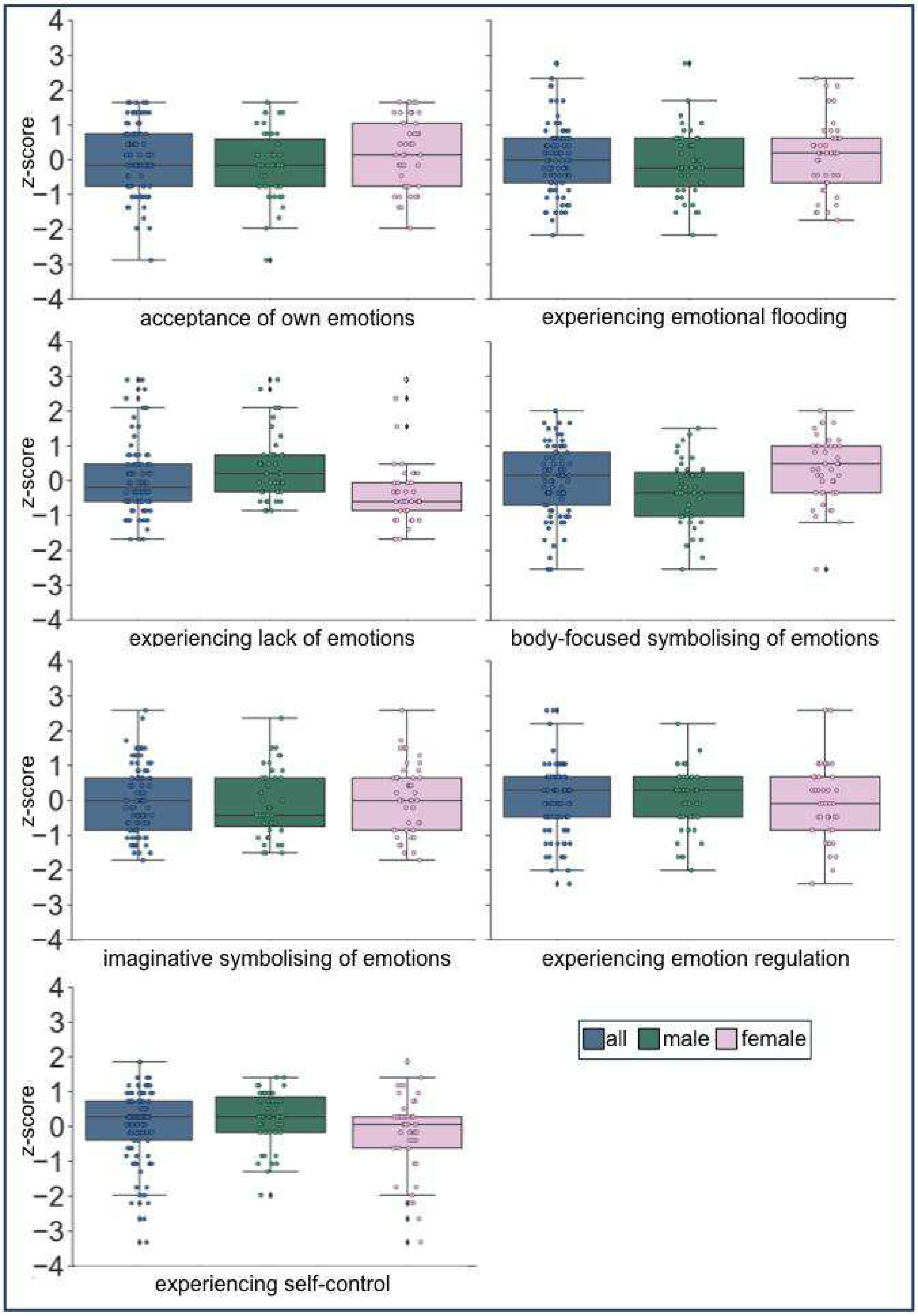


Figure 15. **Z-standardized scores of the SEE.**

Unstandardized scores of the SEE subscale imaginative symbolizing of emotions for the complete sample range from 7 to 27 (mean: 15.01, SD: 4.654). Unstandardized scores of the SEE subscale imaginative symbolizing of emotions for male participants range from 8 to 26 (mean: 14.65, SD: 4.56). Unstandardized scores of the SEE subscale imaginative symbolizing of emotions for female participants range from 7 to 27 (mean: 15.39, SD: 4.77).

Unstandardized scores of the SEE subscale experiencing emotion regulation for the complete sample range from 7 to 20 (mean: 13.25, SD: 2.62). Unstandardized scores of the SEE subscale experiencing emotion regulation for male participants range from 8 to 19 (mean: 13.35, SD: 2.38). Unstandardized scores of the SEE subscale experiencing emotion regulation for female participants range from 7 to 20 (mean: 13.15, SD: 2.87).

Unstandardized scores of the SEE subscale experiencing self-control for the complete sample range from 7 to 30 (mean: 21.75, SD: 4.44). Unstandardized scores of the SEE subscale experiencing self-control for male participants range from 13 to 28 (mean: 22.77, SD: 3.36). Unstandardized scores of the SEE subscale experiencing self-control for female participants range from 7 to 30 (mean: 20.68, SD:5.03).

3.2.1.8 Emotionale-Kompetenz-Fragebogen Selbstauskunft (EKF-S)

Unstandardized scores of the EKFS subscale recognizing and understanding own emotions for the complete sample range from -2.13 to 2.43 (mean: 0.51, SD: 1). Unstandardized scores of the EKFS subscale recognizing and understanding own emotions for male participants range from -2.13 to 2.43 (mean: 0.38, SD: 0.95). Unstandardized scores of the EKFS subscale recognizing and understanding own emotions for female participants range from -1.73 to 2.43 (mean: 0.64, SD: 1.05).

Unstandardized scores of the EKFS subscale recognizing others' emotions for the complete sample range from -2.41 to 2.09 (mean: 0.14, SD: 1.03). Unstandardized scores of the EKFS subscale recognizing others' emotions for male participants range from -2.41 to 1.08 (mean: -0.13, SD: 1.06). Unstandardized scores of the EKFS subscale recognizing others' emotions for female participants range from -1.97 to 2.09 (mean: 0.41, SD: 0.93).

Unstandardized scores of the EKFS subscale regulation and control of own emotions for the complete sample range from -2.06 to 2.61 (mean: 0.68, SD: 0.94). Unstandardized scores of the EKFS subscale regulation and control of own emotions for male participants range from -0.62 to 2.49 (mean: 0.87, SD: 0.69). Unstandardized scores of the EKFS subscale regulation and control of own emotions for female participants range from -2.06 to 2.61 (mean: 0.5, SD: 1.12).

Unstandardized scores of the EKFS subscale emotional expressivity for the complete sample range from -1.95 to 1.89 (mean: 0.01, SD: 0.97). Unstandardized scores of the

EKFS subscale emotional expressivity for male participants range from -1.95 to 1.57 (mean: -0.33, SD: 0.86). Unstandardized scores of the EKFS subscale emotional expressivity for female participants range from -1.71 to 1.89 (mean: 0.35, SD: 0.98).

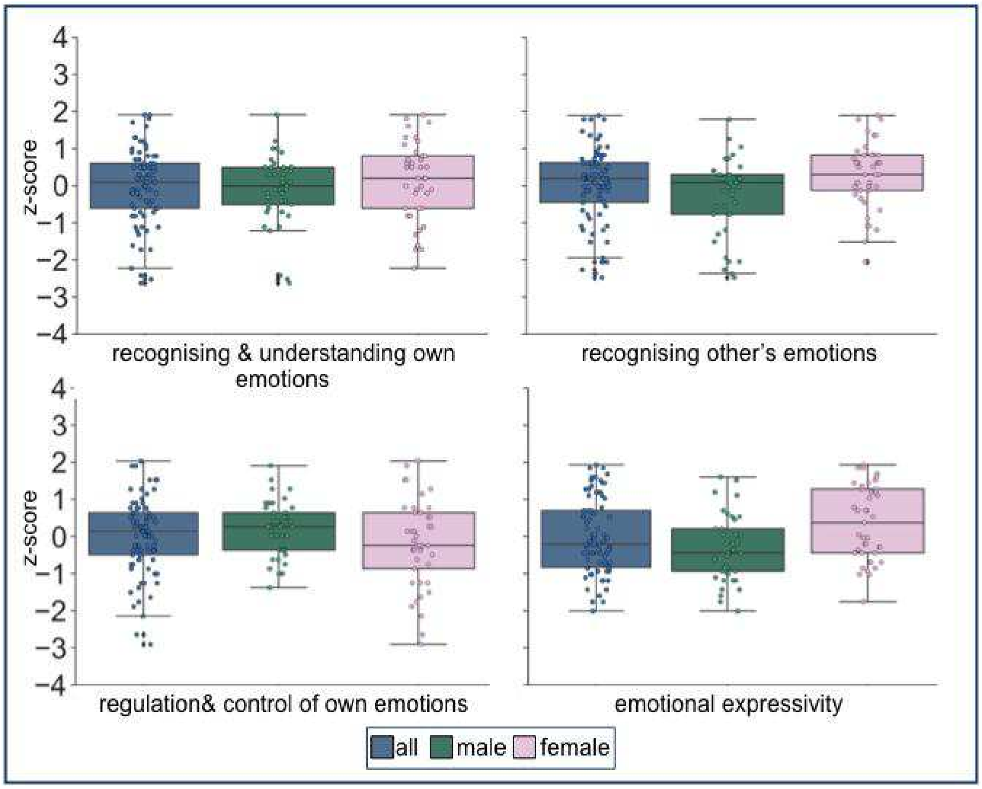


Figure 16. **Z-standardized scores of the EKFS.**

3.2.1.9 Affective Neuroscience Personality Scales (ANPS)

Unstandardized scores of the ANPS subscale SEEKING for the complete sample range from 30 to 53 (mean: 42.31, SD: 4.7). Unstandardized scores of the ANPS subscale SEEKING for male participants range from 30 to 52 (mean: 42.19, SD: 4.46). Unstandardized scores of the ANPS subscale SEEKING for female participants range from 33 to 53 (mean: 42.44, SD: 5).

Unstandardized scores of the ANPS subscale CARE for the complete sample range from 30 to 52 (mean: 41.43, SD: 5.33). Unstandardized scores of the ANPS subscale CARE for male participants range from 30 to 50 (mean: 39.53, SD: 5.27. Unstandardized scores of the ANPS subscale CARE for female participants range from 33 to 52 (mean: 43.41, SD: 4.67).

Unstandardized scores of the ANPS subscale PLAY for the complete sample range from 26 to 54 (mean: 43.35, SD: 6.06). Unstandardized scores of the ANPS subscale PLAY for male participants range from 27 to 53 (mean: 42.58, SD: 5.5). Unstandardized scores of the ANPS subscale PLAY for female participants range from 26 to 54 (mean: 44.15, SD: 6.56).

Unstandardized scores of the ANPS subscale FEAR for the complete sample range from 17 to 50 (mean: 34.56, SD: 6.31). Unstandardized scores of the ANPS subscale FEAR for male participants range from 17 to 49 (mean: 33.65, SD: 6.02). Unstandardized scores of the ANPS subscale FEAR for female participants range from 24 to 50 (mean: 35.49, SD: 6.54).

Unstandardized scores of the ANPS subscale ANGER for the complete sample range from 18 to 52 (mean: 33.08, SD: 6.53). Unstandardized scores of the ANPS subscale ANGER for male participants range from 18 to 43 (mean: 32.26, SD: 5.74). Unstandardized scores of the ANPS subscale ANGER for female participants range from 21 to 52 (mean: 33.95, SD: 7.23).

Unstandardized scores of the ANPS subscale SADNESS for the complete sample range from 23 to 43 (mean: 31.88, SD: 4.25). Unstandardized scores of the ANPS subscale SADNESS for male participants range from 23 to 40 (mean: 30.6, SD: 4.02). Unstandardized scores of the ANPS subscale SADNESS for female participants range from 23 to 43 (mean: 33.22, SD: 4.12).

Unstandardized scores of the ANPS subscale spirituality for the complete sample range from 13 to 42 (mean: 25.06, SD: 6.43). Unstandardized scores of the ANPS subscale spirituality for male participants range from 13 to 33 (mean: 23.42, SD: 5.36). Unstandardized scores of the ANPS subscale spirituality for female participants range from 15 to 42 (mean: 26.78, SD: 7.06).

Unstandardized scores of the ANPS subscale lie for the complete sample range from 7 to 18 (mean: 10.92, SD: 1.97). Unstandardized scores of the ANPS subscale lie for male participants range from 7 to 15 (mean: 10.77, SD: 1.78). Unstandardized scores of the ANPS subscale lie for female participants range from 7 to 18 (mean: 11.07, SD: 2.16).

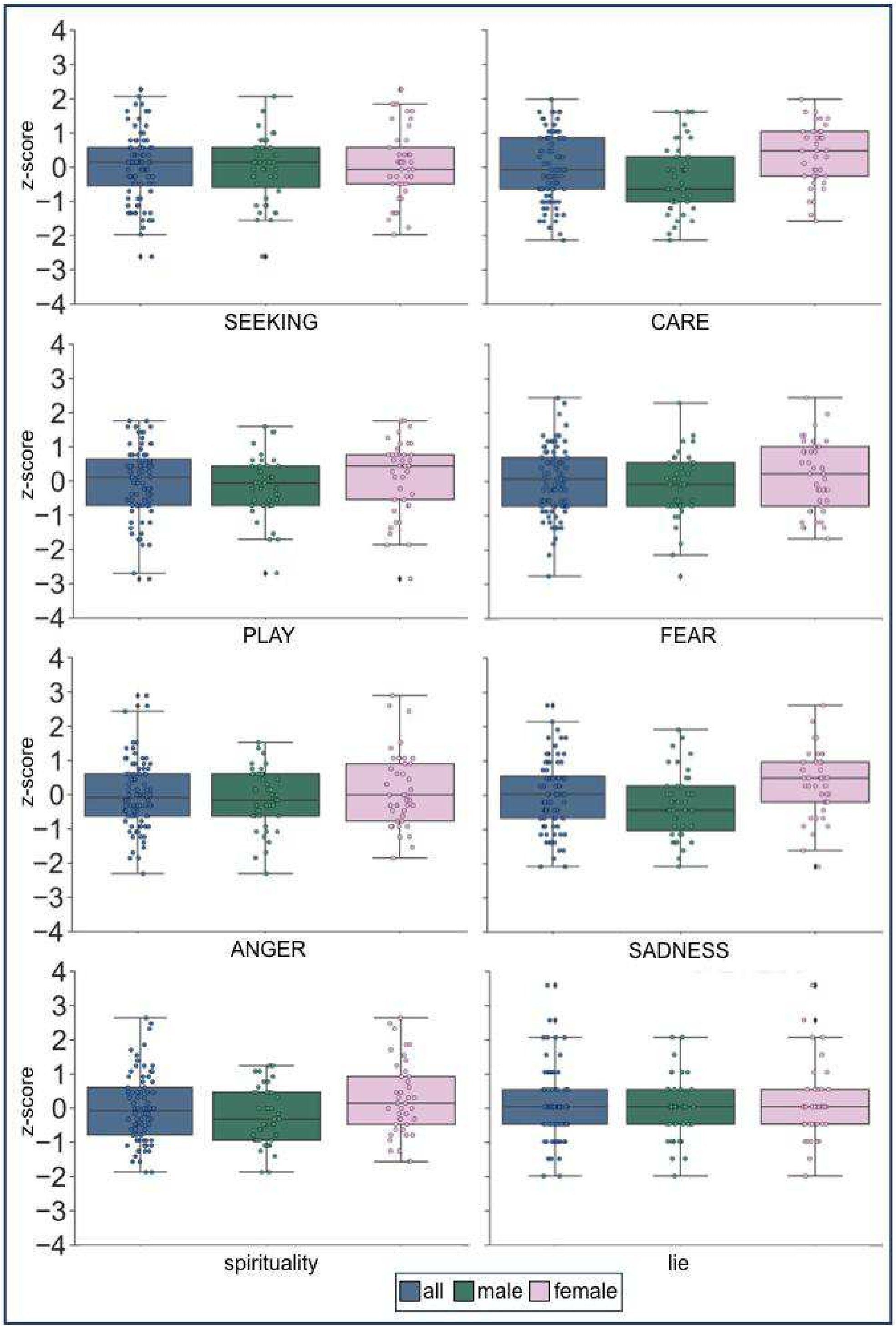


Figure 17. **Z-standardized scores of the ANPS.**

3.2.1.10 Positive and Negative Affect Scales (PANAS)

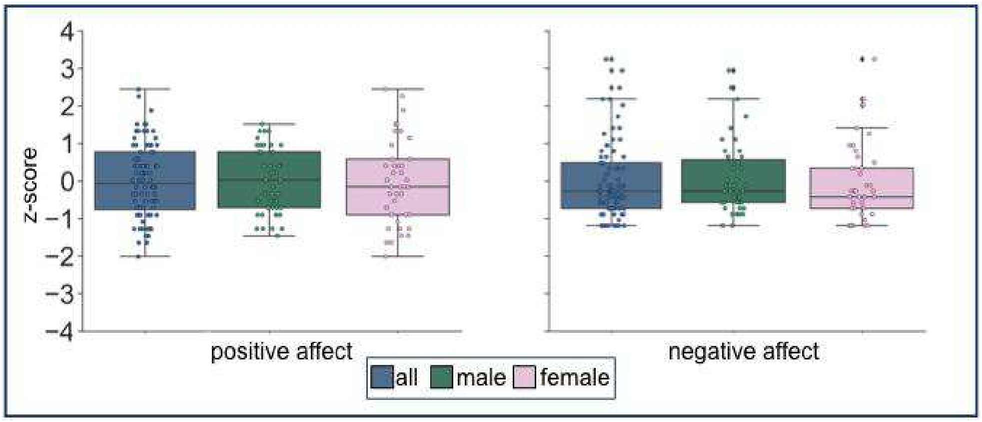


Figure 18. **Z-standardized scores of the PANAS.**

Unstandardized scores of the PANAS subscale positive affect for the complete sample range from 25 to 49 (mean: 35.85, SD: 5.38). Unstandardized scores of the PANAS subscale positive affect for male participants range from 28 to 44 (mean: 35.81, SD: 4.66). Unstandardized scores of the PANAS subscale positive affect for female participants range from 25 to 49 (mean: 35.88, SD: 6.11).

Unstandardized scores of the PANAS subscale negative affect for the complete sample range from 10 to 39 (mean: 17.77, SD: 6.54). Unstandardized scores of the PANAS subscale negative affect for male participants range from 10 to 37 (mean: 18.3, SD: 6.35). Unstandardized scores of the PANAS subscale negative affect for female participants range from 10 to 39 (mean: 17.22, SD: 6.75).

3.2.1.11 State-Trait-Ärgerausdrucks-Inventar (STAXI)

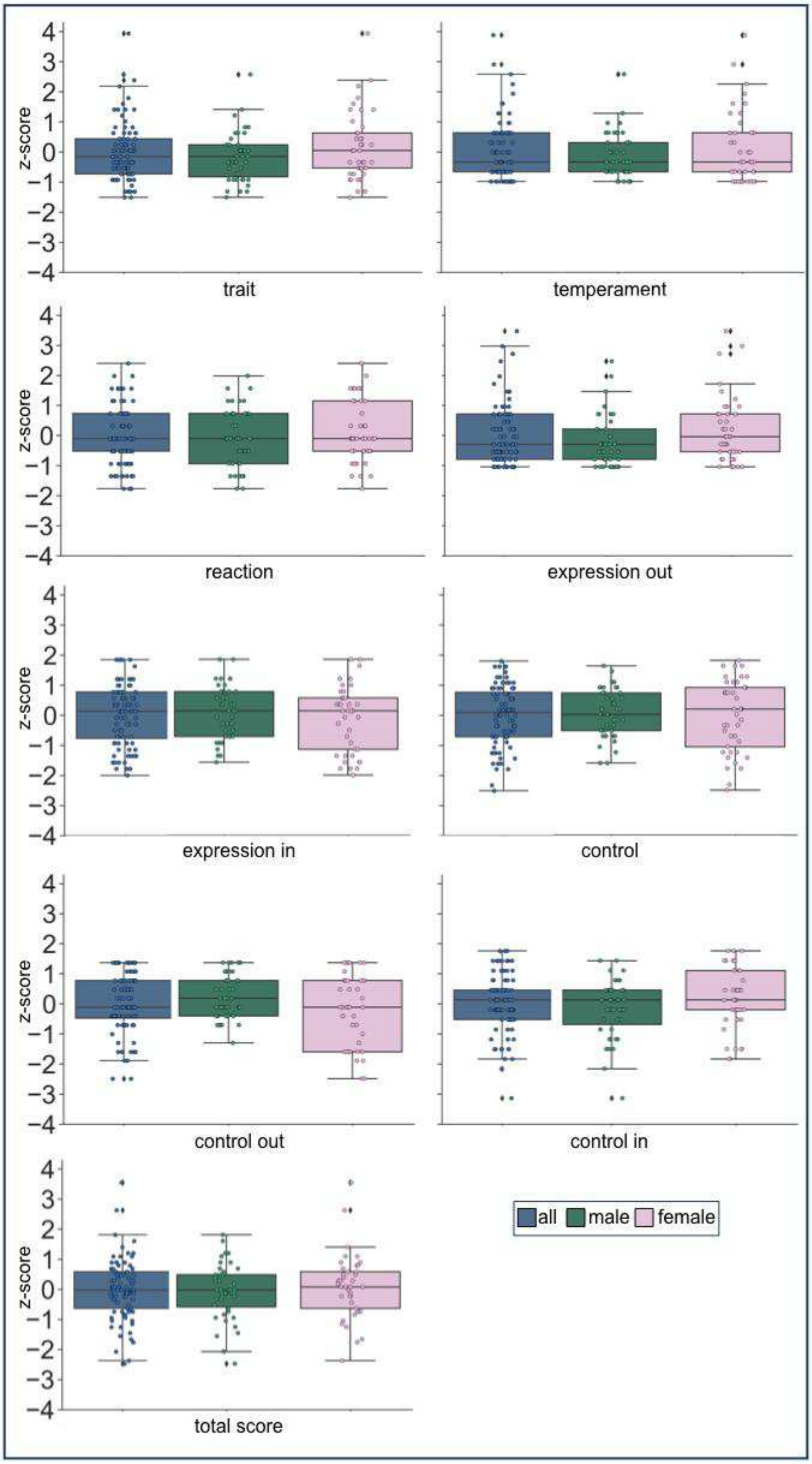


Figure 19. **Z-standardized scores of the STAXI.**

Unstandardized scores of the STAXI subscale trait for the complete sample range from 12 to 40 (mean: 19.77, SD: 5.14). Unstandardized scores of the STAXI subscale trait for male participants range from 12 to 33 (mean: 18.95, SD: 4.21). Unstandardized scores of the STAXI subscale trait for female participants range from 12 to 40 (mean: 20.63, SD: 5.89).

Unstandardized scores of the STAXI subscale temperament for the complete sample range from 5 to 20 (mean: 8.04, SD: 3.08). Unstandardized scores of the STAXI subscale temperament for male participants range from 5 to 16 (mean: 7.56, SD: 2.36). Unstandardized scores of the STAXI subscale temperament for female participants range from 5 to 20 (mean: 8.54, SD: 3.65).

Unstandardized scores of the STAXI subscale anger reaction for the complete sample range from 6 to 16 (mean: 10.25, SD: 2.4). Unstandardized scores of the STAXI subscale anger reaction for male participants range from 6 to 15 (mean: 10, SD: 2.35). Unstandardized scores of the STAXI subscale anger reaction for female participants range from 6 to 16 (mean: 10.51, SD: 2.45).

Unstandardized scores of the STAXI subscale expression out for the complete sample range from 8 to 26 (mean: 12.17, SD: 3.99). Unstandardized scores of the STAXI subscale expression out for male participants range from 8 to 22 (mean: 11.44, SD: 3.45). Unstandardized scores of the STAXI subscale expression out for female participants range from 8 to 26 (mean: 12.93, SD: 4.4).

Unstandardized scores of the STAXI subscale expression in for the complete sample range from 8 to 26 (mean: 17.37, SD: 4.68). Unstandardized scores of the STAXI subscale expression in for male participants range from 10 to 26 (mean: 17.86, SD: 4.15). Unstandardized scores of the STAXI subscale expression in for female participants range from 8 to 26 (mean: 16.85, SD: 5.18).

Unstandardized scores of the STAXI subscale anger control for the complete sample range from 16 to 40 (mean: 29.99, SD: 5.57). Unstandardized scores of the STAXI subscale anger control for male participants range from 21 to 39 (mean: 30.28, SD: 4.39). Unstandardized scores of the STAXI subscale anger control for female participants range from 16 to 40 (mean: 29.68, SD: 6.63).

Unstandardized scores of the STAXI subscale anger control out for the complete sample range from 7 to 20 (mean: 15.38, SD: 3.37). Unstandardized scores of the STAXI subscale anger control out for male participants range from 11 to 20 (mean: 16.16, SD: 2.46). Unstandardized scores of the STAXI subscale anger control out for female participants range from 7 to 20 (mean: 14.56, SD: 3.99).

Unstandardized scores of the STAXI subscale anger control in for the complete sample range from 5 to 20 (mean: 14.61, SD: 3.06). Unstandardized scores of the STAXI subscale anger control in for male participants range from 5 to 19 (mean: 14.12, SD: 2.95). Unstandardized scores of the STAXI subscale anger control in for female participants range from 9 to 20 (mean: 15.12, SD: 3.13).

Unstandardized scores of the STAXI total score for the complete sample range from 55 to 114 (mean: 79.3, SD: 9.8). Unstandardized scores of the STAXI total score for male participants range from 55 to 97 (mean: 78.53, SD: 8.99). Unstandardized scores of the STAXI total score for female participants range from 56 to 114 (mean: 80.1, SD: 10.63).

3.2.1.12 Inventar sozialer Kompetenzen Kurzform (ISK-K)

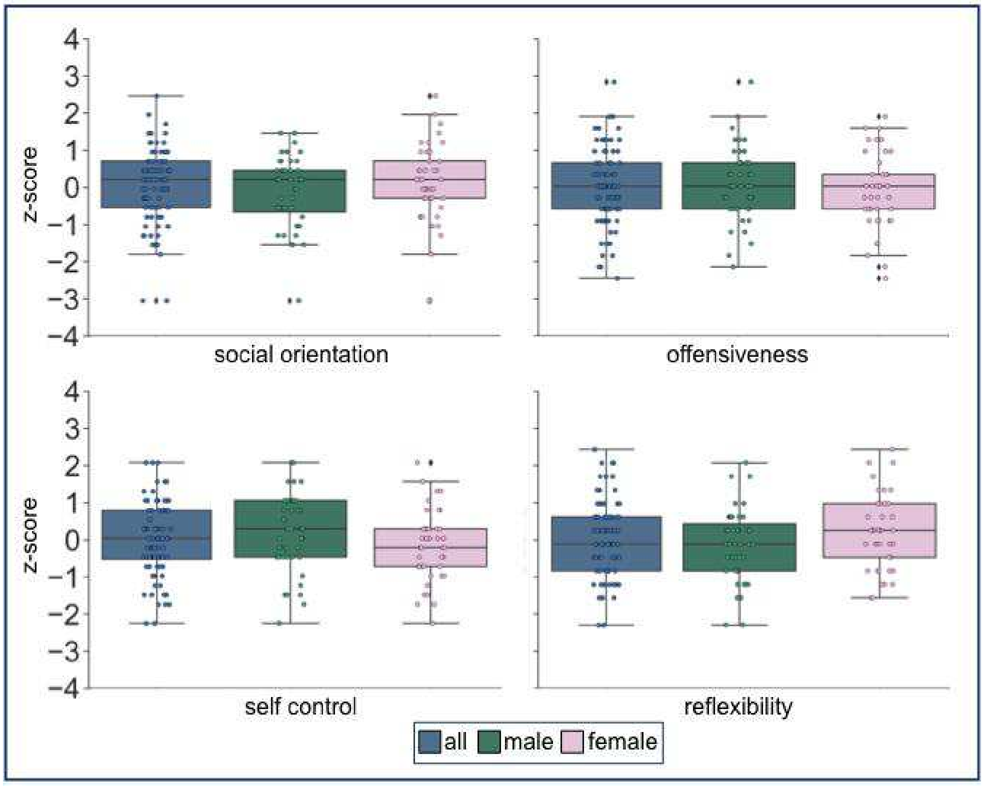


Figure 20. **Z-standardized scores of the ISK-K.**

Unstandardized scores of the ISK-K subscale social orientation for the complete sample range from 18 to 40 (mean: 30.18, SD: 3.99). Unstandardized scores of the ISK-K subscale social orientation for male participants range from 18 to 36 (mean: 29.72, SD: 3.83). Unstandardized scores of the ISK-K subscale social orientation for female participants range from 18 to 40 (mean: 30.66, SD: 4.16).

Unstandardized scores of the ISK-K subscale offensiveness for the complete sample range from 14 to 31 (mean: 21.88, SD: 3.22). Unstandardized scores of the ISK-K subscale offensiveness for male participants range from 15 to 31 (mean: 22.16, SD: 3.17). Unstandardized scores of the ISK-K subscale offensiveness for female participants range from 14 to 28 (mean: 21.59, SD: 3.28).

Unstandardized scores of the ISK-K subscale self-control for the complete sample range from 15 to 32 (mean: 23.85, SD: 3.94). Unstandardized scores of the ISK-K subscale self-control for male participants range from 15 to 32 (mean: 24.47, SD: 4.07). Unstandardized scores of the ISK-K subscale self-control for female participants range from 15 to 32 (mean: 23.2, SD: 3.74).

Unstandardized scores of the ISK-K subscale reflexibility for the complete sample range from 14 to 27 (mean: 20.31, SD: 2.75). Unstandardized scores of the ISK-K subscale reflexibility for male participants range from 14 to 26 (mean: 19.84, SD: 2.69). Unstandardized scores of the ISK-K subscale reflexibility for female participants range from 16 to 27 (mean: 20.8, SD: 2.76).

3.2.1.13 ADHS-Screening für Erwachsene (ADHS-E)

Unstandardized scores of the ADHS-E subscale emotion and affect for the complete sample range from 0 to 13 (mean: 4.62, SD: 2.54). Unstandardized scores of the ADHS-E subscale emotion and affect for male participants range from 0 to 8 (mean: 4.21, SD: 2.2). Unstandardized scores of the ADHS-E subscale emotion and affect for female participants range from 1 to 13 (mean: 5.05, SD: 2.82).

Unstandardized scores of the ADHS-E subscale attention control for the complete sample range from 0 to 11 (mean: 5.08, SD: 2.95). Unstandardized scores of the ADHS-E subscale attention control for male participants range from 0 to 11 (mean: 6.14, SD: 2.77). Unstandardized scores of the ADHS-E subscale attention control for female participants range from 0 to 10 (mean: 3.98, SD: 2.74).

Unstandardized scores of the ADHS-E subscale restlessness and hyperactivity for the complete sample range from 0 to 14 (mean: 5.31, SD: 2.76). Unstandardized scores of the ADHS-E subscale restlessness and hyperactivity for male participants range from 1 to 14 (mean: 5.3, SD: 2.7). Unstandardized scores of the ADHS-E subscale restlessness and hyperactivity for female participants range from 0 to 11 (mean: 5.32, SD: 2.86).

Unstandardized scores of the ADHS-E subscale impulse control and disinhibition for the complete sample range from 0 to 8 (mean: 3.43, SD: 1.9). Unstandardized scores of the ADHS-E subscale impulse control and disinhibition for male participants range from 0 to 8 (mean: 3.33, SD: 1.66). Unstandardized scores of the ADHS-E subscale impulse control and disinhibition for female participants range from 0 to 8 (mean: 3.54, SD: 2.13).

Unstandardized scores of the ADHS-E subscale stress intolerance for the complete sample range from 0 to 10 (mean: 4.38, SD: 2.17). Unstandardized scores of the ADHS-E subscale stress intolerance for male participants range from 0 to 10 (mean: 3.95, SD: 1.94). Unstandardized scores of the ADHS-E subscale stress intolerance for female participants range from 0 to 10 (mean: 4.83, SD: 2.33).

Unstandardized scores of the ADHS-E total score for the complete sample range from 8 to 46 (mean: 22.82 SD: 8.63). Unstandardized scores of the ADHS-E total score for male participants range from 8 to 45 (mean: 22.93, SD: 7.48). Unstandardized scores of the ADHS-E total score for female participants range from 8 to 46 (mean: 22.71, SD: 9.79).

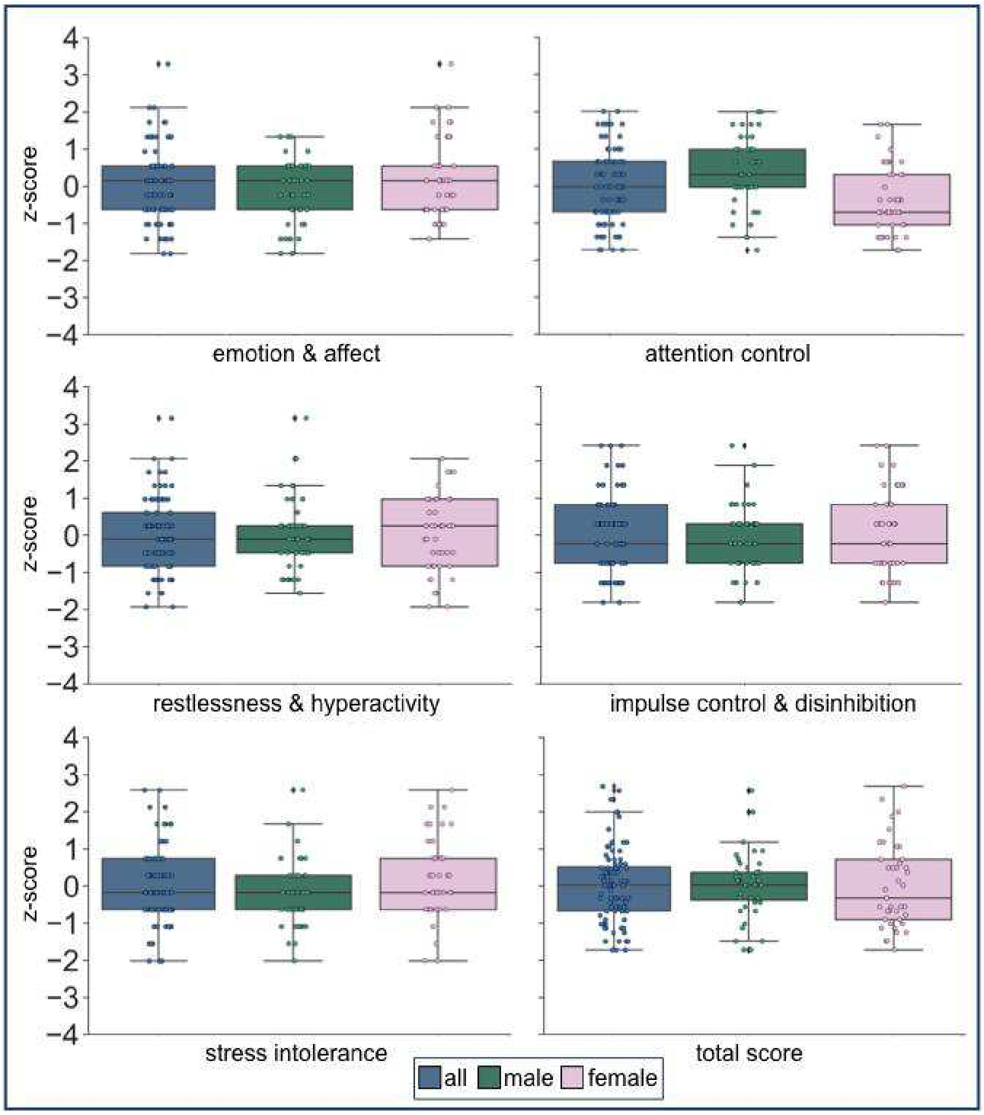


Figure 21. **Z-standardized scores of the ADHS-E.**

3.2.1.14 Schuhfried Theory of Mind (ToM)

Unstandardized scores of the ToM subscale sum wrongly answered control questions for the complete sample range from 0 to 1 (mean: 0.01, SD: 0.11). Unstandardized scores of the ToM subscale sum wrongly answered control questions for male participants range from 0 to 1 (mean: 0.02, SD: 0.15). Unstandardized scores of the ToM subscale sum wrongly answered control questions for female participants range from 0 to 0 (mean: 0, SD: 0).

Unstandardized scores of the ToM subscale implicit ToM for the complete sample range from 1 to 3 (mean: 2.81, SD: 0.42). Unstandardized scores of the ToM subscale implicit ToM for male participants range from 1 to 3 (mean: 2.86, SD: 0.41). Unstandardized scores of the

ToM subscale implicit ToM for female participants range from 2 to 3 (mean: 2.76, SD: 0.43).

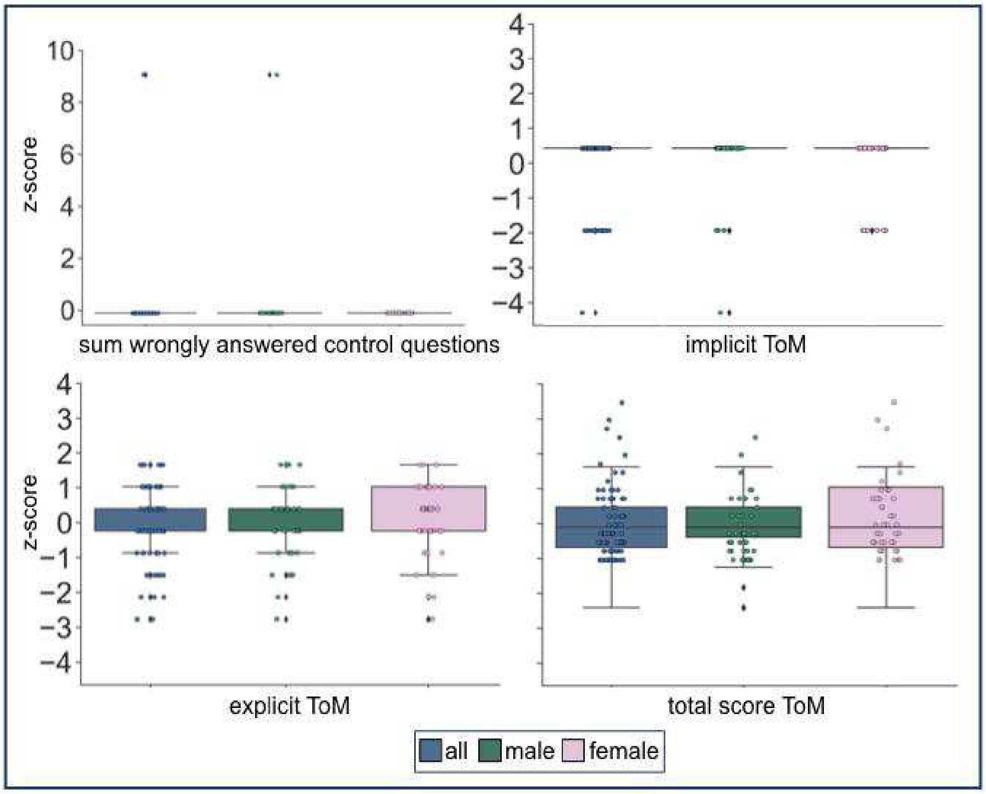


Figure 22. **Z-standardized scores of the ToM.**

Unstandardized scores of the ToM subscale explicit ToM for the complete sample range from 3 to 10 (mean: 7.38, SD: 1.58). Unstandardized scores of the ToM subscale explicit ToM for male participants range from 3 to 10 (mean: 7.35, SD: 1.56). Unstandardized scores of the ToM subscale explicit ToM for female participants range from 3 to 10 (mean: 7.41, SD: 1.63).

Unstandardized scores of the ToM total score for the complete sample range from 6 to 13 (mean: 10.19, SD: 1.74). Unstandardized scores of the ToM total score for male participants range from 6 to 13 (mean: 10.21, SD: 1.71). Unstandardized scores of the ToM total score for female participants range from 6 to 13 (mean:10.176, SD: 1.79).

3.2.1.15 State questionnaires pre MRI measurement

Unstandardized scores of the state BDI for the complete sample range from 0 to 16 (mean: 2.99, SD: 3.68). Unstandardized scores of the state BDI for male participants range from 0 to 16 (mean: 3.21, SD: 4.03). Unstandardized scores of the state BDI for female participants range from 0 to 14 (mean: 2.76, SD: 3.31).

Unstandardized scores of the state STAI for the complete sample range from 20 to 47 (mean: 29.76, SD: 5.92). Unstandardized scores of the state STAI for male participants range from 21 to 46 (mean: 29.93, SD: 5.35). Unstandardized scores of the state STAI for female participants range from 20 to 47 (mean: 29.59, SD: 6.53).

Unstandardized scores of the STAXI subscale state anger for the complete sample range from 15 to 35 (mean: 15.58, SD: 2.48). Unstandardized scores of the STAXI subscale state anger for male participants range from 15 to 35 (mean: 15.6, SD: 3.06). Unstandardized scores of the STAXI subscale state anger for female participants range from 15 to 24 (mean: 15.56, SD: 1.7).

Unstandardized scores of the STAXI subscale feeling of anger for the complete sample range from 5 to 16 (mean: 5.38, SD: 1.5). Unstandardized scores of the STAXI subscale feelings of anger for male participants range from 5 to 16 (mean: 5.35, SD: 1.7). Unstandardized scores of the STAXI subscale feelings of anger for female participants range from 5 to 12 (mean: 5.41, SD: 1.26).

Unstandardized scores of the STAXI subscale verbal anger impulse for the complete sample range from 5 to 12 (mean: 5.18, SD: 0.84). Unstandardized scores of the STAXI subscale verbal anger impulse for male participants range from 5 to 12 (mean: 5.21, SD: 1.08). Unstandardized scores of the STAXI subscale verbal anger impulse for female participants range from 5 to 7 (mean: 5.15, SD: 0.48).

Unstandardized scores of the STAXI subscale physical anger impulse for the complete sample range from 5 to 7 (mean: 5.02, SD: 0.22). Unstandardized scores of the STAXI subscale physical anger impulse for male participants range from 5 to 7 (mean: 5.05, SD: 0.3). Unstandardized scores of the STAXI subscale physical anger impulse for female participants range from 5 to 5 (mean: 5, SD: 0).

Unstandardized scores of the SAM subscale arousal for the complete sample range from 1 to 7 (mean: 2.34, SD: 1.24). Unstandardized scores of the SAM subscale arousal for male participants range from 1 to 7 (mean: 2.6, SD: 1.34). Unstandardized scores of the SAM subscale arousal for female participants range from 1 to 5 (mean: 2.07, SD: 1.08).

Unstandardized scores of the SAM subscale valence for the complete sample range from 2 to 9 (mean: 5.3, SD: 1.44). Unstandardized scores of the SAM subscale valence for male participants range from 2 to 8 (mean: 5.37, SD: 1.29). Unstandardized scores of the SAM subscale valence for female participants range from 2 to 9 (mean: 5.22, SD: 1.59).

Unstandardized scores of the SAM subscale dominance for the complete sample range from 3 to 9 (mean: 7.38, SD: 1.22). Unstandardized scores of the SAM subscale dominance for male participants range from 5 to 9 (mean: 7.35, SD: 1.19).

Unstandardized scores of the SAM subscale dominance for female participants range from 3 to 9 (mean: 7.41, SD: 1.26).

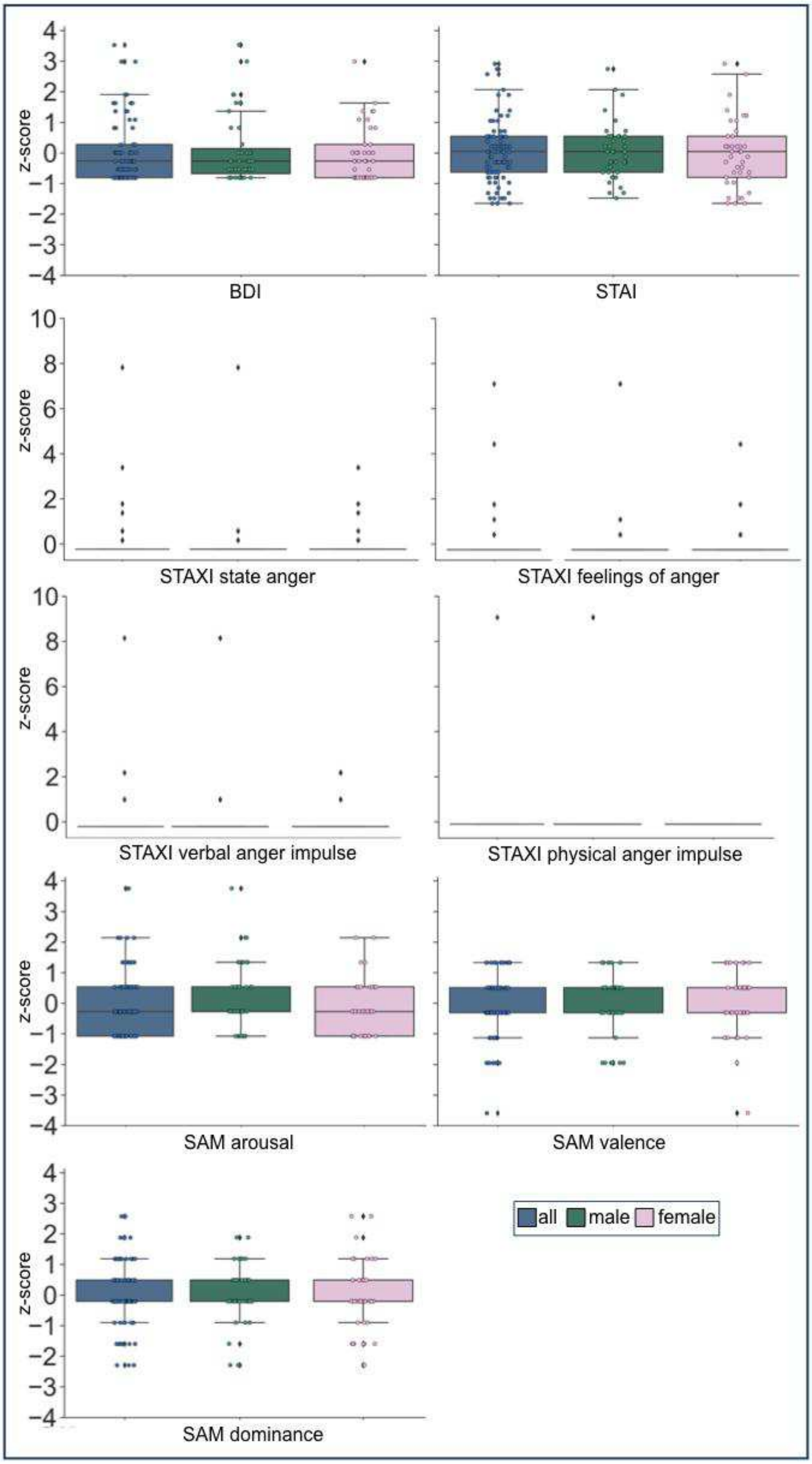


Figure 23. **Z-standardized scores of the state versions of the BDI, STAI, STAXI and SAM questionnaires assessed prior to MRI measurement.**

3.2.1.16 State questionnaires pre and post MRI measurement

Unstandardized scores of the PANAS subscale positive affect pre MRI measurement for the complete sample range from 19 to 46 (mean: 33.55, SD: 6.76). Unstandardized scores of the PANAS subscale positive affect pre MRI measurement for male participants range from 19 to 46 (mean: 34.12, SD: 5.82). Unstandardized scores of the PANAS subscale positive affect pre MRI measurement for female participants range from 19 to 46 (mean: 32.95, SD: 7.65). Unstandardized scores of the PANAS subscale positive affect post MRI measurement for the complete sample range from 13 to 46 (mean: 30.23, SD: 7.01). Unstandardized scores of the PANAS subscale positive affect post MRI measurement for male participants range from 16 to 43 (mean: 30.1, SD: 6.22). Unstandardized scores of the PANAS subscale positive affect post MRI measurement for female participants range from 13 to 46 (mean: 30.37, SD: 7.82). A paired samples t-test calculated on the full sample revealed a significant difference between the pre and post PANAS subscale positive affect ($t(82) = 5.18$, $p < .001$).

Unstandardized scores of the PANAS subscale negative affect pre MRI measurement for the complete sample range from 10 to 35 (mean: 12.67, SD: 3.69). Unstandardized scores of the PANAS subscale negative affect pre MRI measurement for male participants range from 10 to 35 (mean: 13.09, SD: 4.3). Unstandardized scores of the PANAS subscale negative affect pre MRI measurement for female participants range from 10 to 23 (mean: 12.22, SD: 2.9). Unstandardized scores of the PANAS subscale negative affect post MRI measurement for the complete sample range from 10 to 28 (mean: 12.55, SD: 3.69). Unstandardized scores of the PANAS subscale negative affect post MRI measurement for male participants range from 10 to 28 (mean: 12.43, SD: 4.07). Unstandardized scores of the PANAS subscale negative affect post MRI measurement for female participants range from 10 to 25 (mean: 12.68, SD: 3.31). A paired samples t-test calculated on the full sample revealed no significant difference between the pre and post PANAS subscale negative affect ($t(82) = 0.29$, $p = 0.776$).

Unstandardized scores of the DSSQ subscale engagement pre MRI measurement for the complete sample range from 16 to 32 (mean: 25.19, SD: 4.06). Unstandardized scores of the DSSQ subscale engagement pre MRI measurement for male participants range from 16 to 30 (mean: 24.86, SD: 3.47). Unstandardized scores of the DSSQ subscale engagement pre MRI measurement for female participants range from 16 to 32 (mean: 25.54, SD: 4.6). Unstandardized scores of the DSSQ subscale engagement post MRI measurement for the complete sample range from 10 to 32 (mean: 22.99, SD: 4.86).

Unstandardized scores of the DSSQ subscale engagement post MRI measurement for male participants range from 10 to 31 (mean: 22.26, SD: 4.69). Unstandardized scores of the DSSQ subscale engagement post MRI measurement for female participants range from 11 to 32 (mean: 23.73, SD: 4.96). A paired samples t-test calculated on the full sample revealed a significant difference between the pre and post DSSQ subscale engagement ($t(82) = 4.74$, $p < .001$).

Unstandardized scores of the DSSQ subscale distress pre MRI measurement for the complete sample range from 3 to 16 (mean: 7.8, SD: 3.21). Unstandardized scores of the DSSQ subscale distress pre MRI measurement for male participants range from 3 to 12 (mean: 7.35, SD: 2.64). Unstandardized scores of the DSSQ subscale distress pre MRI measurement for female participants range from 3 to 16 (mean: 8.27, SD: 3.7). Unstandardized scores of the DSSQ subscale distress post MRI measurement for the complete sample range from 3 to 23 (mean: 9.49, SD: 4.26). Unstandardized scores of the DSSQ subscale distress post MRI measurement for male participants range from 3 to 23 (mean: 9.45, SD: 3.82). Unstandardized scores of the DSSQ subscale distress post MRI measurement for female participants range from 3 to 23 (mean: 9.54, SD: 4.71). A paired samples t-test calculated on the full sample revealed a significant difference between the pre and post DSSQ subscale distress ($t(82) = -3.64$, $p < .001$).

Unstandardized scores of the DSSQ subscale worry pre MRI measurement for the complete sample range from 8 to 29 (mean: 17.76, SD: 5.6). Unstandardized scores of the DSSQ subscale worry pre MRI measurement for male participants range from 8 to 29 (mean: 17.47, SD: 5.79). Unstandardized scores of the DSSQ subscale worry pre MRI measurement for female participants range from 8 to 29 (mean: 18.07, SD: 5.45). Unstandardized scores of the DSSQ subscale worry post MRI measurement for the complete sample range from 8 to 28 (mean: 14.51, SD: 5.35). Unstandardized scores of the DSSQ subscale worry post MRI measurement for male participants range from 8 to 28 (mean: 15, SD: 5.32). Unstandardized scores of the DSSQ subscale worry post MRI measurement for female participants range from 8 to 25 (mean: 14, SD: 5.39). A paired samples t-test calculated on the full sample revealed a significant difference between the pre and post DSSQ subscale engagement ($t(82) = 6.09$, $p < .001$).

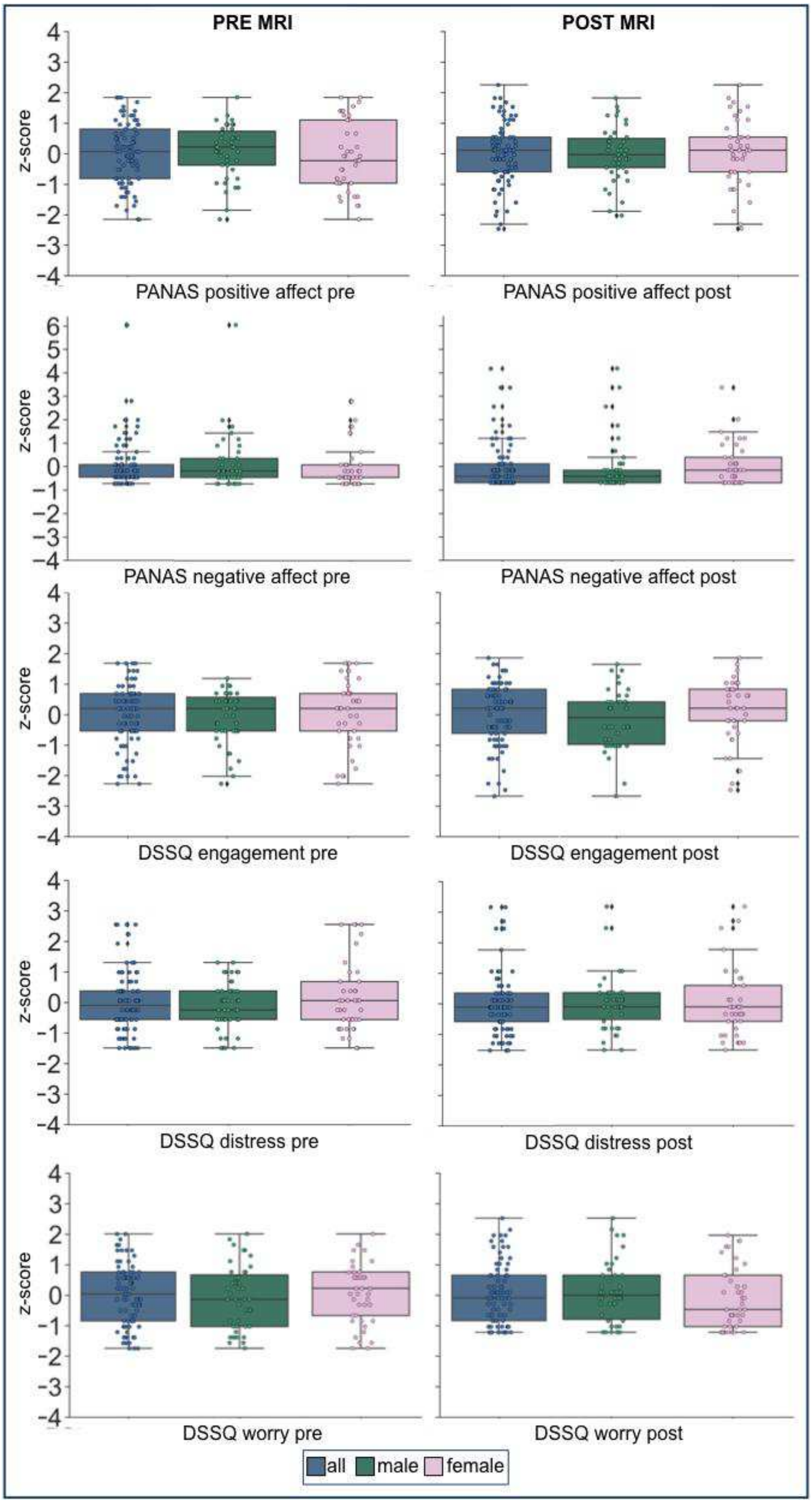


Figure 24 **Z-standardized scores of the state versions of the PANAS and DSSQ questionnaires assessed before and after MRI measurement.**

3.2.1.17 Movies-Questionnaire

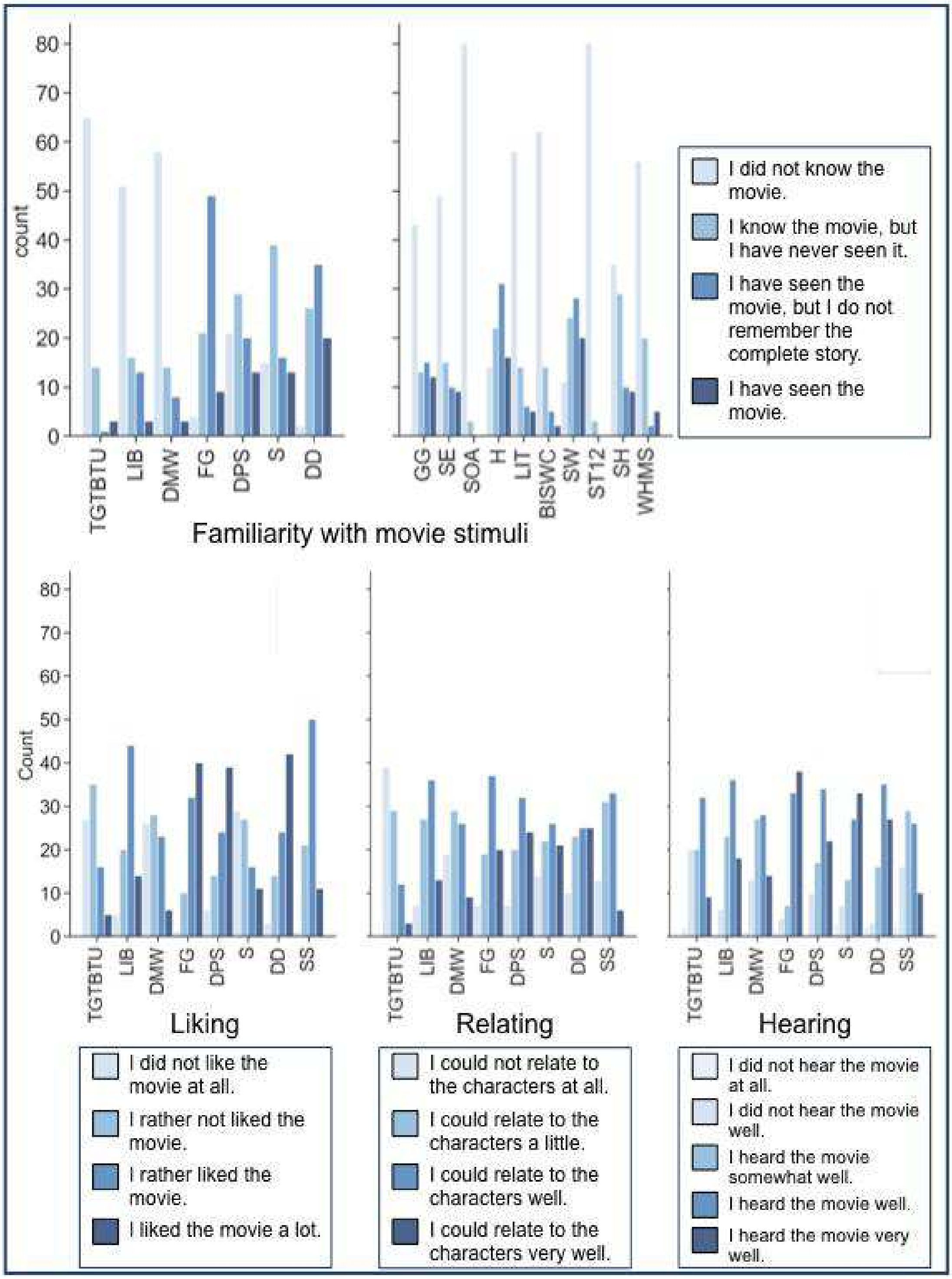


Figure 25. **Counts of participants that responded to questions about the familiarity and liking of the movie stimuli, how well they could relate to the characters and how well they could acoustically understand the movie stimuli.**

The Movies-questionnaires assessed various information about the viewer's knowledge and perception of the movie stimuli and their general movie or TV watching habits and preferences.

Figure 25 presents the counts of participants that reported on the familiarity with the movie stimulus (upper row), how they liked the movie stimuli (lower row, left figure), how well they could relate to the characters in the movie stimuli (lower row, middle figure), and how well they could acoustically understand the movie stimuli (lower row, right figure). Participants reported when they had last seen the movies that they were already familiar with, with dates ranging from days to years ago. Participants were asked whether and when they closed their eyes during the movie stimuli. While most participants reported not closing their eyes during the movie stimuli, those that looked away named the needle injection in DMW as the most common reason. Further reasons for closing their eyes or looking away were depictions of blood, rape or male homosexuality. Participants were also asked which character's position they could put themselves into best and worst. Responses varied greatly, but overall the characters from TGTBTU seem to have been the hardest to relate to.

The Movies questionnaire further assessed whether, and if yes, during which MRI scans, participants self-reported drowsing off or falling asleep. 4 participants reported drowsing or sleeping during the first resting state scan. 17 participants reported dowsing or sleeping during the anatomical T1 scan, 16 during the second resting state scan, none during the test movie scan, 7 during TGTBTU, 3 during LIB, 0 during DMW, 2 during FG, 1 during DPS, 0 during S, 2 during DD, 1 during SS and 50 participants reported never drowsing or falling asleep during any MRI measurement.

Additionally, the Movies questionnaire asked which movie stimulus participants perceived as most fitting to a series of adjectives. When asked which movie stimulus was the most exciting (orig. "am spannendsten"), 34 reported S, 16 reported DMW, 10 reported TGTBTU, 8 reported LIB, 8 reported SS, 5 reported FG, 2 reported DPS, and 0 reported DD. When asked which movie stimulus was most happy (orig. "am fröhlichsten"), 61 participants reported DD, 8 reported LIB, 8 reported FG, 4 reported SS, 2 reported DPS and none reported TGTBTU, DMW, or S. When asked which movie stimulus was most sad (orig. "am traurigsten"), 35 reported DMW, 29 reported LIB, 11 reported DPS, 5 reported FG, 2 reported SS, 1 reported DD and none reported TGTBTU or S. When asked which movie stimulus was most positive (orig. "am positivsten), 43 participants reported DD, 15 reported DPS, 13 reported FG, 9 reported LIB, 2 reported SS, 1 reported S and none reported TGTBTU or DMW. When asked which movie stimulus was most negative (orig. "am negativsten"), 56 participants reported DMW, 9 reported LIB, 9 reported S, 5 reported TGTBTU, 3 reported SS, 1 reported DPS, and none reported FG or DD. When asked which movie stimulus was the most inspiring (orig. "am inspirierendsten"), 34 participants reported DPS, 26 reported FG, 9 reported DD, 7 reported LIB, 3 reported SS, 2 reported DMW, and 1 each reported TGTBTU and S.

When asked which movie stimulus was the most disgusting (orig. “am ekligsten”), 52 participants reported DMW, 12 reported S, 11 reported SS, § reported each FG and LIB, 2 reported TGTBTU, and none reported DPS or DD. When asked which movie stimulus was the most entertaining (orig. “am unterhaltsamsten”), 33 participants reported FG, 24 reported DD, 7 reported SS, 6 reported each DPS and S, 3 reported each LIB and DMW, and 1 reported TGTBTU. When asked which movie stimulus was the most surprising (orig. “am überraschendsten”), 23 participants reported LIB, 20 reported SS, 12 each reported DPS and S, 6 reported DMW, 5 reported FG, 4 reported TGTBTU and 1 reported DD. When asked which movie stimulus had the most action (orig. “am actionreichsten”), 32 participants reported SS, 22 reported S, 16 reported TGTBTU, 4 each reported LIB and DMW, 3 reported DD, 2 reported FG and non reported DPS. When asked which movie stimulus was the scariest (orig. “am beängstigendsten”), 45 participants reported S, 27 reported DMW, 6 reported LIB, 3 reported SS, 2 reported FG and none reported TGTBTU, DPS or DD. When asked which movie stimulus was the most angering (orig. “am Wut erregendsten”), 47 participants reported DMW, 13 reported LIB, 9 reported S, 5 reported SS, 4 reported DPS, 3 reported FG and 1 each reported TGTBTU and DD. When asked which movie stimulus was the most boring (orig. “am langweiligsten”), 48 participants reported TGTBTU, 10 reported LIB, 8 reported DD, 6 reported FG, 5 reported SS, 4 reported DPS and 1 each reported DMW and S. Taken together, TGTBTU was the movie stimulus that was rated the most boring, LIB was the movie stimulus was rated the most surprising, DMW was rated the most sad, most negative, most disgusting and most angering, FG was rated the most entertaining, DPS was rated the most inspiring, S was rated the most exciting and most scary, DD was rated the happiest and most positive, and SS was rated as having the most action.

Furthermore, the Movies questionnaire assessed which genre of movies participants preferred the most. Multiple options were given, of which multiple could be picked, and additionally a blank field was presented for further remarks. 15 participants reported that horror movies are their preferred genre. 52 participants picked comedy as the preferred genre, 39 participants reported thriller, 26 reported romance, 36 reported drama, 53 reported action and adventure, 41 reported fantasy and science fiction, and 4 reported that they do not watch movies. In the field for additional remarks, 3 participants reported that they prefer documentaries, one added (anti-) war movies and anime, and another participants mentioned preferably watching comedy series.

The Movies questionnaire then asked participants about what aspects of movies they are paying most attention to while watching. Again, some options were given, with an additional free form field for more comments. Participants could pick multiple answers. 19 participants responded that they paid most attention to technical aspects such as cuts, camera settings, lighting and more. 35 participants responded with the soundtrack, 57 participants picked social interactions or relationships between characters, 29 participants picked places or landscapes, 24 participants picked closeness to reality or continuity (including aspects such as “plotholes”, errors in logic or presentation, and more), 46 picked content such as the plot or story, 9 picked actors/actresses (fame, which character is portrayed by which actor/actress, in which other productions did these actors/actresses appear). As additional remarks, one participant mentioned political messages or represented values, one mentioned the general mood of the movie, one added the meaning of the scene image, and one participant paid most attention to how complex technologies are and the social structures in the movies.

Afterwards, participants were asked to report how many hours per week they spent watching movies or series. 10 participants responded that they watch less than 1 hours per week, 18 participants reported 1 to 3 hours per week, 26 participants reported 3 to 6 hours per week, 17 participants reported 6 to 9 hours per week, 7 participants reported 9 to 12 hours per week, 4 participants reported 12 to 15 hours, 1 reported 15 hours and more, and one participants did not report any movies or series watching.

Last, participants were asked about the language in which they watch movies or series. 16 participants reported that they watch movies or series in the original language, if that language is English. 2 participants reported that they watch movies or series in the original language, even if it is not English. 28 reported that they watch movies or series partially in German, partially in other languages. 1 reported that they watch movies or series in other languages in German with German subtitles. 36 of participants reported that they watch movies or series only in German. 1 participant reported no movie or series watching.

3.2.1.18 Hormone sampling

Because the evaluation of the hormone samples is done by an external organization and samples are sent in for evaluation in bigger batches, evaluation has taken place only for a subsample of participants.

3.3 Annotation of emotions

3.3.1 Sample

The annotation of emotions was done by 44 participants (23 male, 21 female, age range: 20 - 30 years, mean: 25.05 years, SD: 3.23 years). Male participants were aged 20 - 30 years (mean age 25.22 years, SD 2.94 years). Female participants were aged 20 - 30 years (mean age 24.86 years, SD 3.6 years). Highest education was distributed among the sample as follows: 2 participants graduated from “Realschule” (1 male, 1 female), 10 participants completed vocational training (4 male, 6 female), 14 graduated from school with “Abitur” (4 male, 10 female) and 18 participants graduated from university (14 male, 4 female).

3.3.2 Annotation of emotions

Figure 26 shows the ratings of emotions induced by the movie stimuli of the Movies study over the course of each movie stimulus. The ratings were sampled every 100ms. The movie stimuli show distinct progressions and profiles of elicited emotions. DD shows continuously high ratings of joy and continuously lower ratings of surprise, while all other emotions hardly were perceived by viewers at all. Other movie stimuli induce a more complex profile of emotions. DMW induces medium high ratings of sadness throughout the stimulus, while anger and disgust increase rapidly in the second half of the stimulus. Both emotions show a parallel progression marked by a few distinct spikes. Fear and surprise are present at lower intensities continuously throughout the stimulus. DPS induced a profile of mixed emotions of low intensity. About two-thirds into the stimulus, sadness and anger spike and then decrease. During this decrease, surprise is felt with higher intensity and especially joy. FG shows three parts with two distinct profiles over the course of the stimulus: for about the first half of the stimulus, joy of moderate intensity is felt by viewers, accompanied by surprise of lower intensity. Then comes a period of more mixed emotions: anger and disgust spike simultaneously, followed by a spike in sadness, while joy was felt less intensely over that period, it increases again afterwards, repeating the pattern of the first half of the stimulus. LIB ratings include a mix of emotions of low intensity felt by viewers for the most part of the stimulus, with anger, fear and sadness being dominant. About a third into the stimulus, anger and sadness spiked. At the end of the stimulus, negative emotions dropped strongly in intensity, while joy increased to high intensity. S is dominated by fear throughout the stimulus, followed by anger. Other emotions were induced at lower intensities. SS was marked by rapid changes in emotion profiles,

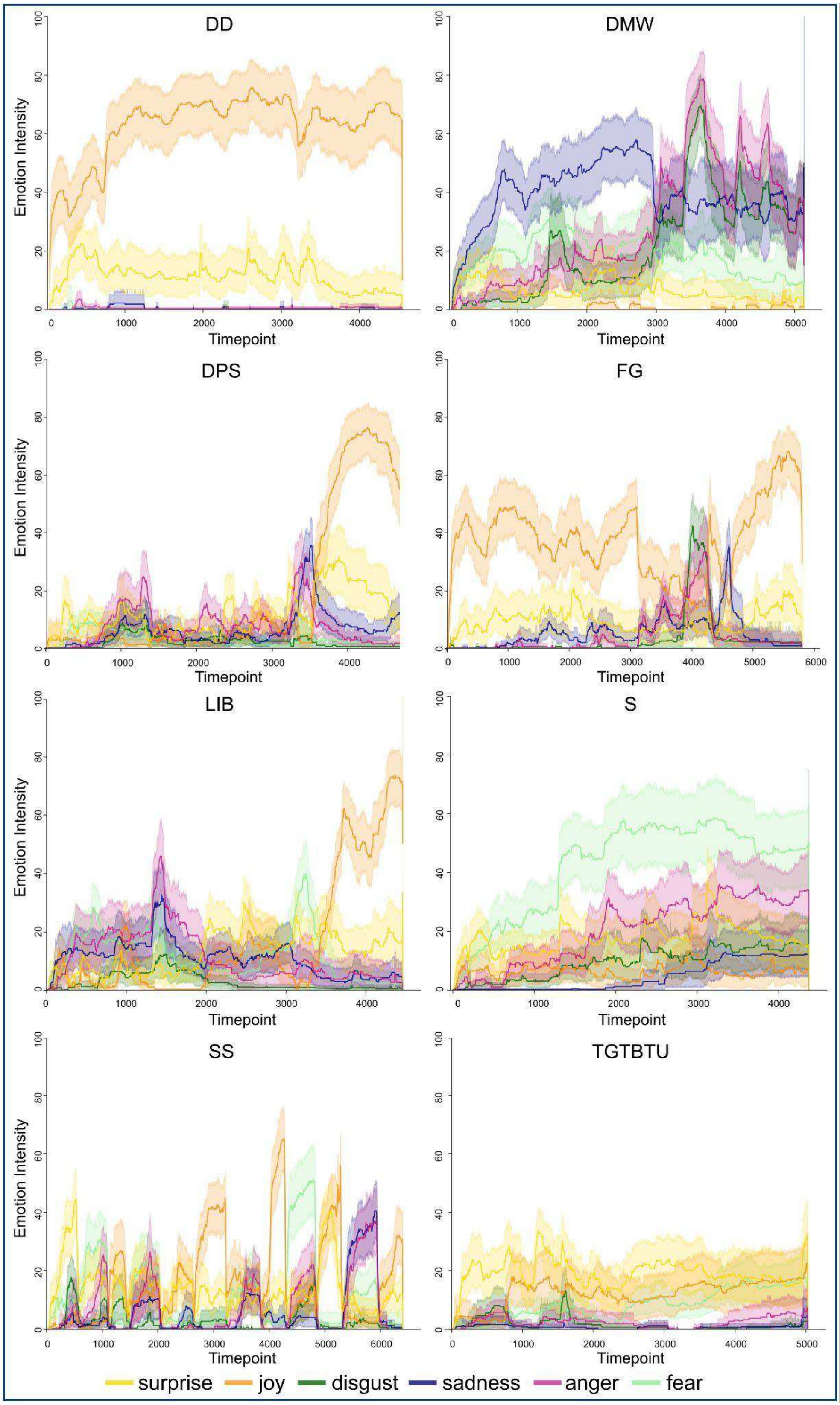


Figure 26. **Emotions induced by movie stimuli.** The x axis shows the timepoints of the emotion rating across each movie stimulus with each timepoint corresponding to 100ms. The y axis shows the perceived intensity of each emotion with 0 indicating absence of the emotion and 100 indicating emotion was felt at full intensity. DD = Dirty Dancing, DMW = Dead Men Walking, DPS = Dead

Poets Society, FG = Forrest Gump, LIB = Life is Beautiful, S = Scream, SS = short sequences, TGTBTU = The Good, the Bad, the Ugly.

indicating that each movie stimulus in the sequence induced different emotions. Different spiking emotions vary distinctly between the 12 short sequences. The emotions felt most intensely by order are: surprise, fear, joy, a mix with no clearly dominant emotion, joy and surprise, joy, again a mix without a dominant emotion, joy, fear, joy and surprise, anger and sadness, and joy. TGTBTU overall induced the least intense emotions with surprise and joy dominating throughout the stimulus. Fear increases in the second half of the stimulus.

Confidence intervals indicating the similarities in rating among viewers vary between movie stimuli. Across movies, there is a trend towards usually more variation in the perceived intensity of emotions among viewers when an emotion was perceived with stronger intensity. However, if an emotion is perceived with low intensity, there is less variation among viewers. For example, DD, S and TGTBTU show distinct profiles across viewers (e.g. joy being felt intensely in DD or fear in S), but the perceived intensity of emotions still varied comparatively much among viewers. Meanwhile, the less intensely perceived emotions in DPS or FG show smaller confidence intervals.

3.4 Covariation of trait and trait anxiety with NFC under naturalistic stimulation

3.4.1 Trait and state anxiety

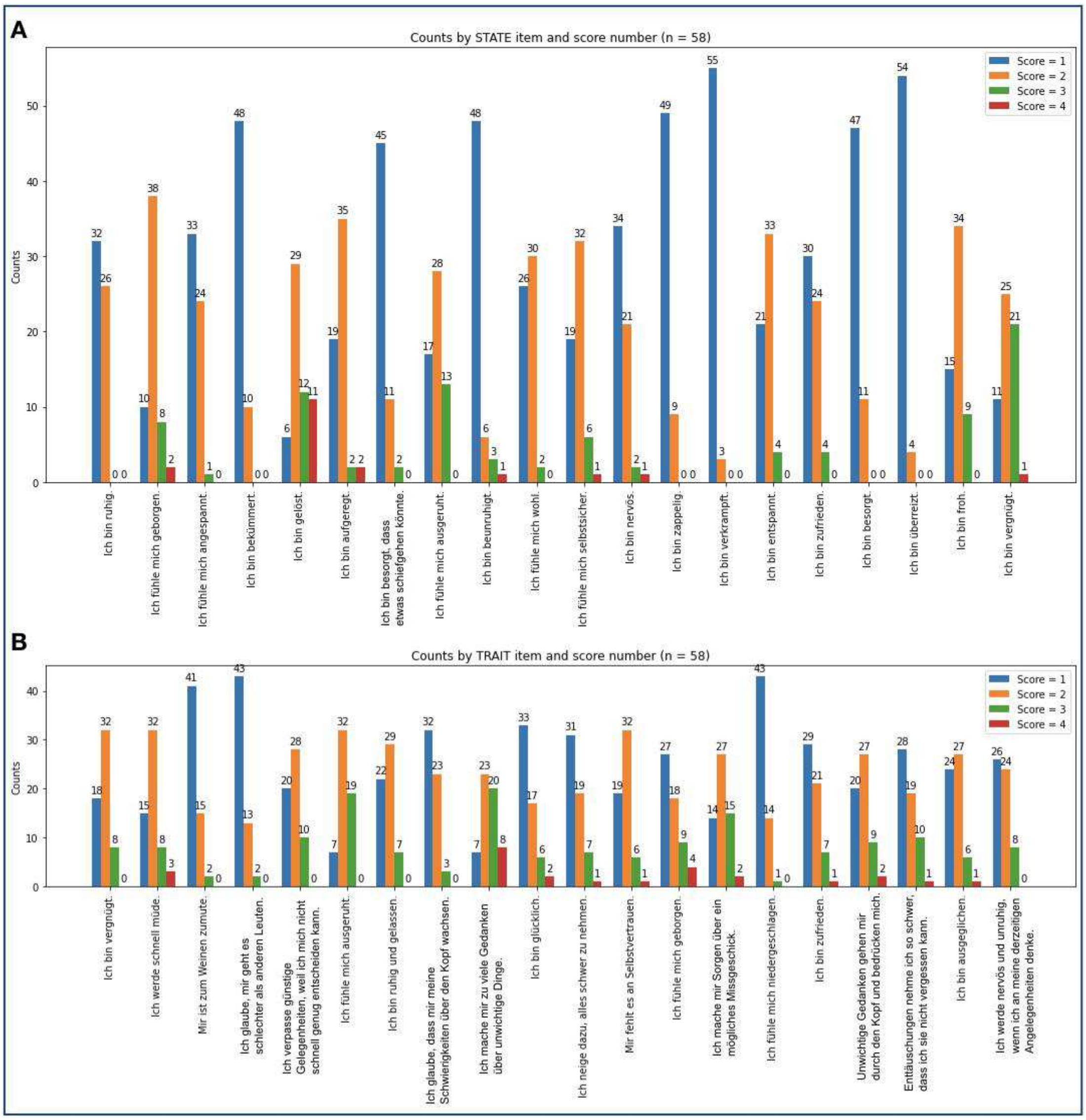


Figure 27. **Results on state and trait STAI.** Frequencies of scores participants rated on each item are displayed in A) of the state STAI and in B) of the trait STAI. Lower scores represent lower anxiety. When necessary, item scores were reversed, so that higher scores represent higher levels of anxiety even if the item is of positive valence (e.g. item 1 of state STAI: “I am calm”.).

Item-level responses of the state and trait versions of the STAI were used in our analysis. Figure 27 A and B show that responses representing lower levels of anxiety were most frequent across all items in our sample in the state and trait STAI. Lower scores here universally represent lower anxiety, which might correspond to higher presence or intensity (e.g. orig. response “almost always”/”very much so”) or absence (e.g. orig. response “almost never”/”not at all”) of the feeling described in the respective item, because some items are reverse coded. For example, lower scores of the state item “I am calm.” represent greater feelings of calmness, while lower scores of the item “I am tense” represent lower feelings of tension.

The pairwise distances between subjects’ scores across all items of the state STAI (Figure 28 A) and trait STAI (Figure 28 B) show higher variability in distances between subjects in the state compared to the trait STAI.

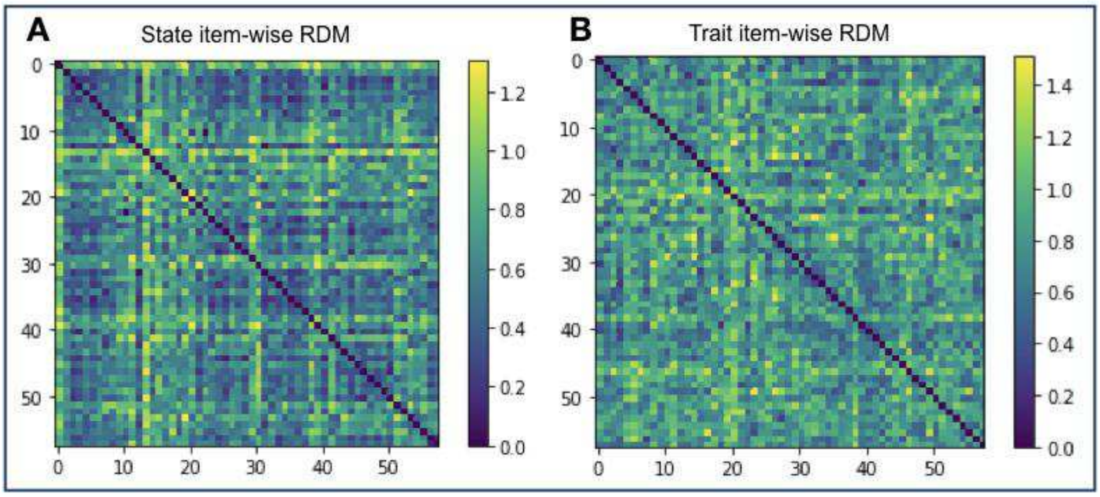


Figure 28. **RDMs between subjects are displayed in A) for state STAI and in B) for trait STAI item responses.** Distances are calculated as 1 - Spearman correlation values. Each row and each column represents one subject (N = 58). The lower the values, the smaller is the distance between subjects and the more similar two subjects were in their item-level responses to the respective STAI. Higher values indicate greater distances and therefore more dissimilarity between subject pairs.

3.4.2 Representational similarity analysis

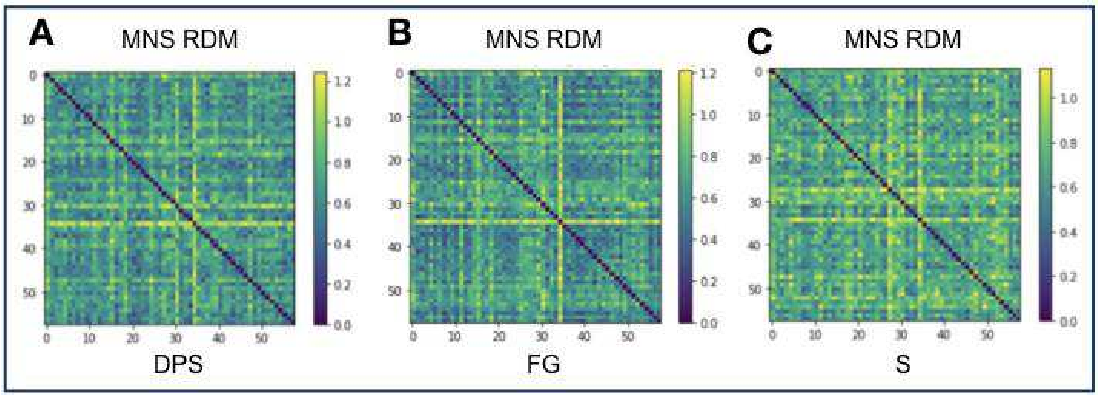


Figure 29. **RDMs between subjects in the MNS network are displayed based on A) DPS, B) FG, and C) S.** These RDMs showed significant correlations with the RDM based on the state STAI. Distances are calculated as 1 - Pearson correlation values. Each row and each column represents one subject (N = 58). Lower values indicate smaller distances and more similarity between pairs of subjects, while higher values indicate greater distances and more dissimilarity between subject pairs.

We compared the results of the state and trait RSA against their respective null distributions with a significance threshold of $p < .05$.

The RSAs based on the trait STAI were not significant for any of the networks in any of the RS or movie conditions.

The state STAI resulted in three significant positive correlations between the behavioural RDM and and a NFC RDM. In the MNS network, the NFC RDMs based on the Dead Poets Society (Figure 29 left, DPS, correlation coefficient: 0.16, distance coefficient: 0.84, $p = 0.046$), Forrest Gump (Figure 29 middle, FG, correlation coefficient: 0.21, distance coefficient: 0.79, $p = 0.007$) and Scream (Figure 29 right, S, correlation coefficient: 0.18, distance coefficient: 0.82, $p = 0.031$) movies were significantly correlated with the state STAI RDM (Figure 28 A).

4. Discussion

In this thesis, I have argued for the usefulness of NV in individual differences and socio-affective research. Movies are complex, multimodal, dynamic stimuli that allow for studying the continuous processing of input that approximates real-life. Movies might amplify individual differences (Finn et al., 2017) and benefit the detection of brain-behavior relationships (Finn et al., 2020; Chen et al., 2020; van der Meer et al., 2020; Finn & Bandettini, 2020). Socio-affective traits and behaviors might be studied particularly well under naturalistic stimulation (Sonkusare et al., 2019; Saarimäki, 2021).

However, realizing the full potential of NV necessitates careful considerations regarding stimulus choice and study design. Although some effects of NV, for example on the consistency of FC, might generalize across different movie stimuli (Tian et al., 2021), often specific stimuli are needed to detect individual differences or specific brain-behavior relationships (Lahnakoski et al., 2014; Bacha-Trams et al., 2017; Jääskeläinen et al., 2016; Bacha-Trams et al., 2018; Finn et al., 2018; Chen et al., 2020; Kröll et al., 2023). Establishing concrete connections between movie stimuli and their features, brain measures and phenotypes of interest is essential (Finn et al., 2017; Finn et al., 2020; Eickhoff, Milham & Vanderwal, 2020). This need in NV research makes large-scale efforts to create NV datasets including stimulus annotations and subject phenotyping indispensable. The creation of such a dataset lies at the heart of this thesis.

4.1 Inter- and intra-subject variability in NFC across a full narrative movie

The first project investigated the inter- and intra-subject variability of NFC over the course of a full narrative movie. By analyzing a publicly available dataset that contains fMRI data spanning a full narrative movie, we investigated changes in variability over multiple, consecutive movie segments. Inter- and intra-subject variability were best explained when accounting for network, movie segment, valence and a movie segment by valence interaction. Additionally, arousal played a role in explaining intra-subject variability by interacting with movie segment and valence. The effect of the movie stimulus on changes in inter- and intra-subject variability was network dependent. Comparing portrayed valence and arousal across movie segments showed that both varied across the segments, indicating differences in emotional content that we could relate to the content of the different movie segments.

4.1.1 Portrayed valence and arousal

Emotions are an important feature of movie stimuli. Shorter movies have been used to study emotion processing (Westermann et al., 1996; Carvalho et al., 2012; Schaefer at al., 2010) and longer movies might allow studying emotions across a larger timescale. Emotions are a major factor in narration (Cutting, 2016; Aldama, 2015), change over time, and dynamically interact with social context (Redcay & Moraczewski, 2019). Therefore, full narrative movies have clear advantages for studying emotions in a naturalistic setting. Additionally, emotions portrayed in movies affect inter-subject synchronization (Dziura et al., 2021) and inter-subject alignment of brain states (Chang et al., 2021), which makes them a relevant factor for studying individual differences using naturalistic viewing.

Here, we used a previously reported description (Labs et al., 2019) of portrayed valence and arousal for comparisons between the emotional content of different movie segments. Our results showed that movie segments differed in the direction (i.e.: positive/negative valence ; high/low arousal) and the extent of agreement between observers concerning these measures (Figure 6). In our results a pattern emerged in which segments that were marked by high concordance in positive valence also showed good agreement in low arousal evaluations (segments 1, 6, 7, 8), whereas the reversed pattern was observable in segment 4. This indicates a potential negative relationship between valence and arousal as depicted in our movie stimuli. Valence and arousal are the bipolar dimensions in circumplex models of affect (Yik, Russell, Barrett, 1999), and the relationship between valence and arousal seems to be highly individual and related to personality and culture (Kuppens et al., 2016).

Overall, the pattern of results implies that the segments of the chosen movie stimulus differed in emotional content, which makes it valuable for inducing variability in functional networks associated with socio-emotional processing. Specifically, the Forrest Gump movie features a broad range of themes (love, friendship, politics, fate), settings (varying historical events, places, times and roles of the protagonist) and situations portraying a wide spectrum of emotions in different contexts. Our results are thus in line with studies showing that movies can elicit complex and mixed states of emotions (Schaefer et al., 2010; Carvalho et al., 2012). In particular, the Forrest Gump movie stimulus has been shown to induce distinct affective states throughout the movie, which was used to map the topographic organization of these states (Lettieri et al., 2019). Hence in accordance with the proposal by Finn and colleagues (2017), the chosen movie can evoke brain states in a meaningful manner, and thus represents a fitting stimulus for studying variability in and between subjects over time.

We investigated portrayed valence and arousal as important emotional features of the movie stimulus. A movie can be characterized by a number of features and characteristics that might modulate the cognitive and emotional effects of a movie. Studies on a movie stimulus depicting a moral dilemma showed that inter-subject correlation of brain activity in various brain areas was influenced by prior knowledge about relationships between movie characters (Bacha-Trams et al., 2017), thinking styles of participants (Bacha-Trams et al., 2018) and the social perspective taken by participants (Bacha-Trams et al., 2020), which was also shown in a previous study (Lahnakoski et al., 2014). Furthermore, inter-subject correlation in brain activity in the lateral orbitofrontal cortex and ventromedial prefrontal cortex were shown to be influenced by participants watching movies featuring either just or unjust violence (Adebimpe, Bassett, Jamieson & Romer, 2019). These findings highlight the complexity of movies and the importance of careful consideration of underlying features of the stimuli. Therefore, characterizing the features of a movie stimulus is crucial for naturalistic viewing research (Grall & Finn, 2022).

4.1.2 Inter-subject variability

Across networks, inter-subject correlations increased over the course of the movie, indicating a general tendency of subjects' NFC to become more similar (Figure 7A).

By using LMM, we found several factors contributing to changes in inter-subject variability, including network, movie segment, valence and interaction of movie content and valence. Specifically, the model that best explains inter-subject variability comprises the fixed effects network, movie segment, valence and a movie segment by valence interaction, with subject identity as a random effect.

Looking more closely at the networks, we see that some networks, such as the CogAC, MNS, Motor, ToM and eSAD, do not contribute significantly to changes in inter-subject variability. This might indicate that these networks are not sensitive to the effects of a full narrative movie and the emotions portrayed within. For those networks that are significantly modulated by movie content, we observed large variations across networks in inter-subject variability. The AM, ER and SM networks are associated with higher inter-subject variability, which can be seen in lower inter-subject correlation values (Figure 7A) and negative model coefficients (Figure 7B). This indicates that emotion regulation and long term memory processes are most sensitive to a full narrative movie. This might reflect the stimulus containing a highly emotional narrative and many references to real world events and history. Contrarily, the EmoSF, Empathy, Rew, VigAtt, WM and eMDN networks are associated with lower inter-subject variability. Across all networks, the coefficients exhibit a wide range in values, with the ER network showing the highest absolute coefficient, indicating the greatest effect on inter-subject variability. Movie segment had a small negative effect on inter-subject variability, indicating that inter-subject variability decreases over the course of a full narrative movie, which is also reflected in a slight increase in inter-subject correlation values (Figure 7B). Previous research has shown high inter-subject variability in response to professionally produced and conventional movies that were much shorter (< 20min) than a full narrative movie (Hasson, Malach & Heeger, 2009; Vanderwal et al., 2015). It is likely that the change towards more similarity in NFC over the course of the movie results from the shared experience, which is created to evoke certain reactions and feelings in the audience. Indeed, viewers' emotional and cognitive states can be affected and synchronized through director's decisions, such as the camera settings, light, performance of actors, scripts and dialog, and more (Tarvainen et al., 2015; Baranowski & Hecht, 2017). Studying viewers' emotions while watching the identical stimulus used here, Lettieri et al. (2019) showed that ratings of basic emotions were consistent across viewers, indicating an overall highly similar emotional experience induced by the movie. Emphasizing the relevance of affective states in movie fMRI, previous studies showed higher alignment of brain states between subjects during highly affective events in a TV show (Chang et al., 2021) and more synchronization of amygdala activity between subjects during positive events in a "shared watching" condition (Dziura et al., 2021).

In our study, valence was associated with higher inter-subject variability. A study by Nummenmaa et al. (2012) studied the relationship between perceived valence and arousal and inter-subject synchronization of brain activity during movie watching. They found that more negative valence was associated with increased inter-subject synchronization in an emotion-processing network and the default-mode network, while high arousal was associated with increased inter-subject synchronization in somatosensory cortices, and visual and dorsal attention networks (Nummenmaa et al., 2012). This is in line with the pattern of positive valence being associated with higher variability in our results.

However, the movie segment by valence interaction has a positive coefficient, indicating that positive valence is associated with lower inter-subject variability across the course of a full narrative movie. This might represent the effects of a conventional Hollywood movie orchestrating higher similarity in viewers' experience by using positive portrayed emotions.

The random factor subject identity had a very small negative effect on inter-subject variability, indicating that there were no great differences between subjects.

4.1.3 Intra-subject variability

Similar to inter-subject variability, intra-subject variability increases over the course of a full narrative movie across networks (Figure 7A).
When selecting the best model in our LMM analysis to explain intra-subject variability, network, movie segment, arousal, and valence emerged as relevant fixed effects. Additionally, the model included a movie segment by valence and a movie segment by arousal by valence interaction. Again, subject identity was included as a random effect.
Of all networks, only the MNS and Rew networks did not affect intra-subject variability significantly. The AM, ER, SM, ToM and eSAD networks were associated with increased intra-subject variability, while the CogAC, EmoSF, Empathy, Motor, VigAtt, WM and eMDN networks were associated with decreased intra-subject variability. Similar to the results on inter-subject variability, emotion regulation and long-term memory were most sensitive to the effects of a full narrative movie, showing the highest intra-subject variability across movie segments. Additionally, networks processing self- and other-related social cognition showed high intra-subject variability, indicating that a full narrative movie might tax introspection and relating to others in a way that varies along the narrative.
Similar to the pattern of results seen in inter-subject variability, valence was associated with higher intra-subject variability while the movie segment by valence interaction was associated with lower intra-subject variability. Additionally, the three-way interaction between movie segment, valence and arousal was associated with lower intra-subject variability. This might indicate that positive valence is generally associated with higher intra-subject variability, although the progression of movie segments and higher portrayed arousal decrease intra-subject variability. While arousal did not significantly influence inter-subject variability, it interacts with movie segment and valence when influencing intra-subject variability. This might indicate that arousal is a more relevant factor when investigating variability within subjects and might prompt future comparisons on the effects of movies with different levels of arousal on single subjects. Arousal seems to be influenced by various stylistic features of a movie and can be further differentiated into subdimensions such as energetic and tense mood (Tarvainen et al., 2015).
The random factor subject identity had a very small positive effect on intra-subject variability, indicating that there were no great differences between subjects.

4.1.4 Limitations

By using an unusually long naturalistic stimulus - a full-length movie - our study offers important insights into inter- and intra-subject variability in NFC across a two hour acquisition period. However, this comes at the cost of investigating only the effects of a single movie. Comparisons with an equally long resting state acquisition or a movie stimulus without a narrative would have given stronger evidence for the effect of full narrative movies. However, there were no such scans available in this dataset. Analyses of additional full narrative movies might expand the insights gained into the effects of different narratives. The choice of a conventional Hollywood movie might have led to higher inter-subject similarity (Hasson,et al., 2010; Vanderwal et al., 2015; Chang et al., 2021; Tarvainen et al., 2015; Baranowski & Hecht, 2017), while more ambiguous or emotionally and socially equivocal movies could enhance inter-individual differences to a greater degree.

Familiarity with a movie stimulus has been discussed as a potential factor for driving individual differences. An effect of repeated movie watching in functional connectivity on the network level has been shown before (Andric et al., 2016). However, such effects can be assumed to be low in our sample. All participants were familiar with the narrative of the movie, and only one participant reported to never have seen the movie (Hanke et al., 2014).

4.2 Movies Dataset

Various research efforts have started including NV paradigms into their large-scale data acquisitions (for example, HCP, HBN and Cam-CAN) or focussed on NV paradigms to explore this approach in depth, like the *studyforrest* project whose data was used in the first project. When NV is included in large-scale research efforts, it might not be the focus, and therefore implemented as rather short movie clips or clips made up from various movies (see HCP dataset for example). While a lot of phenotypic information about the subjects might be collected in these projects, less information might be available about the stimuli and how they relate to the respective phenotypes. Employing a full narrative movie, on the other hand, as in the *studyforrest* project, and focussing on characterizing that stimulus extensively might reduce resources for subject phenotyping and exploring various different movie stimuli. Every dataset that includes NV allows for learning more about the best use cases, advantages and disadvantages of this approach. However, different datasets and research endeavors focus on different objectives, and no dataset can offer the data needed to answer every research question. Therefore, more datasets are needed to enrich the landscape of NV research.

The Movies dataset aims to offer a dataset suitable for exploring individual differences research particularly in socio-affective phenotypes. To that end we have acquired neuroimaging, including NV, data, extensive phenotyping of the subjects and two annotations of the movie stimuli. Data presented here are a subset of the existing data, as acquisition is still ongoing.

4.2.1 Phenotypic data

Demographically, the Movies dataset up to this point (acquisition is still ongoing) presents a young sample with relatively high education. Distributions of age and education were similar across males and females. Only one participant reported their gender being different from their sex. Although this sample includes few subjects with little to no formal education, it is more varied in educational background than many other psychological studies acquiring data primarily on university campuses.

Comparing Figures 9 - 24 gives insight into the distributions of the phenotypic data that was assessed. These figures present the z-standardized scores of each measure or its respective subscales, allowing for comparison of all measures. Z-standardization transfers all measurements onto a common scale, centering values around a mean of 0 and a standard deviation (SD) of 1. The z-score of each subject now represents the number of standard deviations above or below the mean of the entire sample. Data are presented in boxplots, where the median is surrounded by a box ranging from the lower quartile to the upper quartile and encompassing fifty percent of the data points. In these instances, the whiskers extend to the data points that lie within 1.5 interquartile ranges (distance between lower and upper quartile). Data points outside the whiskers might be outliers.

For all measures, half of the subjects' scores are within 1 SD above or below the mean (as depicted by the box of each boxplot being within the -1 to 1 values on each scale). Skewness and range of the distribution still vary among the measures, indicating that variability in subject responses varies between measures. Some measures might be expected to show less variability in the healthy sample that was acquired, such as the BDI and STAI, but both measures still include datapoints up to or over three standard deviations above the mean.

The focus of the phenotypic assessment was on socio-affective traits, i.e. characteristics, behaviors and skills that are relatively stable over time. To this end, questionnaires were employed that assess personality (NEO-FFI, ANPS), emotional (FEEL-E, SEE, EKF-S, STAXI, TAS-26) and social (ISK-K, ToM) experiencing, processing and skills, affect and mood (PANAS), neurological and clinical traits (ADHS-E, BDI, STAI) and cognitive functioning (verbal fluency). The choice of questionnaires was intended to balance broad and in-depth phenotyping. To characterize the subjects broadly, different traits and skills were assessed. In particular, traits that affect one another or interact with one another were assessed, such as personality, social and emotional traits and related clinical traits

(Soto, et al., 2021; Schoon, 2021; Roberts & Yoon, 2022). For example, subjects scoring higher on the personality trait extraversion might show more social offensiveness. Most measures were self-report questionnaires, which might convey the subjects' self-concepts better than their actual behaviors (Breil et al., 2022). However, many similar behaviors and traits were assessed in multiple measures here, allowing for assessing certain concepts from different angles. This in-depth phenotyping approach increases the robustness of the phenotyping of this sample. Indeed, some questionnaires include highly similar subscales. For example, both SEE and EKFS assess emotion regulation and self control, as well as similar concepts such as recognizing, understanding and accepting one's own emotions. The TAS-26 complements these concepts with a subscale difficulty describing emotions, here approaching similar skills from an angle of deficits, focusing on problems in that area. An advantage of this robust assessment of similar traits through multiple measures is that it allows for analyzing the latent structure underlying these traits, for example through factor analysis. Factor analysis reduces the number of observed variables to fewer underlying factors, aiding in interpretability and the discovery of hidden structures in the data (Spearman, 1904; Treiblmaier & Filzmoser, 2010). For example, it is presumable that the aforementioned subscales of different measures (e.g. the SEE and EKFS subscales on emotion regulation and self control) assess the same underlying construct, although it is measured using different items.

Abstracting from single measures to underlying constructs also aids in the comparability of different datasets and samples. This might alleviate a limitation of some of the measures used in this study, which is their language specificity. Some measures are originally in German, have been validated in German samples only and not yet translated to other languages. This inherently hinders the transferability of this specific study design to other samples, which is important for testing the generalizability of results based on analysis of this dataset. However, the use of different measures across studies is common. Socio-affective traits were also assessed in the HCP dataset, for example, although different questionnaires and tasks were used. Here, a factorization approach might help draw similar conclusions from both datasets: it allows for relating the underlying socio-affective constructs to NV imaging data, instead of specific questionnaire scores or item responses.

Thorough assessment of traits is highly relevant when studying inter-individual differences in functional brain and network architectures. For example, individual differences in subjective appraisal of emotions when watching movies modulate functional brain connectivity (Liu et al., 2023), personality traits such a neuroticism or paranoia can modulate brain activity and connectivity (Cremers et al., 2010, Gentili et al., 2017; Finn et al., 2018), individual differences in trait anxiety correspond to individual differences in FC in an emotion processing network (Markett et al., 2016), and emotion regulation traits are associated with efficiency in fronto-parietal and default-mode networks (Pan et al., 2018). Individual differences in functional brain organization and individual differences in behavior are related (Kong et al., 2018; Gordon et al., 2020; Finn & Bandettini, 2021; Grady et al., 2023; Li et al., 2023) and phenotyping of subjects is needed to advance single subject, “precision medicine” approaches to understanding brain-behavior relationships on an individual level (Michon et al., 2022; Zwir et al., 2023; Kraus et al., 2023). FC profiles are unique enough to allow for the identification of individual subjects (Finn et al., 2015), indicating that stable individual differences between subjects exist (Gratton et al., 2018). However, individual identification based on structural MRI data can outperform identification based on fMRI data, because fMRI data is not only shaped by stable traits, but is also state-dependent (Cole et al., 2014; Geerligs et al., 2015; Salehi et al., 2020; Nakuci et al., 2023; Wang et al., 2023; Lee et al., 2023). Moreover, these state-dependent changes include both common aspects that are generalizable across subjects, and idiosyncratic aspects that are specific to individuals (Porter et al., 2023), again highlighting the need to characterize subjects on multiple levels.

In addition to the phenotyping of traits, the Movies dataset includes state questionnaires assessing the current affective state and mood of subjects directly before MRI measurement. fMRI and FC specifically can be influenced by vigilance (Liu & Falahour, 2020), attention (Song & Rosenberg, 2021), mood (Mirchi et al., 2019), mind wandering (Kucyi, 2018) and emotional states (Lettieri et al., 2019; Lettieri et al., 2022; Dan et al., 2023; for review see Martin et al., 2021). Given that FC acquired during RS can be affected by prior tasks and emotional events (Gordon et al., 2014; Gaviria et al., 2020), arguably the mental and affective state of participants immediately preceding study participation might likewise influence MRI measurements. NV is well suited to induce emotions, but emotions depend on context, previous experiences (including prior emotions) and personal relevance (Saarimäki, 2021).

To track the affective state of the subjects across study participation, the SAM was administered to assess arousal, valence and dominance directly before MRI measurement and after each movie stimulus. Comparison of the responses will allow for studying the effects each movie stimulus had on the affective state of the participants and the effect of the affective state prior to MRI measurement. Additionally, responses to the state versions of the BDI and STAI can give insights into the relationship between participants’ current mental well-being and their responses to NV. Lastly, the state STAXI was administered to control for current anger. However, state anger is strongly context-dependent and shows little variance in a neutral situation (Rohrmann et al., 2013), and the resulting left-skewed distribution can be found in this dataset, too. The situation of study participation should ideally not induce anger, so that variance between state anger is expected to be low. Nevertheless, a few subjects seem to have been in a state of heightened state anger compared to the rest of the sample (Figure 23), indicating that it is still meaningful to screen for state anger. Future studies using this dataset need to determine if state anger might influence the respective research question and data analysis, and might exclude these subjects from analysis if deemed necessary. The extensive phenotyping of the Movies dataset allows researchers to study the sample in depth, making informed choices about subject inclusion and exerting control about potential confounds.

4.2.1.1 Pre- and post-MRI state questionnaires

Some state questionnaires were applied directly before and after the NV MRI measurement to assess subjects' affective and stress-related states and measure the impact of the MRI measurement on these. T-tests revealed significant differences between pre and post measures of the PANAS subscale positive affect, but not the subscale negative affect. Ratings of positive affect had decreased after the MRI measurement, while ratings of negative affect remained similar. Overall, ratings of positive affect were much higher than ratings of negative affect. This indicates that the MRI measurement decreased the subjects' positive affect slightly, but did not induce negative affect. A decrease in positive mood might plausibly be related to the prolonged MRI acquisition and related physical discomfort, but an effect of the movie stimuli cannot be ruled out. However, it is encouraging that the long acquisition and the movie stimuli with partially negative content did not increase negative affect in subjects.

T-tests of the DSSQ revealed significant differences between pre and post measures of all three subscales engagement, distress and worry. Response values to the engagement and worry subscales decreased after MRI measurement, while response values to the distress subscale increased. The differences were not very large, indicating that even long NV MRI acquisitions are tolerable for subjects. A reduction in engagement might be attributed to the by then long ongoing acquisition appointment, while a reduction in worry might be related to the familiarization to the MRI measurement. Not all participants had undergone MRI scanning prior to study participation, so that the unknown situation might have induced worry in these subjects. However, some of the items in the DSSQ, including items of the distress subscale, are specifically assessing worry and distress about task performance. Because there was no classic task involved in the NV paradigm, subjects might have been unsure how to assess their performance, leading to more distress.

4.2.1.2 Movies-Questionnaire

The self-drafted Movies questionnaire gives insight into the viewers appraisal of the movie stimuli as well as their habits and preferences regarding movie watching in general.

Familiarity with the movie stimuli was mixed across the stimuli. While TGTBTU, LIB and DMW were unknown to most subjects, the other stimuli received more mixed responses. DD and FG were the most well known stimuli in this sample. The movie clips making up the SS were mostly unfamiliar to the subjects.

Subjects liked the movie stimuli to differing extents. Overall, TGTBTU was the least liked movie stimulus. DMW and S received more mixed responses, while LIB, FG, DPS, DD and DD were liked well overall.

A similar pattern is seen in the relatability of the movies. TGTBTU and DMW appear to be less relatable than all other movie stimuli, where responses were more balanced. This indicated that for most movies, relatability seems to be subject-specific, as responses were similar and balances across response categories for most movies. At this level of analysis, however, it is not clear which underlying responses of individual subjects make up this pattern. Potentially, some subjects could not relate to the movies at all while some subjects related a lot to most of them, or subgroups of subjects could relate well to some movies, but not others, with other subgroups responding in the opposite direction.

Acoustic understanding, that is, how well subjects could hear the stimuli, was assessed to control for how well the audio features of the stimuli (speech, music, sounds) were perceived by the subjects. Overall, the majority of subjects heard the movies at least somewhat well, with similar responses across most movies.

Drowsing or sleeping was assessed for all MRI measurements. The majority of subjects reported never drowsing or falling asleep. Drowsing or sleeping was reported most often during the anatomical and second RS scan. Presumably, subjects had familiarized themselves with the MRI environment at that point and were most prone to sleepiness in the absence of any stimulation or task. Drowsing and sleeping drastically reduced during the NV measurements, highlighting that even during longer MRI measurements, NV successfully reduces sleepiness and increases participant engagement. The highest number of subjects reported drowsing or sleeping during TGTBTU.

Fittingly, TGTBTU was the movie stimulus that was rated the most boring and, as described above, was overall liked the least. Movies were rated across different affective adjectives, where subjects could rate which movie is best described by the respective adjective. DMW was rated with the most negative adjectives (most sad, negative, disgusting and angering), S was simultaneously the most exciting and scariest and DD was the happiest and most positive. Other movies were rated with distinct adjectives, showing that overall, each stimulus contributed different affective aspects most intensely.

The chosen movies represented different affective categories. However, the response format does not allow for inferences about the absolute intensity each affective adjective was perceived as, only the relative intensity compared to all other movies. That is, a movie might be rated as scariest movie only when compared to even less scary other stimuli.

The questionnaire further assessed which genre of movies participants prefer to watch, what aspects of the movies they are paying attention to the most, how many hours per week they watch movies or series and which language they watch movies or series in. These responses, as well as the responses about personal appraisal of the stimuli, will be useful for contextualizing subjects' responses to the movie stimuli. Conventional movies are designed in a way that makes them easy to understand for the viewer, that enables rapid processing and that steers the viewer's attention and emotions (Cutting, 2016). However, it is not yet clear how this processing is modulated by personal preferences, experiences and habits surrounding movies and media consumption. The responses to the Movies-questionnaire offer the opportunity to relate differences in neuroimaging data, phenotypic data and the stimuli and media related information on a subject level.

4.2.2. Future data sharing of Movies dataset

The cost associated with acquiring big neuroimaging samples and investing in additional characterizations of subjects and stimuli makes the resulting data very valuable. Additionally, one dataset can never include all stimuli and phenotypic assessments of interest, so that research efforts across multiple datasets are needed to reach robust conclusions about the potential and limitations of NV. These factors result in a great need for data sharing and collaboration (Liu et al., 2014; Moon & Lee, 2017; Vanderwal, Eilbott & Castellanos, 2019; Eickhoff, Milham & Vanderwal, 2020). However, the sharing of stimuli based on conventional movies might be restricted by the copyright of the movies. Even if precise information about stimulus creation is included in a study, Gross & Levenson inadvertently offer an illuminating example of the problems that may still arise: "We are committed to sharing the fruits of our labours with others working in this area. Given that most of the film clips in our set were extracted from commercial films, we cannot provide copies of the films themselves. All of the commercial films in the set, however, are currently available on videotape, and, upon request, we will provide the editing instructions needed to produce the same film clips that we used." (Gross & Levenson, 1995). While the authors were foresighted in their efforts to share their work, the decline of videotapes as a common media format limits the practicability of their approach over time. Similar developments might apply to the currently used DVD media format, and NV researchers will have to work on solutions for this problem. Concerns regarding copyright might be circumvented by the creation of purpose-built movie stimuli for research (see Inscapes as an example, Vanderwal et al., 2015), but the cost of creating long, complex socio-affective movies similar to conventional movies is a serious constraint.

Sharing the data acquired under NV and additions such as annotations comes with fewer limitations and has gained popularity (Van Essen et al., 2013; Hanke et al., 2014; Shafto et al., 2014; Hanke et al., 2016; Alexander et al., 2017), and the Movies dataset has acquired informed consent of subjects to share their data upon request, so that the dataset and annotations are planned to be made available to other researchers in the future.

4.3 Annotations of emotions induced by the movie stimuli

In an additional study, emotions induced by the movie stimuli were measured. Subjects rated the six basic emotions (happiness, fear, surprise, sadness, disgust and anger, Ekman & Friesen, 1971; Ekman et al., 1983) continuously while watching the movie stimuli. This study consists of a sample that is largely independent of the main Movies dataset sample, although some subjects also participated in the main Movies study before or after participating in the emotion annotation study.

Comparison of the induced emotions measured in this annotation study and responses regarding emotional appraisal of the movie stimuli in the Movies dataset show remarkable similarities. For example, TGTBTU was rated as the most boring movie stimulus by the Movies sample, and likewise shows the lowest and most uniform emotions ratings in the annotation study (Figure 26). Across other movie stimuli, specific events in the movies seem to trigger specific emotion responses. However, few of such events seem to exist in TGTBTU, and none induce emotion responses comparable to the other stimuli. Similarly, DD was rated as the happiest and most positive stimulus by the Movies sample, in line with the continuously high ratings of happiness in the annotation study. S was rated the most scary and exciting in the Movies sample and displays high ratings of fear and anger in the emotion annotation. DMW was rated the most sad, negative, disgusting and angering, and likewise induces high feelings of sadness with distinct and co-occurring bouts of fear, anger and disgust in the annotation study. These examples highlight that emotions can be induced reliably across participants with movie stimuli that are longer and more complex than those traditionally used in emotion induction studies (Gross & Levenson, 1995; Schaefer et al., 2010; Jenkins & Andrews, 2012; Soleymani et al., 2009;

Soleymani et al., 2012; Carvalho et al., 2012; Gilman et al., 2017; Samson et al., 2015). In line with research documenting and analyzing the structure of and directing techniques applied in conventional movies (Bordwell & Thompson, 2010; Cutting, 2016; Baranowski & Hecht, 2017), the stimuli used in this dataset succeed in inducing specific emotion profiles across subjects.

Nevertheless, there is no complete agreement to the emotional appraisal of the stimuli in the Movies sample, and Figure 26 displays variance among the subject emotion annotation ratings. Here, the emotion profiles induced by the stimuli over time - that is, the specific composition of emotion elements - seem to be highly similar across subjects, but the subjective intensity of each emotion still varies considerably among subjects. Further analysis is needed to differentiate between factors residing within the movie stimuli and factors residing within the subjects themselves affecting the intensity and composition of emotions induced by the movies. Given the evident similarity between emotions ratings of the Movies and annotation samples, the in-depth phenotyping and neuroimaging data of the Movies sample will be useful in relating emotional states induced by NV to subject characteristics. In a study likewise assessing continuous emotion ratings of the NV stimulus in a sample separate from the neuroimaging sample, Lettieri et al., 2019, previously established the concept of “emotionotopy”, which describes the topographic organization of affective states in the right temporo-parietal cortex. In a later study, Lettieri et al., 2022, used the same data to investigate the relationship between changes in default mode network connectivity and temporal dynamics of the emotion ratings. These studies highlight two factors that are relevant for studying the emotional effects of NV: emotions are frequently co-occurring and vary across time.

Studying emotional profiles as compositions of co-occurring emotions rather than studying each emotion separately is one approach to account for the interdependence and simultaneity of the emotional experience. Some emotions co-occur more often with specific other emotions, such as happiness and amusement or anger and disgust (Gilman et al., 2017). Anger might co-occur in particular with moral disgust, but not so much physical disgust (Gross & Levenson, 1995; Salerno & Peter-Hagene, 2013; Whitton et al., 2014). This specific co-occurrence becomes evident in the near parallel progression of ratings of anger and disgust in response to DMW in the Movies emotion annotation study (Figure 26), although moral and physical disgust were not differentiated here. Additionally, the emotion ratings show that the intensity of induced emotions varies across the movie stimuli and seem to be tied to specific timepoints or events in the movies. Some of these events seem to increase the intensity of specific emotions temporarily, while other events seem to induce an increase in intensity of multiple emotions.

The emergence and change of emotions across time can be studied well using NV, which offers complex and continuous, by variable emotional stimulation (Saarimäki, 2021). The ratings of the emotion annotation show distinct progressions of induced emotions across time, unique to each movie. While some movie stimuli, such as DD and TGTBTU, display relatively stable emotion profiles across time, other movies are marked by various changes in the emotions they induce. Some movies induce steady changes of emotions over longer time periods (e.g. the increase of fear ratings in S), or the rise and fall of emotions in a certain part of the movie (e.g. the inverted U curve of joy ratings at the end of DPS), or short, defined bursts of emotional responses (e.g. the spike of anger and sadness in LIB). To better analyze the temporal dynamics induced by NV stimuli, some studies have adapted time-varying or dynamic analysis approaches (Hutchison et al., 2013; Calhoun et al., 2014; Raz et al., 2016; Preti, Bolton & Van De Ville, 2017; Gilson et al., 2018; Demirtaş et al., 2019; Betzel et al., 2020; van der Meer et al., 2020; Lettieri et al., 2022; Sun et al., 2022; Levakov et al., 2023). The emotion annotation can enrich the analysis of NV data on a finer temporal scale by offering insights into the emotional effects of the movie stimuli over time.

The Movies emotion annotation study used the same paradigm as Lettieri et al., 2019, (the ReMoTa toolbox kindly provided by Luca Cecchetti, Giada Lettieri and Giacomo Handjaras), increasing comparability between datasets and research endeavors in NV research, and allowing for future replication studies on different movie stimuli.

4.4 Annotation of features of the movie stimuli

Movies are complex, multimodal stimuli that can be characterized on multiple levels. Characterizations of movie stimuli are needed to disentangle which movie features are related to which phenotypes of interest and/or which brain measures. Understanding how movies amplify individual differences and brain-behavior relationships will make stimulus choice easier for future studies, allowing a more precise and deliberate employment of NV.

Incidentally, movies do not exist in a vacuum. Conventional movies are consistent in their structure, how features of the movies are used to convey a story to the viewers, and show reliable links in their narration and narrative (Cutting, 2016). However, movies are also a product of their time. Technical possibilities and directing techniques have evolved over time (Cutting, 2016; Baranowski & Hecht, 2017), as well as content, depictions of characters, speech and more aspects situating a movie in a specific time and place. Therefore, the appraisal of movie stimuli by subjects might change over time, over samples and might even be impacted by current events. For example, Aaron et al., 2018, used a movie stimulus first established by Gross & Levenson, 1995, depicting a stand up comedy sketch by Robin Williams that was originally used to induce amusement. Data acquisition took place shortly after the actor's death, and 28% of subjects reported sadness as a secondary emotion occurring during the comedy bit (Aaron et al., 2018). The primary induced emotion was still amusement and the stimulus' capacity to induce this emotion had not changed, but current events influenced the subjects' associations with and emotional responses to the stimulus. Moreover, especially humor might be prone to change across time, cultures, sex and gender, and along with societal norms, and might therefore induce different reactions across studies (Kotthoff, 2006; Mendiburo-Seguel et al., 2015; Kuipers & van der Ent, 2016; Jiang et al., 2019; Schermer et al., 2023). Biases, stereotypes and discrimination are reflected in popular media, as seen in the representation of specific social groups based on, for example, age, sex, gender, sexual orientation, religion or ethnicity in movies and other media, and the effects of their portrayals on viewers (Harwood & Anderson, 2002; Shaheen, 2003; Lauzen & Dozier, 2005; Mastro & Behm-Morawitz, 2005; Ramasubramanian, 2011; Bond, 2014; Mastro, 2015; Leavitt et al., 2015; Appel & Weber, 2017; Ramasubramanian et al., 2017; Neville & Anastasio, 2019; Appel & Gnambs, 2022; Scharrer et al., 2022). It might be argued that these reflections of societal issues are part of what makes socio-affective movies "naturalistic", but the effects on viewers of different social groups might lead to unintentional confounds. Careful consideration of movie stimuli on multiple levels is needed in NV research.

To this end, the Movies dataset was extended by an annotation of features characterizing the movie stimuli by two raters. Both social and non-social features were documented in their presence and number of occurrences across each movie stimulus. Human annotations are always prone to subjectivity, an effect that could be overcome through big samples, for example by crowdsourcing (Knautz & Stock, 2011; Tian et al., 2017). However, the use of conventional movies limits the option to freely present the movie stimuli because of copyright reasons. Some features can be quantified through machine-based annotations (e.g. McNamara et al., 2017), but not all features of interest are currently captured by these approaches. Machine-based approaches might be more futile in the documentation of low-level features (e.g. auditory or visual features such as pitch or brightness), while human annotations might be more valuable when studying high-level features (e.g. the representation of social groups or the character of social interactions). Altogether, the human feature annotations in the Movies dataset present a useful first characterization of the stimuli, although an addition of more raters or machine-based annotations is highly recommended for future extensions of this dataset.

4.5 Covariation of trait & state anxiety with NFC under naturalistic stimulation

In a first application of the Movies dataset, we investigated the covariation of state and trait anxiety with NFC. As discussed above, both traits and states might have significant influence on brain activity and connectivity, but it is not clear yet how this relationship might be mediated by NV. Movies might function as a “stress test” of brain functioning (Eickhoff et al., 2020), creating naturalistic environments that engage processing related to certain traits or states. For example, movies inducing anxiety and related negative emotions might target neural processes similar to those involved in the processing of anxiety in the real world. FC has been shown to be affected by the experienced tension (Sun et al., 2022) and other negative emotions such as sadness, fear and anger (Raz et al., 2016) during movie watching. The inclusion of NFC data based on various movie stimuli and RS data allowed us to study which conditions might elicit covariation between state and/or trait anxiety and NFC, and whether this covariation depends on specific movie stimuli.

RSA has been established as a useful approach to study brain-behavior relationships, with some studies indicating that item-level responses might be especially meaningful for uncovering these relationships. (Finn et al., 2020; Chen et al., 2020, van der Meer et al., 2020). Similarly, we have here studied the covariation between item-level responses of the trait and state STAI and NFC in various meta-analytical networks.

The STAI has been established as a reliable, short, self-report measure of both state and trait anxiety. State anxiety here is defined as the current intensity of feelings of tension, apprehension, nervousness and worry, combined with arousal of the autonomic nervous system, while trait anxiety is defined as relatively stable inter-individual differences in the predisposition for anxiety measure by the frequency of past anxious states and probability of future states of anxiety (Spielberger & Reheiser, 2009). Consequently, the state version assesses the intensity of feelings or anxiety, while the trait version assesses the frequency of feelings and symptoms of anxiety.

Higher scores in the trait STAI indicate a more frequent experience of state anxiety in situations perceived as dangerous or threatening (Spielberger & Reheiser, 2009). There has been discussion about the validity of the trait STAI for measuring anxiety as opposed to depression (Bieling, Antony & Swinson, 1998; Caci et al., 2003; Bados et al., 2010; Knowles & Olatunji, 2020), indicating that it is not a pure measure of trait anxiety. However, trait anxiety has been shown to have a moderate positive correlation with state anxiety (Leal et al., 2017), suggesting that both measures can still give insight into state-trait relationships of anxiety. In this study, trait anxiety scores were skewed toward lower predispositions for anxiety, but displayed higher variability than state scores (Figure 27). As a reminder, both state and trait STAI include reverse-coded items that have been accounted for, so that lower scores represent lower anxiety instead of level of agreement to the item. An additional limitation is that subjects participating in fMRI studies might have lower trait anxiety than subjects participating in behavior-only research, potentially constricting generalizability of results across different samples (Charpentier et al., 2021).

Higher scores of the state STAI are related to situational factors such as current or expected danger or thoughts of past events relating to the current situation (Spielberger & Reheiser, 2009). In the current analysis, state anxiety scores were overwhelmingly low, predominantly scoring in the lower 2 of 4 possible item responses. The items that reflected higher state anxiety in more subjects' responses were items relating to feeling secure (item 2), at ease (item 5), rested (item 8), joyful (item 19) and pleasant (item 20). State anxiety during study participation could have been related to concerns about the MRI procedure, study participation and performance in general or personal worries and concerns unrelated to the study. Even in the absence of these concerns, long and taxing study participation might not be related to particularly positive feelings such as pleasure and joy. Nevertheless, study participation here can be assumed to be a rather neutral situation (compared to intentionally stressful study environments, as, for example, the Trier Social Stress Test, Kirschbaum et al., 1993), which is overall associated with skewed distributions of the state STAI (Spielberger & Reheiser, 2009).

Similarities between subjects in item-level responses to the trait and state STAI were calculated as RDMs (Figure 28). The RDM based on state anxiety shows more defined clusters than the RDM based on trait anxiety, indicating that some subjects display particularly great distances, i.e. are particularly dissimilar to most other subjects in their state anxiety.

Similarity between subjects in state anxiety was related to similarity in NFC in the MNS network based on NV data from three different movies. RDMs of NFC measures in the MNS based on RS data were not significantly correlated with RDMs of behavioral measures, indicating that the brain-behavior relationship between state anxiety and NFC in the MNS is not intrinsic, but rather elicited by the movie stimuli. In a previous RSA study, Arbula, Pisanu & Rumiati, 2021, found that social and non-social video stimuli led to similar activation patterns in dorsomedial prefrontal cortex in subjects with low agreeableness, but more dissimilar patterns in subjects with high agreeableness. These results indicate that uncovering brain-behavior relationships depends on specific combinations of brain measures and the conditions they are acquired under, and behaviors of interest. Finding the right combination of stimuli and brain measure is a crucial aspect of individual differences research using NV (Finn et al., 2017; Eickhoff, Milham & Vanderwal, 2020; Kröll et al., 2023).

Here, three out of eight movie stimuli could elicit patterns of NFC in the MNS network that covaried with patterns of state anxiety. S is a horror movie that carries great face validity for revealing brain-behavior relationships surrounding anxiety. The emotion annotation of this stimulus shows the highest ratings of fear across all stimuli and compared to all other emotions induced by S. In the Movies sample, S was rated the scariest and most exciting movie stimulus. The movie stimulus presents the beginning of the movie during which a young woman receives increasingly threatening phone calls by a stranger who murders her boyfriend just at the end of the movie excerpt. The scenes are marked by high tension and rising danger. In comparison, DPS and FG seem less anxiety-inducing. DPS might induce tension and anxiety by following the unjust persecution of a teacher and the involuntary involvement of his students, but the movie (and stimulus) ends in a bittersweet display of solidarity. FG follows the beginning of the movie, introducing the protagonist and parts of his biography. The movie stimulus revisits various memories, including some showing problems or conflict in the protagonist's childhood that may be more alarming to the viewer than to the seemingly unsuspecting protagonist (e.g. a family history involving the Ku Klux Klan, physical impairments, his mother giving in to sexual advances of his headmaster to secure his education). However, in the emotion annotation study, joy was rated the most intensely induced emotion throughout the majority of the movie stimulus. A closer investigation into the movie stimuli is needed to understand why these movies elicit covariation between NFC in the MNS network and state anxiety.

The mirror neuron system underlies action observation and imitation (Gallese et al., 1996). Because studying mirror neurons directly requires single-cell recordings, which are rare in humans, establishing and investigating the action observation network through neuroimaging studies became a useful proxy (Caspers et al., 2010). The mirror neuron system creates a link between observing actions and understanding their underlying intentions, which supports theory of mind processes, empathy and complex social interactions (Rizzolatti, 2005; Rizzolatti & Fabbri-Destro, 2008; Fabbri-Destro & Rizzolatti, 2008; Iacoboni, 2008; Rizzolatti & Sinigaglia, 2016). The MNS network standing out in this analysis is delineated from quantitative meta-analysis, converging the results of many neuroimaging studies and representing the peak activations revealed in a conjunction analysis on action observation and action imitation (Caspers et al., 2010). Empathy can be related to anxiety through worry and rumination, and a sensitivity to socio-affective information (Knight et al., 2019). The movie stimuli all represent social stimuli presenting interpersonal interactions and conflict and state anxiety might modulate the responses of the MNS to these stimuli.

This analysis shows that the Movies dataset can be used to uncover brain-behavior relationships and study how inter-individual differences in socio-affective phenotypes are related to inter-individual differences in the functional architecture of the brain. The stimuli are distinguishable in their content, features and emotions induced, and might be well suited to uncover relationships between different phenotypes and brain measures. This study highlights the importance of the specific combination of stimulus, phenotype and brain measure.

4.6 Conclusions

In this thesis, I have discussed benefits and limitations of NV in individual differences and socio-affective research. I have presented two studies on NV and a comprehensive dataset including extensive subject phenotyping and two additional annotations of the movie stimuli.

In the first project, I have studied the influence of movie watching on inter-individual differences in NFC in 14 functional networks. Specifically, a publicly available dataset was used to investigate the effects of a full narrative movie, i.e. a movie stimulus that spans the full narrative of a conventional movie. This stimulus reflects many of the benefits associated with NV. It offers long, dynamic, continuous and complex stimulation encompassing various social relationships and their dynamics over time, semantic knowledge about the world and historical events embedded into the narrative, diverse emotionally evocative events, and the employment of different filmmaking techniques (e.g. flashbacks, voice over narration). Results show that inter- and intra-subject variability in NFC were sensitive to the progressing narrative and emotions portrayed in the movie. The emotion regulation network displayed the highest variability within and between subjects in NFC, followed by networks associated with long-term memory processing. The sensitivity of these networks to the full narrative movie might be explained by the highly emotional narrative and continuous references to real world historical events, highlighting the importance of specific features and content of the chosen movie stimulus. The overarching narrative gives a unique possibility to study emotions in a social context and how they develop over time. These socio-cognitive aspects seem to specifically influence variability within subjects, as high intra-subject variability was additionally seen in networks involved in self- and other-related cognition. Altogether, these results show that a network perspective might help to elucidate the effects of different movie stimuli on specific cognitive domains. However, the limitation of using a full-narrative movie is the lack of comparison with different movie stimuli or RS acquisition. Additionally, no subject characterizations were available that would allow the investigation of relationships between individual differences in socio-affective phenotypes and NFC in response to this unique socio-affective movie stimulus.

Stimulus choice is an integral part of NV research and depends on the focus of each research endeavor. No one stimulus can answer all research questions. Efforts to understand the relationship between stimulus features and effects on subjects will greatly support future informed choices of stimuli (Finn et al., 2020). In-depth characterization of stimuli and subjects are needed to create effective NV paradigms. To this end, the Movies dataset was created, encompassing RS and anatomical (f)MRI data, fMRI data based on eight different movie stimuli intended to induce socio-affective processing, extensive phenotyping of socio-affective traits and states of the subjects, an annotation of emotions induced by the movie stimuli in a separate sample, and an annotation of features of the movie stimuli by two raters.

The phenotypic assessments capture individual variability in the Movies sample, a requirement for studying brain-behavior relationships and in individual differences research (Finn et al., 2017). The robust assessment of socio-affective traits lends itself to the use of factorization approaches to uncover general structures underlying the specific measures and increase comparability with other datasets assessing the same constructs using different measures. The additional state measures allow for differentiating between trait and state effects influencing brain-behavior relationships induced by NV. State measures administered pre and post MRI measurement help gauge the effects of NV on subjects' mood and stress levels, with results indicating that even long NV MRI acquisitions are tolerable for subjects. Additional information about the subjects' appraisal of the movie stimuli and their general preferences and habits regarding movie watching can help identify further factors that might drive individual differences in the reception of NV.

The annotation of emotions induced by the movie stimuli in a separate sample gives valuable information about the effects of the stimuli used in this dataset. The resulting emotion profiles resemble the appraisal of the stimuli given by the Movies sample, indicating reliability of induced emotions across different samples, as is expected from conventional movies (Cutting, 2016). Nevertheless, variability in the induced emotions support the existence of individual differences in the processing and appraisal of movie stimuli, which might be related to individual differences in brain functional architecture. The temporal progression of emotion profiles across the stimuli in time highlight the opportunities lying in time-varying and dynamic FC analysis approaches to uncover the brain's processing of complex socio-affective information on a finer temporal scale and across meaningful changes in the stimulus material (progression of the narrative, changing context of social relationships and emotional events). The emotional profiles reveal the co-occurence and dynamics of emotions induced by NV, emphasizing the complexity of naturalistic stimuli.

The annotation of features of the movie stimuli is needed to better study how movie stimuli induce emotions, elicit individual differences and help uncover brain-behavior relationships. Feature annotations support the disentanglement of the complexity of NV stimuli and might reveal valuable information about how NV works that can guide the stimulus choice of future NV studies. Feature annotations additionally allow for studying the implicit biases, stereotypes and potential discrimination reflected in movies. Human-based annotations should be complemented by machine-based annotations to cover a wide array of low-level (e.g. auditory and visual features such as pitch and brightness) and high-level (e.g. representation of social groups or character of social interactions) features.

A first application of the Movies dataset investigated the covariation in trait and state anxiety with NFC in 14 functional networks. This project aimed to differentiate the potential relationships between NFC and either trait- or state-level behavior, investigating if and how NV might function as a “stress test” of brain functioning (Eickhoff et al., 2020). Comparison between eight different movie stimuli and RS fMRI could reveal that significant correlations between state anxiety and NFC in the MNS network were not intrinsic, but elicited by three specific movie stimuli. These results show that individual variability in behavior covaries with individual variability in brain measures under NV and highlight that specific combinations of stimulus, phenotypes of interest and brain measure are required to uncover brain-behavior relationships. The Movies dataset offers rich data for studying individual differences specifically in socio-affective applications.

NV is a promising approach to study functional brain architecture in a more ecologically valid environment in particular in individual differences and socio-affective research. Since the advent of NV, many studies evidenced its usefulness and benefits, and sometimes great expectations are posed for the future of NV and neuroimaging (Finn et al., 2017; Sonkusare et al.,, 2019; Finn, 2021; Saarimäki, 2021; Jääskeläinen et al., 2022). NV is no panacea, but it could prove itself to be an indispensable addition to task-based and RS studies.

References

A. T. BECK, M.D. C. H. WARD, M. D., M. MENDELSON, M.D. J. MOCK, M. D. A., J. ERBAUGH, M. D. P., & 561. (1961). An Inventory for Measuring Depression. *Library*, 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Aaron, R. V., Snodgress, M. A., Blain, S. D., & Park, S. (2018). Affect labeling and other aspects of emotional experiences in relation to alexithymia following standardized emotion inductions. *Psychiatry Research*, *262*(July 2017), 115–123. https://doi.org/10.1016/j.psychres.2018.02.014

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., … Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*(FEB), 1–10. https://doi.org/10.3389/fninf.2014.00014

Adebimpe, A., Bassett, D. S., Jamieson, P. E., & Romer, D. (2019). Intersubject Synchronization of Late Adolescent Brain Responses to Violent Movies: A Virtue-Ethics Approach. *Frontiers in Behavioral Neuroscience*, *13*(November), 1–13. https://doi.org/10.3389/fnbeh.2019.00260

Aldama, F. L. (2015). The Science of Storytelling: Perspectives from Cognitive Science, Neuroscience, and the Humanities. *Projections*, *9*(1). https://doi.org/10.3167/proj.2015.090106

Alexander, L. M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., … Milham, M. P. (2017). Data Descriptor: An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, *4*, 1–26. https://doi.org/10.1038/sdata.2017.181

Amft, M., Bzdok, D., Laird, A. R., Fox, P. T., Schilbach, L., & Eickhoff, S. B. (2015). Definition and characterization of an extended social-affective default network. *Brain Structure and Function*, *220*(2), 1031–1049. https://doi.org/10.1007/s00429-013-0698-0

Andric, M., Goldin-Meadow, S., Small, S. L., & Hasson, U. (2016). Repeated movie viewings produce similar local activity patterns but different network configurations. *NeuroImage*, *142*, 613–627. https://doi.org/10.1016/j.neuroimage.2016.07.061

Appel, M., & Gnambs, T. (2023). Women in Fiction: Bechdel-Wallace Test Results for the Highest-Grossing Movies of the Last Four Decades. *Psychology of Popular Media*, *12*(4), 499–504. https://doi.org/https://doi.org/10.1037/ppm0000436

Appel, M., & Weber, S. (2021). Do Mass Mediated Stereotypes Harm Members of Negatively Stereotyped Groups? A Meta-Analytical Review on Media-Generated Stereotype Threat and Stereotype Lift. *Communication Research*, *48*(2), 151–179. https://doi.org/10.1177/0093650217715543

Arbula, S., Pisanu, E., & Rumiati, R. I. (2021). Representation of social content in dorsomedial prefrontal cortex underlies individual differences in agreeableness trait. *NeuroImage*, *235*(April), 118049. https://doi.org/10.1016/j.neuroimage.2021.118049

Arifin, S., & Cheung, P. Y. K. (2008). Affective level video segmentation by utilizing the pleasure-arousal- dominance information. *IEEE Transactions on Multimedia*, *10*(7), 1325–1341. https://doi.org/10.1109/TMM.2008.2004911

Aubert, O., Prié, Y., Aubert, O., & Prié, Y. (2017). Advene : active reading through hypervideo To cite this version : Advene : active reading through hypervideo, (Hypertext 2005), 235–244.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Bach, M., Bach, D., de Zwaan, M., & Serim, M. (1996). Validierung der deutschen Version der 20-Item Toronto-Alexithymie-Skala bei Normalpersonen und psychiatrischen Patienten [Validation of the German version of the 20-item Toronto Alexithymia Scale in normal adults and psychiatric inpatients]. *PPmP: Psychotherapie Psychosomatik Medizinische Psychologie*, 46(1), 23–28.

Bacha-Trams, M., Alexandrov, Y. I., Broman, E., Glerean, E., Kauppila, M., Kauttonen, J., … Jääskeläinen, I. P. (2018). A drama movie activates brains of holistic and analytical thinkers differentially. *Social Cognitive and Affective Neuroscience*, *13*(12), 1293–1304. https://doi.org/10.1093/scan/nsy099

Bacha-Trams, M., Glerean, E., Dunbar, R., Lahnakoski, J. M., Ryyppö, E., Sams, M., & Jääskeläinen, I. P. (2017). Differential inter-subject correlation of brain activity when kinship is a variable in moral dilemma. *Scientific Reports*, *7*(1), 1–16. https://doi.org/10.1038/s41598-017-14323-x

Bacha-Trams, M., Ryyppo, E., Glerean, E., Sams, M., & Jaaskelainen, I. P. (2020). Social perspective-takingshapes brain hemodynamic activity and eye movements during movie viewing. *Social Cognitive and Affective Neuroscience*, *15*(2), 175–191. https://doi.org/10.1093/scan/nsaa033

Bados, A., Gómez-Benito, J., & Balaguer, G. (2010). The state-trait anxiety inventory, trait version: Does it really measure anxiety? *Journal of Personality Assessment*, *92*(6), 560–567. https://doi.org/10.1080/00223891.2010.513295

Bagby, M., Parker, J. D. A., & Taylor, G. J. (1994). THE TWENTY-ITEM TORONTO ALEXITHYMIA SCALE-I. ITEM SELECTION AND CROSS-VALIDATION OF THE FACTOR STRUCTURE. *Journal of Psychosomatic Research*, *38*(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Baranowski, A., & Hecht, H. (2017). One Hundred Years of Photoplay: Hugo Münsterberg's Lasting Contribution to Cognitive Movie Psychology. *Projections*, *11*(2), 1–21. https://doi.org/10.3167/proj.2017.110202

Bartels, A., & Zeki, S. (2004). Functional Brain Mapping during Free Viewing of Natural Scenes. *Human Brain Mapping*, *21*(2), 75–85. https://doi.org/10.1002/hbm.10153

Baveye, Y., Dellandréa, E., Chamaret, C., & Chen, L. (2015). LIRIS-ACCEDE: A video database for affective content analysis. *IEEE Transactions on Affective Computing*, *6*(1), 43–55. https://doi.org/10.1109/TAFFC.2015.2396531

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck Depression Inventory–II (BDI-II) [Database record]. APA PsycTests.
https://doi.org/10.1037/t00742-000

Behr, M., Becker, M. (2004). Skalen zum Erleben von Emotionen. 1. Auflage. Göttingen: Hogrefe.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *Journal of Vision*, *9*(5), 1–15. https://doi.org/10.1167/9.5.19

Betti, V., DellaPenna, S., de Pasquale, F., Mantini, D., Marzetti, L., Romani, G. L., & Corbetta, M. (2013). Natural scenes viewing alters the dynamics of functional connectivity in the human brain. *Neuron*, *79*(4), 782–797. https://doi.org/10.1016/j.neuron.2013.06.022

Betzel, R. F., Byrge, L., Esfahlani, F. Z., & Kennedy, D. P. (2020). Temporal fluctuations in the brain's modular architecture during movie-watching. *NeuroImage*, *213*(February), 116687. https://doi.org/10.1016/j.neuroimage.2020.116687

Bieling, P. J., Antony, M. M., & Swinson, R. P. (1998). The state-trait anxiety inventory, trait version: Structure and content re-examined. *Behaviour Research and Therapy*, *36*(7–8), 777–788. https://doi.org/10.1002/humu.21034

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Boeving, E. R., Flannery, J. S., Riedel, M. C., Sutherland, M. T., Laird, A. R., Bottenhorn, K. L., & Eickhoff, S. B. (2018). Cooperating yet distinct brain networks engaged during naturalistic paradigms: A meta-analysis of functional MRI results. *Network Neuroscience*, *3*(1), 27–48. https://doi.org/10.1162/netn_a_00050

Bond, B. J. (2014). Sex and Sexuality in Entertainment Media Popular With Lesbian, Gay, and Bisexual Adolescents. *Mass Communication and Society*, *17*(1), 98–120. https://doi.org/10.1080/15205436.2013.816739

Bordwell, David, and Kristin Thompson. 2010. Film Art: An Introduction. 9th ed. New York: McGraw-Hill.

Borkenau, P., Ostendorf, F. (2008). NEO-Fünf-Faktoren-Inventar nach Costa und Mc Crae. 2., neu normierte und vollständig überarbeitete Auflage. Göttingen: Hogrefe.

Breil, S. M., Mielke, I., Ahrens, H., Geldmacher, T., Sensmeier, J., Marschall, B., & Back, M. D. (2022). Predicting Actual Social Skill Expression from Personality and Skill Self-Concepts. *Journal of Intelligence*, *10*(3). https://doi.org/10.3390/jintelligence10030048

Buhle, J. T., Silvers, J. A., Wage, T. D., Lopez, R., Onyemekwu, C., Kober, H., … Ochsner, K. N. (2014). Cognitive reappraisal of emotion: A meta-analysis of human neuroimaging studies. *Cerebral Cortex*, *24*(11), 2981–2990. https://doi.org/10.1093/cercor/bht154

Brüne, A. (2016). TOM: Theory of Mind Test. SCHUHFRIED GmbH.

Byrge, L., Dubois, J., Tyszka, J. M., Adolphs, R., & Kennedy, D. P. (2015). Idiosyncratic Brain Activation Patterns Are Associated with Poor Social Comprehension in Autism. *Journal of Neuroscience*, *35*(14), 5837–5850. https://doi.org/10.1523/JNEUROSCI.5182-14.2015

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function*, *217*(4), 783–796. https://doi.org/10.1007/s00429-012-0380-y

Caci, H., Baylé, F. J., Dossios, C., Robert, P., & Boyer, P. (2003). The Spielberger trait anxiety inventory measures more than anxiety. *European Psychiatry*, *18*(8), 394–400. https://doi.org/10.1016/j.eurpsy.2003.05.003

Calhoun, V. D., Miller, R., Pearlson, G., & Adali, T. (2014). The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery. *Neuron*, *84*(2), 262–274. https://doi.org/10.1016/j.neuron.2014.10.015

Camilleri, J. A., Müller, V. I., Fox, P., Laird, A. R., Hoffstaedter, F., Kalenscher, T., & Eickhoff, S. B. (2018). Definition and characterization of an extended multiple-demand network. *NeuroImage*, *165*(October 2017), 138–147. https://doi.org/10.1016/j.neuroimage.2017.10.020

Cantlon, J. F., & Li, R. (2013). Neural Activity during Natural Viewing of Sesame Street Statistically Predicts Test Scores in Early Childhood. *PLoS Biology*, *11*(1). https://doi.org/10.1371/journal.pbio.1001462

Carvalho, S., Leite, J., Galdo-Álvarez, S., & Gonçalves, Ó. F. (2012). The emotional movie database (EMDB): A self-report and psychophysiological study. *Applied Psychophysiology Biofeedback*, *37*(4), 279–294. https://doi.org/10.1007/s10484-012-9201-6

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2016). ALE meta-analysis of action observation and imitation in the human brain. *Physiology & Behavior*, *176*(1), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

Chang, L. J., Jolly, E., Cheong, J. H., Rapuano, K. M., Greenstein, N., Chen, P. H. A., & Manning, J. R. (2021). Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Science Advances*, *7*(17). https://doi.org/10.1126/sciadv.abf7129

Charpentier, C. J., Faulkner, P., Pool, E. R., Ly, V., Tollenaar, M. S., Kluen, L. M., … O'Doherty, J. P. (2021). How representative are neuroimaging samples? Large-scale evidence for trait anxiety differences between fMRI and behaviour-only research participants. *Social Cognitive and Affective Neuroscience*, *16*(10), 1057–1070. https://doi.org/10.1093/scan/nsab057

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

Cieslik, E. C., Mueller, V. I., Eickhoff, C. R., Langner, R., & Eickhoff, S. B. (2015, January 1). Three key regions for supervisory attentional control: Evidence from neuroimaging meta-analyses. *Neuroscience and Biobehavioral Reviews*. Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2014.11.003

Costa, P. (1992). *Neo PI-R professional manual*. Psychological Assessment Resources, Inc.

Cox, R. W., & Hyde, J. S. (1997). Software Tools for Analysis and Visualization of FMRI Data NMR in Biomedicine, in press. *NMR Biomed*, *10*(4–5), 171–178. Retrieved from https://afni.nimh.nih.gov/pub/dist/doc/papers/afni_paper2.pdf

Cremers, H. R., Demenescu, L. R., Aleman, A., Renken, R., van Tol, M. J., van der Wee, N. J. A., … Roelofs, K. (2010). Neuroticism modulates amygdala-prefrontal connectivity in response to negative emotional facial expressions. *NeuroImage*, *49*(1), 963–970. https://doi.org/10.1016/j.neuroimage.2009.08.023

Critchley, H. D., & Harrison, N. A. (2013). Visceral Influences on Brain and Behavior. *Neuron*, *77*(4), 624–638. https://doi.org/10.1016/j.neuron.2013.02.008

Cutting, J. E. (2014). Event segmentation and seven types of narrative discontinuity in popular movies. *Acta Psychologica*, *149*, 69–77. https://doi.org/10.1016/j.actpsy.2014.03.003

Cutting, J. E. (2016). Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin and Review*, *23*(6), 1713–1743. https://doi.org/10.3758/s13423-016-1051-4

Cutting, J. E., & Armstrong, K. L. (2018). Cryptic Emotions and the Emergence of a Metatheory of Mind in Popular Filmmaking. *Cognitive Science*, *42*(4), 1317–1344. https://doi.org/10.1111/cogs.12586

Cutting, J. E., Brunick, K. L., & Candan, A. (2012). Perceiving event dynamics and parsing Hollywood films. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1476–1490. https://doi.org/10.1037/a0027737

Cutting, J. E., Brunick, K. L., & DeLong, J. (2012). On Shot Lengths and Film Acts: A Revised View. *Projections*, *6*(1). https://doi.org/10.3167/proj.2012.060106

Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C., & Candan, A. (2011). Quicker, faster, darker: Changes in hollywood film over 75 years. *I-Perception*, *2*(6), 569–576. https://doi.org/10.1068/i0441aap

Cutting, J. E., & Candan, A. (2015). Shot durations, shot classes, and the increased pace of popular movies. *Projections (New York)*, *9*(2), 40–62. https://doi.org/10.3167/proj.2015.090204

Cutting, J. E., DeLong, J. E., & Brunick, K. L. (2011). Visual Activity in Hollywood Film: 1935 to 2005 and Beyond. *Psychology of Aesthetics, Creativity, and the Arts*, *5*(2), 115–125. https://doi.org/10.1037/a0020995

Cutting, J. E., Delong, J. E., & Nothelfer, C. E. (2010). Attention and the evolution of hollywood film. *Psychological Science*, *21*(3), 432–439. https://doi.org/10.1177/0956797610361679

Cutting, J., & Iricinschi, C. (2015). Re-presentations of space in hollywood movies: An event-indexing analysis. *Cognitive Science*, *39*(2), 434–456. https://doi.org/10.1111/cogs.12151

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Dan, R., Weinstock, M., & Goelman, G. (2023). Emotional states as distinct configurations of functional brain networks. *Cerebral Cortex*, *33*(9), 5727–5739. https://doi.org/10.1093/cercor/bhac455

Davis, K. L., & Panksepp, J. (2011). The brain's emotional foundations of human personality and the Affective Neuroscience Personality Scales. *Neuroscience and Biobehavioral Reviews*, *35*(9), 1946–1958. https://doi.org/10.1016/j.neubiorev.2011.04.004

Davis, K. L., Panksepp, J., & Normansell, L. (2003). The Affective Neuroscience Personality Scales: Normative Data and Implications. *Neuropsychoanalysis*, *5*(1), 57–69. https://doi.org/10.1080/15294145.2003.10773410

Dayan, E., Barliya, A., De Gelder, B., Hendler, T., Malach, R., & Flash, T. (2018). Motion cues modulate responses to emotion in movies. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-29111-4

Delgado-Herrera, M., Reyes-Aguilar, A., & Giordano, M. (2021). What Deception Tasks Used in the Lab Really Do: Systematic Review and Meta-analysis of Ecological Validity of fMRI Deception Tasks. *Neuroscience*, *468*, 88–109. https://doi.org/10.1016/j.neuroscience.2021.06.005

Demirtaş, M., Ponce-Alvarez, A., Gilson, M., Hagmann, P., Mantini, D., Betti, V., … Deco, G. (2019). Distinct modes of functional connectivity induced by movie-watching. *NeuroImage*, *184*(June 2018), 335–348. https://doi.org/10.1016/j.neuroimage.2018.09.042

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 1–17. https://doi.org/10.1167/10.10.28

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., … Petersen, S. E. (2006). A Core System for the Implementation of Task Sets. *Neuron*, *50*(5), 799–812. https://doi.org/10.1016/j.neuron.2006.04.031

Douglas-Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M., … Karpouzis, K. (2011). The HUMAINE Database: Addressing the Collection and Annotation of Naturalistic and Induced Emotional Data. In R. Cowie, C. Pelachaud, & P. Petta (Eds.), *Emotion-Oriented Systems* (pp. 488–500). Springer Verlag.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., … Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, *36*(5), 623–636. https://doi.org/10.1007/s10803-006-0107-0

Ebert, D., Krause, J., & Roth-Sackenheim, C. (2003). ADHS im Erwachsenenalter--Leitlinien auf der Basis eines Expertkonsensus mit Unterstützung der DGPPN. *Der Nervenarzt*, *74*(10), 939–946.

Eickhoff, S. B., & Grefkes, C. (2011). Approaches for the integrated analysis of structure, function and connectivity of the human brain. *Clinical EEG and Neuroscience*, *42*(2), 107–121. https://doi.org/10.1177/155005941104200211

Eickhoff, S. B., Milham, M., & Vanderwal, T. (2020). Towards clinical applications of movie fMRI. *NeuroImage*, *217*(October 2019), 116860. https://doi.org/10.1016/j.neuroimage.2020.116860

Ekman, P. (2014). Consants Across Cultures In the Face and Emotion. *Psycnet.Apa.Org*, *18*(2), 114311. Retrieved from http://psycnet.apa.org/record/1971-07999-001

Ekman, P., Levenson, R. W., & Friesen, W. V. (2013). Autonomic Nervous System Activity Distinguishes among Emotions Author (s): Paul Ekman , Robert W . Levenson and Wallace V . Friesen. *Science*, *221*(4616), 1208–1210.

Esteban, Oscar, Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, et al. 2018. “FMRIPrep.” Software. Zenodo. https://doi.org/10.5281/zenodo.852659.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., … Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Evans, A. C., Janke, A. L., Collins, D. L., & Baillet, S. (2012). Brain templates and atlases. *NeuroImage*, *62*(2), 911–922. https://doi.org/10.1016/j.neuroimage.2012.01.024

Fabbri-Destro, M., & Rizzolatti, G. (2008). Mirror neurons and mirror systems in monkeys and humans. *Physiology*, *23*(3), 171–179. https://doi.org/10.1152/physiol.00004.2008

Fernández-Aguilar, L., Ricarte, J., Ros, L., & Latorre, J. M. (2018). Emotional differences in young and older adults: Films as mood induction procedure. *Frontiers in Psychology*, *9*(July). https://doi.org/10.3389/fpsyg.2018.01110

Finn, E. S. (2021). Is it time to put rest to rest? *Trends in Cognitive Sciences*, *25*(12), 1021–1032. https://doi.org/10.1016/j.tics.2021.09.005

Finn, E. S., & Bandettini, P. A. (2021). Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *NeuroImage*, *235*(March), 117963. https://doi.org/10.1016/j.neuroimage.2021.117963

Finn, E. S., Corlett, P. R., Chen, G., Bandettini, P. A., & Constable, R. T. (2018). Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social

narrative. *Nature Communications*, *9*(1), 1–13. https://doi.org/10.1038/s41467-018-04387-2
Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., & Bandettini, P. A. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, *215*, 1–31. https://doi.org/10.1016/j.neuroimage.2020.116828
Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017, October 15). Can brain state be manipulated to emphasize individual differences in functional connectivity? *NeuroImage*. Academic Press Inc. https://doi.org/10.1016/j.neuroimage.2017.03.064
Fonov, VS, AC Evans, RC McKinstry, CR Almli, and DL Collins. 2009. “Unbiased Nonlinear Average Age-Appropriate Brain Templates from Birth to Adulthood.” NeuroImage 47, Supplement 1: S102. https://doi.org/10.1016/S1053-8119(09)70884-5.
Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102
Frazier, T. W., Strauss, M. E., & Steinhauer, S. R. (2004). Respiratory sinus arrhythmia as an index of emotional response in young adults. *Psychophysiology*, *41*(1), 75–83. https://doi.org/10.1046/j.1469-8986.2003.00131.x
Gaines, S. O. (2019). Personality Psychology. *Personality Psychology*. https://doi.org/10.4324/9780429056031
Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609. https://doi.org/10.1093/brain/119.2.593
Gao, J., Chen, G., Wu, J., Wang, Y., Hu, Y., Xu, T., … Yang, Z. (2020). Reliability map of individual differences reflected in inter-subject correlation in naturalistic imaging. *NeuroImage*, *223*(July), 117277. https://doi.org/10.1016/j.neuroimage.2020.117277
Gao, W., & Lin, W. (2012). Frontal parietal control network regulates the anti-correlated default and dorsal attention networks. *Human Brain Mapping*, *33*(1), 192–202. https://doi.org/10.1002/hbm.21204
Gaviria, J., Rey, G., Bolton, T., Ville, D. Van De, & Vuilleumier, P. (2021). Dynamic functional brain networks underlying the temporal inertia of negative emotions. *NeuroImage*, *240*(May), 118377. https://doi.org/10.1016/j.neuroimage.2021.118377
Geerligs, L., Rubinov, M., Cam-CAN, & Henson, R. N. (2015). State and Trait Components of Functional Connectivity: Individual Differences Vary with Mental State. *Journal of Neuroscience*, *35*(41), 13949–13961. https://doi.org/10.1523/JNEUROSCI.1324-15.2015
Gentili, C., Cristea, I. A., Ricciardi, E., Vanello, N., Popita, C., David, D., & Pietrini, P. (2017). Not in one metric: Neuroticism modulates different resting state metrics within distinctive brain regions. *Behavioural Brain Research*, *327*, 34–43. https://doi.org/10.1016/j.bbr.2017.03.031
Gilman, T. L., Shaheen, R., Nylocks, K. M., Halachoff, D., Chapman, J., Flynn, J. J., … Coifman, K. G. (2017). A film set for the elicitation of emotion in research: A comprehensive catalog derived from four decades of investigation. *Behavior Research Methods*, *49*(6), 2061–2082. https://doi.org/10.3758/s13428-016-0842-x
Gilson, M., Deco, G., Friston, K. J., Hagmann, P., Mantini, D., Betti, V., … Corbetta, M. (2017). Effective connectivity inferred from fMRI transition dynamics during movie

viewing points to a balanced reconfiguration of cortical interactions. *NeuroImage*, *180*(October 2017), 534–546. https://doi.org/10.1016/j.neuroimage.2017.09.061

Goldberg, L. R. (1990). Personality Processes and Individual Differences - an Alternative “Description of Personality”: The Big-Five Factor Structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229.

Goldin, P. R., Hutcherson, C. A. C., Ochsner, K. N., Glover, G. H., Gabrieli, J. D. E., & Gross, J. J. (2005). The neural bases of amusement and sadness: A comparison of block contrast and subject-specific emotion intensity regression approaches. *NeuroImage*, *27*(1), 26–36. https://doi.org/10.1016/j.neuroimage.2005.03.018

Goldstein, R. B., Woods, R. L., & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine*, *37*(7), 957–964. https://doi.org/10.1016/j.compbiomed.2006.08.018

Gonzalez-Castillo, J., Kam, J. W. Y., Hoy, C. W., & Bandettini, P. A. (2021). How to interpret resting-state fMRI : ask your participants . *Journal of Neuroscience*, *41*(6), 1130–1141.

Gordon, E. M., Breeden, A. L., Bean, S. E., & Vaidya, C. J. (2014). Working memory-related changes in functional connectivity persist beyond task disengagement. *Human Brain Mapping*, *35*(3), 1004–1017. https://doi.org/10.1002/hbm.22230

Gordon, E. M., Laumann, T. O., Marek, S., Raut, R. V., Gratton, C., Newbold, D. J., … Nelson, S. M. (2020). Default-mode network streams for coupling to language and control systems. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(29), 17308–17319. https://doi.org/10.1073/pnas.2005238117

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*(August). https://doi.org/10.3389/fninf.2011.00013

Gorgolewski, Krzysztof J., Oscar Esteban, Christopher J. Markiewicz, Erik Ziegler, David Gage Ellis, Michael Philipp Notter, Dorota Jarecka, et al. 2018. “Nipype.” Software. Zenodo. https://doi.org/10.5281/zenodo.596855.

Grady, C. L., Rieck, J. R., Baracchini, G., & Desouza, B. (2023). Relation of resting brain signal variability to cognitive and socioemotional measures in an adult lifespan sample. *Social Cognitive and Affective Neuroscience*, *18*(1). https://doi.org/10.1093/scan/nsad044

Grall, C., & Finn, E. S. (2022). Leveraging the power of media to drive cognition: A media-informed approach to naturalistic neuroscience. *Social Cognitive and Affective Neuroscience*, *17*(6), 598–608. https://doi.org/10.1093/scan/nsac019

Grandgirard, J., Poinsot, D., Krespi, L., Nénon, J. P., & Cortesero, A. M. (2002). Costs of secondary parasitism in the facultative hyperparasitoid Pachycrepoideus dubius: Does host size matter? *Entomologia Experimentalis et Applicata*, *103*(3), 239–248. https://doi.org/10.1023/A

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., … Petersen, S. E. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, *98*(2), 439–452. https://doi.org/10.1016/j.neuron.2018.03.035

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Grob, A. , Horowitz, D. (2014). Fragebogen zur Erhebung der Emotionsregulation bei Erwachsenen (FEEL-E). Bern: Huber.

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, *74*(1), 224–237. https://doi.org/10.1037//0022-3514.74.1.224

Gross, J. J., & Levenson, R. W. (1995). Emotion Elicitation using Films. *Cognition and Emotion*, *9*(1), 87–108. https://doi.org/10.1080/02699939508408966

Gross, J., & Kreibig, S. (2012). Understanding mixed emotions: paradigms and measures. *Physica Scripta*, *85*(5), 62–71. https://doi.org/10.1016/j.cobeha.2017.05.016

Guo, C. C., Nguyen, V. T., Hyett, M. P., Parker, G. B., & Breakspear, M. J. (2015). Out-of-sync: Disrupted neural activity in emotional circuitry during film viewing in melancholic depression. *Scientific Reports*, *5*(June), 1–12. https://doi.org/10.1038/srep11605

Hagemann, D., Naumann, E., Maier, S., Becker, G., Lürken, A., & Bartussek, D. (1999). The assessment of affective reactivity using fims: Validity, reliability and sex differences. *Personality and Individual Differences*, *26*, 627–639. https://doi.org/10.1017/CBO9781139193436.003

Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F. R., … Stadler, J. (2016). A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Scientific Data*, *3*, 160092. https://doi.org/10.1038/sdata.2016.92 LK - https://tdnetdiscover.com/shibboleth/go/773/resolver/full?sid=EMBASE&issn=20524463&id=doi:10.1038%2Fsdata.2016.92&atitle=A+studyforrest+extension%2C+simultaneous+fMRI+and+eye+gaze+recordings+during+prolonged+natural+stimulation&stitle=Sci+Data&title=Scientific+data&volume=3&issue=&spage=160092&epage=&aulast=Hanke&aufirst=Michael&auinit=M.&aufull=Hanke+M.&coden=&isbn=&pages=160092-&date=2016&auinit1=M&auinitm=

Hanke, Michael, Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., … Stadler, J. (2014). A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Scientific Data*, *1*. https://doi.org/10.1038/sdata.2014.3

Hanke, Michael, & Ibe, P. (2016). Lies, irony, and contradiction - an annotation of semantic conflict in the movie “Forrest Gump.” *F1000Research*, *5*, 1–9. https://doi.org/10.12688/f1000research.9635.1

Hanke, Michael, Labs, A., Reich, T., Schulenburg, H., Boennen, M., Mareike, G., … von Sobbe, F. R. (2015). Portrayed emotions in the movie “Forrest Gump.” *F1000Research*, *4*. https://doi.org/10.12688/f1000research.6230.1

Hart, B. M. t., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., & Einhäuser, W. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, *17*(6–7), 1132–1158. https://doi.org/10.1080/13506280902812304

Harwood, J., & Anderson, K. (2002). The presence and portrayal of social groups on prime–time television. *International Journal of Phytoremediation*, *21*(1), 81–97. https://doi.org/10.1080/08934210209367756

Hasson, U., Furman, O., Clark, D., Dudai, Y., & Davachi, L. (2008). Enhanced Intersubject Correlations during Movie Viewing Correlate with Successful Episodic Encoding. *Neuron*, *57*(3), 452–462. https://doi.org/10.1016/j.neuron.2007.12.009

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The Neuroscience of Film. *Projections*, *2*(1), 1–26. https://doi.org/10.3167/proj.2008.020102
Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48. https://doi.org/10.1016/j.tics.2009.10.011
Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science*, *303*(MARCH), 1634–1640. https://doi.org/10.1126/science.1089506
Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, *28*(10), 2539–2550. https://doi.org/10.1523/JNEUROSCI.5487-07.2008
Häusler, C. O., & Hanke, M. (2021). A studyforrest extension, an annotation of spoken language in the German dubbed movie "Forrest Gump" and its audio-description. *F1000Research*, *10*. https://doi.org/10.12688/f1000research.27621.1
Hautzinger, M., Keller, F., Kühner, C. (2009). Beck-Depressions-Inventar Revision. 2. Auflage. Frankfurt am Main: Pearson Assessment.
Hlinka, J., Děchtěrenko, F., Rydlo, J., Androvičová, R., Vejmelka, M., Jajcay, L., … Horáček, J. (2022). The intra-session reliability of functional connectivity during naturalistic viewing conditions. *Psychophysiology*, (March), 1–13. https://doi.org/10.1111/psyp.14075
Hoffner, C. A., & Levine, K. J. (2013). Enjoyment of mediated fright and violence: A meta-analysis. *Mass Media Effects Research: Advances Through Meta-Analysis*, (September 2013), 215–244. https://doi.org/10.4324/9780203823453
Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., … Chang, C. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, *80*, 360–378. https://doi.org/10.1016/j.neuroimage.2013.05.079
Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, *60*(October 2008), 653–670. https://doi.org/10.1146/annurev.psych.60.110707.163604
Jääskeläinen, I. P., Ahveninen, J., Klucharev, V., Shestakova, A. N., & Levy, J. (2022). Behavioral Experience-Sampling Methods in Neuroimaging Studies With Movie and Narrative Stimuli. *Frontiers in Human Neuroscience*, *16*(January), 1–7. https://doi.org/10.3389/fnhum.2022.813684
Jääskeläinen, I. P., Koskentalo, K., Balk, M. H., Autti, T., Kauramäki, J., Pomren, C., & Sams, M. (2008). Inter-Subject Synchronization of Prefrontal Cortex Hemodynamic Activity During Natural Viewing. *The Open Neuroimaging Journal*, *2*(1), 14–19. https://doi.org/10.2174/1874440000802010014
Jääskeläinen, I. P., Pajula, J., Tohka, J., Lee, H. J., Kuo, W. J., & Lin, F. H. (2016). Brain hemodynamic activity during viewing and re-viewing of comedy movies explained by experienced humor. *Scientific Reports*, *6*(December 2015), 1–14. https://doi.org/10.1038/srep27741
Jang, C., Knight, E. Q., Pae, C., Park, B., Yoon, S. A., & Park, H. J. (2017). Individuality manifests in the dynamic reconfiguration of large-scale brain networks during movie viewing. *Scientific Reports*, *7*(January), 1–14. https://doi.org/10.1038/srep41414

Jenkins, L. M., & Andrewes, D. G. (2012). A new set of standardised verbal and non-verbal contemporary film stimuli for the elicitation of emotions. *Brain Impairment*, *13*(2), 212–227. https://doi.org/10.1017/BrImp.2012.18

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jiang, T., Li, H., & Hou, Y. (2019). Cultural differences in humor perception, usage, and implications. *Frontiers in Psychology*, *10*(JAN), 1–8. https://doi.org/10.3389/fpsyg.2019.00123

Kanning, U.P. (2002). Soziale Kompetenz - Definition, Strukturen und Prozesse. *Zeitschrift Für Psychologie-Journal of Psychology*, 210, 154-163.

Kanning, U. P. (2003). Diagnostik sozialer Kompetenzen. Kompendien Psychologische Diagnostik Bd. 4. Göttingen: Hogrefe.

Kauppi, J. P., Jääskeläinen, I. P., Sams, M., & Tohka, J. (2010). Inter-subject correlation of brain hemodynamic responses during watching a movie: Localization in space and frequency. *Frontiers in Neuroinformatics*, *4*(MAR). https://doi.org/10.3389/fninf.2010.00005

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

Kim, D., Kay, K., Shulman, G. L., & Corbetta, M. (2018). A new modular brain organization of the bold signal during natural vision. *Cerebral Cortex*, *28*(9), 3065–3081. https://doi.org/10.1093/cercor/bhx175

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340. https://doi.org/10.1348/000712607X251243

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The “Trier Social Stress Test” - A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, *3*(28), 76–81.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., … Keshavan, A. (2017). *Mindboggling morphometry of human brains*. *PLoS Computational Biology* (Vol. 13). https://doi.org/10.1371/journal.pcbi.1005350

Knautz, K., & Stock, W. G. (2011). Collective indexing of emotions in videos. *Journal of Documentation*, *67*(6), 975–994. https://doi.org/10.1108/00220411111183555

Knight, L. K., Stoica, T., Fogleman, N. D., & Depue, B. E. (2019). Convergent neural correlates of empathy and anxiety during socioemotional processing. *Frontiers in Human Neuroscience*, *13*(March), 1–15. https://doi.org/10.3389/fnhum.2019.00094

Knowles, K. A., & Bunmi O. Olatunji. (2020). Specificity of Trait Anxiety in Anxiety and Depression: Meta- Analysis of the State-Trait Anxiety Inventory. *Clin Psychol Rev*, *82*. https://doi.org/10.1016/j.cpr.2020.101928. Specificity

Koelstra, S., Mühl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., … Patras, I. (2012). DEAP: A database for emotion analysis; Using physiological signals. *IEEE Transactions on Affective Computing*, *3*(1), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

Kong, R., Li, J., Orban, C., Sabuncu, M. R., Liu, H., Schaefer, A., … Yeo, B. T. T. (2018). Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *Cerebral Cortex*, 1–19. https://doi.org/10.1093/cercor/bhy123

Kotthoff, H. (2006). Gender and humor: The state of the art. *Journal of Pragmatics*, *38*(1 SPEC. ISS.), 4–25. https://doi.org/10.1016/j.pragma.2005.06.003

Kraus, B., Zinbarg, R., Braga, R. M., Nusslock, R., Mittal, V. A., & Gratton, C. (2023). Insights from personalized models of brain and behavior for identifying biomarkers in psychiatry. *Neuroscience and Biobehavioral Reviews*, *152*(May), 105259. https://doi.org/10.1016/j.neubiorev.2023.105259

Kreibig, S. D., Samson, A. C., & Gross, J. J. (2013). The psychophysiology of mixed emotional states. *Psychophysiology*, *50*(8), 799–811. https://doi.org/10.1111/psyp.12064

Kreibig, S. D., Samson, A. C., & Gross, J. J. (2015). The psychophysiology of mixed emotional states: Internal and external replicability analysis of a direct replication study. *Psychophysiology*, *52*(7), 873–886. https://doi.org/10.1111/psyp.12425

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(NOV), 1–28. https://doi.org/10.3389/neuro.06.004.2008

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchung mit einer deutschen Version der “Positive and Negative Affect Schedule (PANAS).” *Diagnostica*, *42*(2), 139–156.

Kröll, J. P., Friedrich, P., Li, X., Patil, K. R., Mochalski, L., Waite, L., … Weis, S. (2023). Naturalistic viewing increases individual identifiability based on connectivity within functional brain networks. *NeuroImage*, *273*(April), 120083. https://doi.org/10.1016/j.neuroimage.2023.120083

Kucyi, A. (2018). Just a thought: How mind-wandering is represented in dynamic brain connectivity. *NeuroImage*, *180*(April 2017), 505–514. https://doi.org/10.1016/j.neuroimage.2017.07.001

Kuipers, G., & Van Der Ent, B. (2016). The seriousness of ethnic jokes: Ethnic humor and social change in the Netherlands, 1995-2012. *Humor*, *29*(4), 605–633. https://doi.org/10.1515/humor-2016-0013

Kupfer, J., Brosig, B., Brähler, E. (2001). Toronto-Alexithymie-Skala-26 Deutsche Version. 1. Auflage. Göttingen: Hogrefe.

Kuppens, P., Tuerlinckx, F., Yik, M., Koval, P., Coosemans, J., Zeng, K. J., & Russell, J. A. (2017). The Relation Between Valence and Arousal in Subjective Experience Varies With Personality and Culture. *Journal of Personality*, *85*(4), 530–542. https://doi.org/10.1111/jopy.12258

Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., & Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, *100*, 316–324. https://doi.org/10.1016/j.neuroimage.2014.06.022

Lahnakoski, J. M., Glerean, E., Salmi, J., Jääskeläinen, I. P., Sams, M., Hari, R., & Nummenmaa, L. (2012). Naturalistic fMRI Mapping Reveals Superior Temporal Sulcus as the Hub for the Distributed Brain Network for Social Perception. *Frontiers in Human Neuroscience*, *6*(August), 1–14. https://doi.org/10.3389/fnhum.2012.00233

Lanczos, C. 1964. “Evaluation of Noisy Data.” Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis 1 (1): 76–85. https://doi.org/10.1137/0701007.

Lang, P. J. The Cognitive Psychophysiology of Emotion: Anxiety and the Anxiety Disorders. Hillsdale, NJ: Lawrence Erlbaum, 1985

Langner, R., & Eickhoff, S. B. (2013). Sustaining Attention to Simple Tasks: A Meta-Analytic Review of the Neural Mechanisms of Vigilant Attention, *71*(2), 233–236. https://doi.org/10.1038/mp.2011.182.doi

Laux, L., Glanzmann, P., Schaffner, P., Spielberger, C. D. (1981). Das State-Trait-Angstinventar. 1. Auflage. Göttingen: Beltz Test.

Lauzen, M. M., & Dozier, D. M. (2005). Maintaining the double standard: Portrayals of age and gender in popular films. *Sex Roles*, *52*(7–8), 437–446. https://doi.org/10.1007/s11199-005-3710-1

Leal, P. C., Goes, T. C., da Silva, L. C. F., & Teixeira-Silva, F. (2017). Trait vs. state anxiety in different threatening situations. *Trends in Psychiatry and Psychotherapy*, *39*(3), 147–157. https://doi.org/10.1590/2237-6089-2016-0044

Leavitt, P. A., Covarrubias, R., Perez, Y. A., & Fryberg, S. A. (2015). “Frozen in time”: The impact of native American media representations on identity and self-understanding. *Journal of Social Issues*, *71*(1), 39–53. https://doi.org/10.1111/josi.12095

Lee, S., Bijsterbosch, J. D., Almagro, F. A., Elliott, L., McCarthy, P., Taschler, B., … Douaud, G. (2023). Amplitudes of resting-state functional networks – investigation into their correlates and biophysical properties. *NeuroImage*, *265*(July 2022), 119779. https://doi.org/10.1016/j.neuroimage.2022.119779

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, *31*(8), 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011

Lettieri, G., Handjaras, G., Ricciardi, E., Leo, A., Papale, P., Betta, M., … Cecchetti, L. (2019). Emotionotopy in the human right temporo-parietal cortex. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-13599-z

Lettieri, G., Handjaras, G., Setti, F., Cappello, E. M., Bruno, V., Diano, M., … Cecchetti, L. (2022). Default and control network connectivity dynamics track the stream of affect at multiple timescales. *Social Cognitive and Affective Neuroscience*, *17*(5), 461–469. https://doi.org/10.1093/scan/nsab112

Levakov, G., Sporns, O., & Avidan, G. (2023). Fine-scale dynamics of functional connectivity in the face-processing network during movie watching. *Cell Reports*, *42*(6), 112585. https://doi.org/10.1016/j.celrep.2023.112585

Li, X., Friedrich, P., Patil, K. R., Eickhoff, S. B., & Weis, S. (2023). A topography-based predictive framework for naturalistic viewing fMRI. *NeuroImage*, *277*(January), 120245. https://doi.org/10.1016/j.neuroimage.2023.120245

Liu, H., Cao, J., Zhang, J., & Ragulskis, M. (2023). Minimum spanning tree brain network topology reflects individual differences in the structure of affective experience. *Neurocomputing*, *521*, 56–64. https://doi.org/10.1016/j.neucom.2022.11.095

Liu, T., Hu, X., Li, X., Chen, M., Han, J., & Guo, L. (2014). Merging neuroimaging and multimedia: Methods, opportunities, and challenges. *IEEE Transactions on Human-Machine Systems*, *44*(2), 270–280. https://doi.org/10.1109/THMS.2013.2296871

Liu, T. T., & Falahpour, M. (2020). Vigilance Effects in Resting-State fMRI. *Frontiers in Neuroscience*, *14*(April). https://doi.org/10.3389/fnins.2020.00321

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1219–1236. https://doi.org/10.1016/j.neubiorev.2010.12.012.Common

Lu, K. H., Hung, S. C., Wen, H., Marussich, L., & Liu, Z. (2016). Influences of high-level features, gaze, and scene transitions on the reliability of BOLD responses to natural movie stimuli. *PLoS ONE*, *11*(8), 1–19. https://doi.org/10.1371/journal.pone.0161797
Mantini, D., Hasson, U., Betti, V., Perrucci, M. G., Romani, G. L., Corbetta, M., … Vanduffel, W. (2012). Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. *Nature Methods*, *9*(3), 277–282. https://doi.org/10.1038/nmeth.1868
Mar, R. A., Tackett, J. L., & Moore, C. (2010). Exposure to media and theory-of-mind development in preschoolers. *Cognitive Development*, *25*(1), 69–78. https://doi.org/10.1016/j.cogdev.2009.11.002
Marchant, P., Raybould, D., Renshaw, T., & Stevens, R. (2009). Are you seeing what I'm seeing? An eye-tracking evaluation of dynamic scenes. *Digital Creativity*, *20*(3), 153–163. https://doi.org/10.1080/14626260903083611
Markett, S., Montag, C., Melchers, M., Weber, B., & Reuter, M. (2016). Anxious personality and functional efficiency of the insular-opercular network: A graph-analytic approach to resting-state fMRI. *Cognitive, Affective and Behavioral Neuroscience*, *16*(6), 1039–1049. https://doi.org/10.3758/s13415-016-0451-2
Martin, C. G., He, B. J., & Chang, C. (2021). State-related neural influences on fMRI connectivity estimation. *NeuroImage*, *244*(May), 118590. https://doi.org/10.1016/j.neuroimage.2021.118590
Mastro, D. (2015). Why the media's role in issues of race and ethnicity should be in the spotlight. *Journal of Social Issues*, *71*(1), 1–16. https://doi.org/10.1111/josi.12093
Mastro, D. E., & Behm-Morawitz, E. (2005). Latino representation on primetime television. *Journalism and Mass Communication Quarterly*, *82*(1), 110–130. https://doi.org/10.1177/107769900508200108
Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state big three. Personality psychology in Europe, 7, 335-350. Tilburg: Tilburg University Press.
Matthews, G., Szalma, J., Panganiban, A. R., Neubauer, C., & Warm, J. S. (2013). *Profiling task stress with the dundee state questionnaire*. *Psychology of Stress: New Research*.
McCrae, R., & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x
McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, *36*(3), 587–596. https://doi.org/10.1016/S0191-8869(03)00118-1
McNamara, Q., De La Vega, A., & Yarkoni, T. (2017). Developing a comprehensive framework for multimodal feature extraction. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F1296*, 1567–1574. https://doi.org/10.1145/3097983.3098075
Meer, J. N. va. der, Breakspear, M., Chang, L. J., Sonkusare, S., & Cocchi, L. (2020). Movie viewing elicits rich and reliable brain state dynamics. *Nature Communications*, *11*(1), 1–14. https://doi.org/10.1038/s41467-020-18717-w
Mendiburo-Seguel, A., Páez, D., & Martínez-Sánchez, F. (2015). Humor styles and personality: A meta-analysis of the relation between humor styles and the Big Five personality traits. *Scandinavian Journal of Psychology*, *56*(3), 335–340. https://doi.org/10.1111/sjop.12209

Metallinou, A., & Narayanan, S. (2013). Annotation and processing of continuous emotional attributes: Challenges and opportunities. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013*, 1–8. https://doi.org/10.1109/FG.2013.6553804

Michon, K. J., Khammash, D., Simmonite, M., Hamlin, A. M., & Polk, T. A. (2022). Person-specific and precision neuroimaging: Current methods and future directions. *NeuroImage*, *263*(July), 119589. https://doi.org/10.1016/j.neuroimage.2022.119589

Mirchi, N., Betzel, R. F., Bernhardt, B. C., Dagher, A., & Mišić ‡, B. (2019). Tracking mood fluctuations with functional network patterns. *Social Cognitive and Affective Neuroscience*, *14*(1), 47–57. https://doi.org/10.1093/scan/nsy107

Mišić, B., & Sporns, O. (2016). From regions to connections and networks: New bridges between brain and behavior. *Current Opinion in Neurobiology*, *40*, 1–7. https://doi.org/10.1016/j.conb.2016.05.003

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion. *Cognitive Computation*, *3*(1), 5–24. https://doi.org/10.1007/s12559-010-9074-z

Moon, S. E., & Lee, J. S. (2017). Implicit Analysis of Perceptual Multimedia Experience Based on Physiological Response: A Review. *IEEE Transactions on Multimedia*, *19*(2), 340–353. https://doi.org/10.1109/TMM.2016.2614880

Morris, J. D. (1995). SAM: The Self-Assessment Manikin - An efficient cross-cultural measurement of emotional response. *Journal of Advertising Research*, *35*(6), 63–68. https://doi.org/http://adsam.com/files/Observations.PDF

Nakuci, J., Wasylyshyn, N., Cieslak, M., Elliott, J. C., Bansal, K., Giesbrecht, B., … Muldoon, S. F. (2023). Within-subject reproducibility varies in multi-modal, longitudinal brain networks. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-33441-3

Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, *14*(6), 669–687. https://doi.org/10.1093/scan/nsz037

Neville, C., & Anastasio, P. (2019). Fewer, Younger, but Increasingly Powerful: How Portrayals of Women, Age, and Power Have Changed from 2002 to 2016 in the 50 Top-Grossing U.S. Films. *Sex Roles*, *80*(7–8), 503–514. https://doi.org/10.1007/s11199-018-0945-1

Nguyen, V. T., Sonkusare, S., Stadler, J., Hu, X., Breakspear, M., & Guo, C. C. (2017). Distinct cerebellar contributions to cognitive- perceptual dynamics during natural viewing. *Cerebral Cortex*, *27*(12), 5652–5662. https://doi.org/10.1093/cercor/bhw334

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, *66*(6), 574–583. https://doi.org/10.1037/h0040291

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., & Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(24), 9599–9604. https://doi.org/10.1073/pnas.1206095109

Nummenmaa, L., Lahnakoski, J. M., & Glerean, E. (2018). Sharing the social world via intersubject neural synchronisation. *Current Opinion in Psychology*, *24*, 7–14. https://doi.org/10.1016/j.copsyc.2018.02.021

Nyström, M., & Holmqvist, K. (2010). Effect of compressed offline foveated video on viewing behavior and subjective quality. *ACM Transactions on Multimedia Computing,*

Communications and Applications, *6*(1), 0–14. https://doi.org/10.1145/1671954.1671958
O'Connor, D., Potler, N. V., Kovacs, M., Xu, T., Ai, L., Pellman, J., … Milham, M. P. (2017). The healthy brain network serial scanning initiative: A resource for evaluating inter-individual differences and their reliabilities across scan conditions and sessions. *GigaScience*, *6*(2), 1–14. https://doi.org/10.1093/gigascience/giw011
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4
Oliveira, E., Martins, P., & Chambel, T. (2013). Accessing movies based on emotional impact. *Multimedia Systems*, *19*(6), 559–576. https://doi.org/10.1007/s00530-013-0303-7
Palomba, D., Sarlo, M., Angrilli, A., Mini, A., & Stegagno, L. (2000). Cardiac responses associated with affective processing of unpleasant film stimuli. *International Journal of Psychophysiology*, *36*(1), 45–57. https://doi.org/10.1016/S0167-8760(99)00099-9
Pan, J., Zhan, L., Hu, C. L., Yang, J., Wang, C., Gu, L., … Wu, X. (2018). Emotion regulation and complex brain networks: Association between expressive suppression and efficiency in the fronto-parietal network and default-mode network. *Frontiers in Human Neuroscience*, *12*(March), 1–12. https://doi.org/10.3389/fnhum.2018.00070
Pervaiz, U., Vidaurre, D., Woolrich, M. W., & Smith, S. M. (2020). Optimising network modelling methods for fMRI. *NeuroImage*, *211*(November 2019), 116604. https://doi.org/10.1016/j.neuroimage.2020.116604
Porter, A., Nielsen, A., Dorn, M., Dworetsky, A., Edmonds, D., & Gratton, C. (2023). Masked features of task states found in individual brain networks. *Cerebral Cortex*, *33*(6), 2879–2900. https://doi.org/10.1093/cercor/bhac247
Power, J. D., Braver, T. S., Petersen, S. E., Cole, M. W., & Bassett, D. S. (2014). Intrinsic and Task-Evoked Network Architectures of the Human Brain. *Neuron*, *83*(1), 238–251. https://doi.org/10.1016/j.neuron.2014.05.014
Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., … Petersen, S. E. (n.d.). Functional network organization of the human brain. *Bone*, *72*(4), 665–678. https://doi.org/10.1016/j.neuron.2011.09.006
Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048
Preti, M. G., Bolton, T. A., & Van De Ville, D. (2017). The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage*, *160*(December 2016), 41–54. https://doi.org/10.1016/j.neuroimage.2016.12.061
Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, *112*, 267–277. https://doi.org/10.1016/j.neuroimage.2015.02.064
Ramasubramanian, S. (2011). The impact of stereotypical versus counterstereotypical media exemplars on racial attitudes, causal attributions, and support for affirmative action. *Communication Research*, *38*(4), 497–516. https://doi.org/10.1177/0093650210384854
Ramasubramanian, S., Doshi, M. J., & Saleem, M. (2017). Mainstream versus ethnic media: How they shape ethnic pride and self-esteem among ethnic minority audiences. *International Journal of Communication*, *11*(January 2017), 1879–1899.

Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., … Barrett, L. F. (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cognitive, Affective and Behavioral Neuroscience*, *16*(4), 709–723. https://doi.org/10.3758/s13415-016-0425-4

Redcay, E., & Moraczewski, D. (2019). Social cognition in context: A naturalistic imaging approach. *NeuroImage*, (October), 116392. https://doi.org/10.1016/j.neuroimage.2019.116392

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Reggente, N., Essoe, J. K. Y., Aghajan, Z. M., Tavakoli, A. V., McGuire, J. F., Suthana, N. A., & Rissman, J. (2018). Enhancing the ecological validity of fMRI memory research using virtual reality. *Frontiers in Neuroscience*, *12*(JUN), 1–9. https://doi.org/10.3389/fnins.2018.00408

Ren, Y., Nguyen, V. T., Sonkusare, S., Lv, J., Pang, T., Guo, L., … Guo, C. C. (2018). Effective connectivity of the anterior hippocampus predicts recollection confidence during natural memory retrieval. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-07325-4

Reuter, M., Panksepp, J., Davis, K., Montag, C. (2017). Affective Neuroscience Personality Scales (ANPS): Deutsche Version. 1. Auflage. Göttingen: Hogrefe.

Rindermann, H. (2009). Emotionale-Kompetenz-Fragebogen: Einschätzung emotionaler Kompetenzen und emotionaler Intelligenz aus Selbst- und Fremdsicht. 1. Auflage. Göttingen: Hogrefe.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology*, *210*(5–6), 419–421. https://doi.org/10.1007/s00429-005-0039-z

Rizzolatti, G., & Fabbri-Destro, M. (2008). The mirror system and its role in social cognition. *Current Opinion in Neurobiology*, *18*(2), 179–184. https://doi.org/10.1016/j.conb.2008.08.001

Rizzolatti, G., & Sinigaglia, C. (2016). The mirror mechanism: A basic principle of brain function. *Nature Reviews Neuroscience*, *17*(12), 757–765. https://doi.org/10.1038/nrn.2016.135

Rohrmann, S., Hodapp, V., Schnell, K., Tibubos, A., Schwenkmezger, P., Spielberger, C.D. (2013). Das State-Trait-Ärgerausdrucks-Inventar - 2: Deutschsprachige Adaptation des State-Trait Anger Expression Inventory-2 (STAXI-2) von Charles D. Spielberger. Bern: Huber.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., … Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, *60*(1), 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050

Saarimäki, H. (2021). Naturalistic Stimuli in Affective Neuroimaging: A Review. *Frontiers in Human Neuroscience*, *15*(June). https://doi.org/10.3389/fnhum.2021.675068

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., … Jeffries, J. (2011). Emotional perception: Meta-analyses of face and natural scene processing. *NeuroImage*, *54*(3), 2524–2533. https://doi.org/10.1016/j.neuroimage.2010.10.011

Salehi, M., Karbasi, A., Barron, D. S., Scheinost, D., & Constable, R. T. (2020). Individualized functional networks reconfigure with cognitive state. *NeuroImage*, *206*(September 2019), 116233. https://doi.org/10.1016/j.neuroimage.2019.116233

Salerno, J. M., & Peter-Hagene, L. C. (2013). The Interactive Effect of Anger and Disgust on Moral Outrage and Judgments. *Psychological Science*, *24*(10), 2069–2078. https://doi.org/10.1177/0956797613486988

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

Samson, A. C., Kreibig, S. D., Soderstrom, B., Wade, A. A., & Gross, J. J. (2016). Eliciting positive, negative and mixed emotional states: A film library for affective scientists. *Cognition and Emotion*, *30*(5), 827–856. https://doi.org/10.1080/02699931.2015.1031089

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., … Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*(1), 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Sawahata, Y., Khosla, R., Komine, K., Hiruma, N., Itou, T., Watanabe, S., … Issiki, N. (2008). Determining comprehension and quality of TV programs using eye-gaze tracking. *Pattern Recognition*, *41*(5), 1610–1626. https://doi.org/10.1016/j.patcog.2007.10.010

Schaefer, Alexander, Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., … Yeo, B. T. T. (2017). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, (July 2017), 1–20. https://doi.org/10.1093/cercor/bhx179

Schaefer, Alexandre, Nils, F., Philippot, P., & Sanchez, X. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, *24*(7), 1153–1172. https://doi.org/10.1080/02699930903274322

Scharrer, E., Ramasubramanian, S., & Banjo, O. (2022). Media, Diversity, and Representation in the U.S.: A Review of the Quantitative Research Literature on Media Content and Effects. *Journal of Broadcasting and Electronic Media*, *66*(4), 723–749. https://doi.org/10.1080/08838151.2022.2138890

Schermer, J. A., Kwiatkowska, M. M., Kowalski, C. M., Aquino, S., Ardi, R., Bolló, H., … Crusius, J. (2019). Humor styles across 28 countries. *Current Psychology*, 1–16.

Schmidt, S., Petermann, F. (2009). ADHS-Screening für Erwachsene: Ein Verfahren zur Erfassung von Symptomen einer ADHS. 2., korrigierte und überarbeitete Auflage. Frankfurt am Main: Pearson Assessment.

Schoon, I. (2021). Towards an Integrative Taxonomy of Social-Emotional Competences. *Frontiers in Psychology*, *12*(March), 1–9. https://doi.org/10.3389/fpsyg.2021.515313

Schwarz, G. (1982). Estimating the Dimension of a Model. *Annals of Statistics*, *6*(2), 461–464.

Sengupta, A., Kaule, F. R., Guntupalli, J. S., Hoffmann, M. B., Häusler, C., Stadler, J., & Hanke, M. (2016). A studyforrest extension, retinotopic mapping and localization of higher visual areas. *Scientific Data*, *3*, 1–14. https://doi.org/10.1038/sdata.2016.93

Shafto, M. A., Tyler, L. K., Dixon, M., Taylor, J. R., Rowe, J. B., Cusack, R., … Matthews, F. E. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, *14*(204), 1—25. https://doi.org/10.1038/nrn1809

Shaheen J. G. (2003). Reel bad Arabs: How Hollywood vilifies a people. A*nnals of the American Academy of Political and Social Science,* 588, 171–193.
https://doi.org/10.1177/0002716203588001011

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspectives on Psychological Science*, *14*(5), 841–859. https://doi.org/10.1177/1745691619856350

Simony, E., & Chang, C. (2020). Analysis of stimulus-induced brain dynamics during naturalistic paradigms. *NeuroImage*, *216*(July 2019), 116461. https://doi.org/10.1016/j.neuroimage.2019.116461

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications*, *7*(May 2015), 1–13. https://doi.org/10.1038/ncomms12141

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., … Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(31), 13040–13045. https://doi.org/10.1073/pnas.0905267106

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*(8), 1–24. https://doi.org/10.1167/13.8.16

Soleymani, M., Chanel, G., Kierkels, J. J. M., & Pun, T. (2009). Movie Scenes Based on Content Analysis and Physiological Changes. *International Journal of Semantic Computing*, *3*(2), 235–254. https://doi.org/10.1142/S1793351X09000744

Soleymani, M., Pantic, M., & Pun, T. (2015). Multimodal emotion recognition in response to videos (Extended abstract). *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, *3*(2), 491–497. https://doi.org/10.1109/ACII.2015.7344615

Song, H., & Rosenberg, M. D. (2021). Predicting attention across time and contexts with functional brain connectivity. *Current Opinion in Behavioral Sciences*, *40*, 33–44. https://doi.org/10.1016/j.cobeha.2020.12.007

Sonkusare, S., Breakspear, M., & Guo, C. (2019, August 1). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences*. Elsevier Ltd. https://doi.org/10.1016/j.tics.2019.05.004

Soto, C. J., Napolitano, C. M., & Roberts, B. W. (2021). Taking Skills Seriously: Toward an Integrative Model and Agenda for Social, Emotional, and Behavioral Skills. *Current Directions in Psychological Science*, *30*(1), 26–33. https://doi.org/10.1177/0963721420978613

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201–292.

Spielberger, C.D., Gorsuch, R.L. & Lushene, R.E.: *Manual for the State-Trait-Anxiety Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1970.

Spielberger, C. D. (1972). *Anxiety As an Emotional State*. *Anxiety*. ACADEMIC PRESS, INC. https://doi.org/10.1016/b978-0-12-657401-2.50009-5

Spielberger, C. D., & Reheiser, E. C. (2009). Assessment of Emotions: Anxiety, Anger, Depression, and Curiosity. *Applied Psychology: Health and Well-Being*, *1*(3), 271–302. https://doi.org/10.1111/j.1758-0854.2009.01017.x

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience*, *17*(5), 652–660. https://doi.org/10.1038/nn.3690

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2008). The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-analysis, 489–510. https://doi.org/10.1162/jocn.2008.21029

Sun, Y., Ma, J., Huang, M., Yi, Y., Wang, Y., Gu, Y., … Dai, Z. (2022). Functional connectivity dynamics as a function of the fluctuation of tension during film watching. *Brain Imaging and Behavior*, *16*(3), 1260–1274. https://doi.org/10.1007/s11682-021-00593-7

Tarvainen, J., Member, S., Sjöberg, M., Member, G. S., & Westman, S. (2014). Content-Based Prediction ofMovie Style, Aesthetics, and Affect: Data Set and Baseline Experiments. *IEEE TRANSACTIONS ON MULTIMEDIA*, *16*(8), 2085–2098. https://doi.org/10.3989/sefarad.2006.v66.i2.415

Tarvainen, J., Westman, S., & Oittinen, P. (2015). The way films feel: Aesthetic features and mood in film. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(3), 254–265. https://doi.org/10.1037/a0039432

Taylor, G. J. ., Ryan, D., & Bagby, R. M. (1985). Toward the Development of a New Self-Report Alexithymia Scale. *Psychotherapy and Psychosomatics*, *44*(4), 191–199. https://doi.org/https://doi.org/10.1159/000287912

Teixeira, R. M. A., Yamasaki, T., & Aizawa, K. (2012). Determination of emotional content of video clips by low-level audiovisual features: A dimensional and categorial experimental approach. *Multimedia Tools and Applications*, *61*(1), 21–49. https://doi.org/10.1007/s11042-010-0702-0

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Tian, Leimin, Muszynski, M., Lai, C., Moore, J. D., Kostoulas, T., Lombardo, P., … Chanel, G. (2018). Recognizing induced emotions of movie audiences: Are induced and perceived emotions the same? *2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017*, *2018-Janua*, 28–35. https://doi.org/10.1109/ACII.2017.8273575

Tian, Lixia, Ye, M., Chen, C., Cao, X., & Shen, T. (2021). NeuroImage Consistency of functional connectivity across different movies. *NeuroImage*, *233*(March), 117926. https://doi.org/10.1016/j.neuroimage.2021.117926

Treiblmaier, H., & Filzmoser, P. (2010). Exploratory factor analysis revisited: How robust methods support the detection of hidden multivariate data structures in IS research. *Information and Management*, *47*(4), 197–207. https://doi.org/10.1016/j.im.2010.02.002

Tupes, E. C., & Christal, R. E. (1992). Recurrent Personality Factors Based on Trait Ratings (Tech. Rep. No. ASD-TR-61-97). *Journal of Personality*, *60*(2), 225–251.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Van Atteveldt, N., Van Kesteren, M. T. R., Braams, B., & Krabbendam, L. (2018). Neuroimaging of learning and development: Improving ecological validity. *Frontline Learning Research*, *6*(3), 186–203. https://doi.org/10.14786/flr.v6i3.366

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Vanderwal, T., Eilbott, J., & Castellanos, F. X. (2019). Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Developmental Cognitive Neuroscience*, *36*(October 2018), 100600. https://doi.org/10.1016/j.dcn.2018.10.004
Vanderwal, T., Eilbott, J., Finn, E. S., Craddock, R. C., Turnbull, A., & Castellanos, F. X. (2017). Individual differences in functional connectivity during naturalistic viewing conditions. *NeuroImage*, *157*(June), 521–530. https://doi.org/10.1016/j.neuroimage.2017.06.027
Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C., & Castellanos, F. X. (2015). Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage*, *122*, 222–232. https://doi.org/10.1016/j.neuroimage.2015.07.069
Van Dijk, Wilco W., and Jaap W. Ouwerkerk. (2014). Schadenfreude: Understanding Pleasure at the Misfortune of Others. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139084246
Vig, E., Dorr, M., & Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. *Spatial Vision*, *22*(5), 397–408. https://doi.org/10.1163/156856809789476065
Viinikainen, M., Glerean, E., Jääskeläinen, I. P., Kettunen, J., Sams, M., & Nummenmaa, L. (2012). Nonlinear neural representation of emotional feelings elicited by dynamic naturalistic stimulation. *Open Journal of Neuroscience*, *2*(1).
Visconti di Oleggio Castello, M., Chauhan, V., Jiahui, G., & Gobbini, M. I. (2020). The Grand Budapest Hotel: an fMRI dataset in response to a socially-rich, naturalistic movie. *Scientific Data*, 1–9. https://doi.org/10.1101/2020.07.14.203257
Wang, H. L., & Cheong, L. F. (2006). Affective understanding in film. *IEEE Transactions on Circuits and Systems for Video Technology*, *16*(6), 689–704. https://doi.org/10.1109/TCSVT.2006.873781
Wang, J., Ren, Y., Hu, X., Nguyen, V. T., Guo, L., Han, J., & Guo, C. C. (2017). Test–retest reliability of functional connectivity networks during naturalistic fMRI paradigms. *Human Brain Mapping*, *38*(4), 2226–2241. https://doi.org/10.1002/hbm.23517
Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. *Revista Brasileira de Psiquiatria*, *35*(4), 416–431. https://doi.org/10.1590/1516-4446-2012-1048
Wang, Z., Yang, J., Zheng, Z., Cao, W., Dong, L., Li, H., … Yao, D. (2023). Trait- and State-Dependent Changes in Cortical–Subcortical Functional Networks Across the Adult Lifespan. *Journal of Magnetic Resonance Imaging*, *58*(3), 720–731. https://doi.org/10.1002/jmri.28599
Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1051/epjconf/201714006017
Wender, P. H. (1995). Attention-Deficit Hyperactivity Disorder in Adults. New York: Oxford University Press.
Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: a meta- analysis RAINER. *European Journal of Social Psychology*, *26*, 557–580.
Whitton, A. E., Henry, J. D., Rendell, P. G., & Grisham, J. R. (2014). Disgust, but not anger provocation, enhances levator labii superioris activity during exposure to moral

transgressions. *Biological Psychology*, *96*(1), 48–56. https://doi.org/10.1016/j.biopsycho.2013.11.012
Witt, S. T., Meyerand, M. E., & Laird, A. R. (2008). Functional neuroimaging correlates of finger tapping task variations: An ALE meta-analysis, *71*(2), 233–236. https://doi.org/10.1038/mp.2011.182.doi
Wolf, I., Dziobek, I., & Heekeren, H. R. (2010). Neural correlates of social cognition in naturalistic settings: A model-free analysis approach. *NeuroImage*, *49*(1), 894–904. https://doi.org/10.1016/j.neuroimage.2009.08.060
Yates, C. & Dennedy, D. MLT framework (Version 0.8.0) [computer program]. http://www.mltframework.org Retrieved from the Debian archive at version 0.8.0-4 (2012).
Yik, M. S. M., Russell, J. A., & Barrett, L. F. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, *77*(3), 600–619. https://doi.org/10.1037/0022-3514.77.3.600
Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424
Zrida, J., Birdwell, J. D., & Cockett, J. R. B. (1987). State Estimation for Static Systems Defined By Stochastic Decision Trees. *Proceedings of the Annual Southeastern Symposium on System Theory*, *38*(7), 487–491.
Zwir, I., Arnedo, J., Mesa, A., del Val, C., de Erausquin, G. A., & Cloninger, C. R. (2023). Temperament & Character account for brain functional connectivity at rest: A diathesis-stress model of functional dysregulation in psychosis. *Molecular Psychiatry*, *28*(6), 2238–2253. https://doi.org/10.1038/s41380-023-02039-6

Supplementary

participant_id	gender	age
sub-01	m	30-35
sub-02	m	30-35
sub-03	f	20-25
sub-04	f	20-25
sub-05	m	25-30
sub-06	m	20-25
sub-09	m	30-35
sub-10	f	20-25
sub-14	f	30-35
sub-15	m	25-30
sub-16	m	35-40
sub-17	m	30-35
sub-18	m	30-35
sub-19	f	20-25
sub-20	f	25-30

Table S1. **Information on participant ID, gender and age range as provided in https://openneuro.org/datasets/ds000113/versions/1.3.0.**

<u>S2 Functional Brain Networks</u>

Autobiographical memory (AM)

A central domain involved in a core network linked to self-projection and scene construction is remembering personal events from one's own past (Spreng, Mar, & Kim, 2008). Within that core network, an autobiographical memory network was found within the medial and lateral temporal cortices, precuneus, posterior cingulate cortex, retrosplenial cortex, temporo-parietal junction, lateral prefrontal and occipital cortices and medial prefrontal cortex (Spreng et al., 2008).

Cognitive attention control (CogAC)

The CogAC network (Cieslik, Mueller, Eickhoff, Langner, & Eickhoff, 2013) is involved in higher level control processes and goal-oriented behaviour and consists of the anterior insula, inferior frontal gyrus, dorsolateral prefrontal cortex, dorsal premotor cortex, bilateral intraparietal sulcus and superior parietal lobe, right temporo-parietal junction, left inferior occipital gyrus, pre-supplementary motor area and anterior midcingulate cortex, as well as the right thalamus and right caudate nucleus.

Extended multiple demand network (eMDN)

Executive functions are fundamental to a variety of behaviours and recruit a core network of brain regions linked to a multiple demand system and additional more task-specific sub-networks. This eMDN consists of the bilateral inferior frontal gyrus, insula, supplementary motor area, intraparietal sulcus, middle frontal gyrus, dorsal pre-motor cortex, putamen, thalamus and the left inferior temporal gyrus (Camilleri et al., 2016).

Emotional scene and face processing (EmoSF)

FMRI studies on emotional processing often make use of visual, emotional face or scene stimuli (Sabatinelli et al., 2011). A meta-analysis on brain regions involved in the visual perception of emotional faces or scenes revealed consistent activation in the medial prefrontal cortex, bilateral inferior, middle and superior frontal gyrus, amygdala, parahippocampal gyrus, fusiform gyrus, medial prefrontal gyrus, orbitofrontal gyrus, lateral occipital cortex, thalamus, pulvinar, right middle temporal gyrus and right anterior cingulate cortex (Sabatinelli et al., 2011).

Empathy

Empathy is the adoption of another's emotional state (Singer & Lamm, 2009). Investigating its involvement in moral decision-making, a meta-analysis revealed an empathy network consisting of the bilateral dorsomedial prefrontal cortex, anterior insula, inferior frontal gyrus, supplementary motor area, cingulate cortex, temporo-parietal junction, right amygdala, right middle temporal gyrus, right posterior superior temporal sulcus, left anterior thalamus, right posterior thalamus, right hippocampus, midbrain and right pallidum (Bzdok et al., 2012).

Theory of mind (ToM)

The same meta-analysis additionally investigated theory of mind, the ability to contemplate another's thoughts, desires and behaviour (Premack and Woodruff, 1978; Frith and Frith, 2003). Brain regions involved in theory of mind were the ventromedial and dorsomedial prefrontal cortex, frontopolar cortex, precuneus, bilateral temporo-parietal junction, temporal pole, middle temporal gyrus, posterior superior temporal sulcus, inferior frontal gyrus and right visual area V5 (Bzdok et al., 2012).

Emotion regulation (ER)

Cognitive reappraisal, which refers to changing one's interpretation of affective stimuli, is an important strategy of emotion regulation (Buhle et al., 2014). A meta-analysis identified the following brain regions as consistently involved in the support or moderation of cognitive reappraisal: the bilateral inferior frontal gyrus, middle frontal gyrus, superior parietal lobe, amygdala, right medial frontal gyrus, left anterior cingulate gyrus, left anterior insula, left superior temporal gyrus and left middle temporal gyrus (Buhle et al., 2014).

Extended socio-affective default network (eSAD)

The default mode network (Gusnard and Raichle, 2001 -> check REF) is closely linked to socio-affective processing (Schilbach et al., 2012 -> check REF). Therefore, Amft et al. (2015) performed a conjunction analysis on regions involved both in the DMN and social or affective tasks, resulting in the eSAD. This network consists of the anterior cingulate cortex, bilateral amygdala and hippocampus, temporo-parietal junction, ventral basal ganglia, precuneus, subgenual cingulate cortex, ventromedial and dorsomedial prefrontal cortex and left middle temporal gyrus and sulcus (Amft et al., 2015).

Mirror neuron system (MNS)

Action observation and imitation tasks can give insight into the human mirror neuron system while avoiding invasive single-cell recordings used to study mirror neurons in non-human primates (Caspers, Zilles, Laird, & Eickhoff, 2016). A meta-analysis on action observation and imitation revealed an underlying network consisting of the bilateral inferior frontal gyrus, primary somatosensory cortex, lateral occipital lobe, right fusiform face and body area, left medial premotor cortex, left posterior middle temporal gyrus and right superior parietal lobe (Caspers et al., 2016).

Motor

To investigate a brain network linked to motor function, a meta-analysis on finger tapping tasks was performed by Witt et al. (2008). The brain regions consistently activated were the bilateral sensorimotor cortex, basal ganglia, anterior cerebellum, inferior parietal cortex, left ventral premotor cortex and supplementary motor area (Witt, Meyerand, & Laird, 2008).

Reward (Rew)

A meta-analysis on reward-related decision making revealed a network consisting of the bilateral insula, thalamus, brain stem, mid-orbitofrontal cortex, middle frontal gyrus, right nucleus accumbens, left pallidum, left dorsomedial prefrontal cortex, left medial orbitofrontal cortex, right amygdala, supplementary motor area, anterior and posterior cingulate cortex, left inferior parietal lobe, right angular gyrus, left frontal pole and left superior frontal gyrus (Liu, Hairston, Schrier, & Fan, 2011).

Semantic memory (SM)

Semantic memory entails the knowledge we gained from experience (Binder, Desai, Graves, & Conant, 2009). A meta-analysis revealed that the angular and supramarginal gyrus, middle temporal gyrus, posterior inferior temporal gyrus, mid-fusiform gyrus, parahippocampus, dorsomedial, ventromedial and orbital prefrontal cortex, superior, middle and inferior frontal gyrus, posterior cingulate gyrus and ventral precuneus were predominantly activated by studies employing semantic memory tasks (Binder et al., 2009).

Vigilant attention (VigAtt)

Vigilant attention describes the ability to maintain attention on repetitive and unengaging tasks for which not much cognitive effort is needed (Langner & Eickhoff, 2013). The network underlying this cognitive function consists of the anterior paracentral lobe, right medial posterior superior frontal gyrus, dorsal midcingulate cortex, bilateral inferior frontal junction, anterior insula, thalamus, right inferior frontal sulcus, left precentral gyrus, left inferior occipital gyrus, right temporo-parietal junction, right middle occipital gyrus, right inferior parietal lobe and cerebellum (Langner & Eickhoff, 2013).

Working memory (WM)

Working memory describes the ability to encode, maintain and retrieve information over a short period of time. Brain regions commonly activated by working memory tasks include the bilateral anterior insula, inferior frontal gyrus, caudal and rostral lateral prefrontal cortex, posterior superior frontal gyrus, thalamus, cerebellum, intraparietal sulcus, superior parietal lobe, left nucleus caudate and left globus pallidum (Rottschy et al., 2012).

S3 Peak coordinates in MNI space of networks

network	node number	x	y	z
AM	1	-1	-53	21
	2	-26	-28	-17
	3	-49	-61	31
	4	-2	51	-11
	5	-60	-9	-18
	6	-50	27	-12
	7	26	-33	-15

	8	-1	20	57
	9	55	-58	30
	10	-47	9	46
	11	-42	53	7
	12	26	-14	-23
	13	52	-5	-18
	14	-39	13	-41
	15	-38	-82	38
	16	-48	29	17
	17	52	31	-11
	18	-11	62	9
	19	4	-8	2
	20	-4	39	16
	21	-5	-34	36
	22	-29	16	51
	23	31	1	-26
CogAC	1	36	22	-4
	2	2	16	48
	3	48	12	30
	4	36	2	54
	5	48	30	24
	6	-38	-44	46
	7	-24	-66	48
	8	40	-46	46
	9	60	-44	24
	10	30	-62	52
	11	-44	10	30

	12	-34	20	-4
	13	-26	2	52
	14	6	-18	-2
	15	-40	-66	-10
	16	48	19	6
	17	8	29	30
	18	-45	27	30
	19	11	7	7
eMDN	1	-46	6	30
	2	50	12	28
	3	-32	20	2
	4	36	22	0
	5	-4	14	44
	6	6	18	46
	7	-32	-52	46
	8	32	-58	48
	9	44	36	20
	10	-28	-4	52
	11	-44	32	22
	12	32	0	52
	13	-20	6	4
	14	10	-12	8
	15	-46	-60	-10
	16	22	6	4
	17	-10	-16	6
EmoSF	1	4	47	7
	2	42	25	3

	3	-42	25	3
	4	48	17	29
	5	-42	13	27
	6	-2	8	59
	7	20	-4	-15
	8	-20	-6	-15
	9	-20	-33	-4
	10	14	-33	-7
	11	53	-50	4
	12	38	-55	-20
	13	-40	-55	-22
	14	38	-76	-16
	15	-40	-78	-21
	16	-4	52	31
	17	36	25	-3
	18	-38	25	-8
	19	2	19	25
	20	0	-15	10
	21	-2	-31	-7
	22	-28	-70	-14
	23	46	-68	-4
	24	-48	-72	-4
Empathy	1	2	56	18
	2	-8	54	34
	3	36	22	-8
	4	-30	20	4
	5	50	12	-8

	6	54	16	20
	7	50	30	4
	8	-44	24	-6
	9	-4	18	50
	10	-2	28	20
	11	-4	42	18
	12	-2	-32	28
	13	52	-58	22
	14	-56	-58	22
	15	22	-2	-16
	16	54	-8	-16
	17	52	-36	2
	18	-12	-4	12
	19	6	-32	2
	20	26	-26	-12
	21	2	-20	-12
	22	14	4	0
ER	1	48	24	9
	2	42	21	45
	3	9	30	39
	4	0	-9	63
	5	-3	24	30
	6	-33	3	54
	7	-36	21	-3
	8	-42	45	-6
	9	63	-51	39
	10	-42	-66	42

	11	-63	-51	-21
	12	-51	-39	3
	13	30	-3	-15
	14	-18	-3	-15
eSAD	1	0	38	10
	2	-24	-10	-20
	3	24	-8	-22
	4	-2	-52	26
	5	-2	32	-8
	6	-46	-66	18
	7	50	-60	18
	8	-2	52	14
	9	-6	10	-8
	10	6	10	-8
	11	-2	50	-10
	12	-54	-10	-20
MNS	1	-56	8	28
	2	-54	6	40
	3	58	16	10
	4	44	-54	-20
	5	-38	-40	50
	6	51	-36	50
	7	-1	16	2
	8	-54	-50	10
	9	-52	-70	6
	10	54	-64	4
	11	30	-62	63

Motor	1	-39	-21	54
	2	41	-16	57
	3	-3	-2	54
	4	-57	2	32
	5	-53	-24	21
	6	45	-38	48
	7	-23	-7	1
	8	25	-8	3
	9	-22	-52	26
	10	18	-54	-22
Rew	1	12	10	-6
	2	-10	8	-4
	3	36	20	-6
	4	-32	20	-4
	5	0	24	40
	6	0	54	-8
	7	24	-2	-16
	8	6	-14	8
	9	-6	-16	8
	10	0	8	48
	11	8	-18	-10
	12	-6	-18	-10
	13	2	44	20
	14	-24	2	52
	15	-38	-4	6
	16	24	40	-14
	17	-16	42	-14

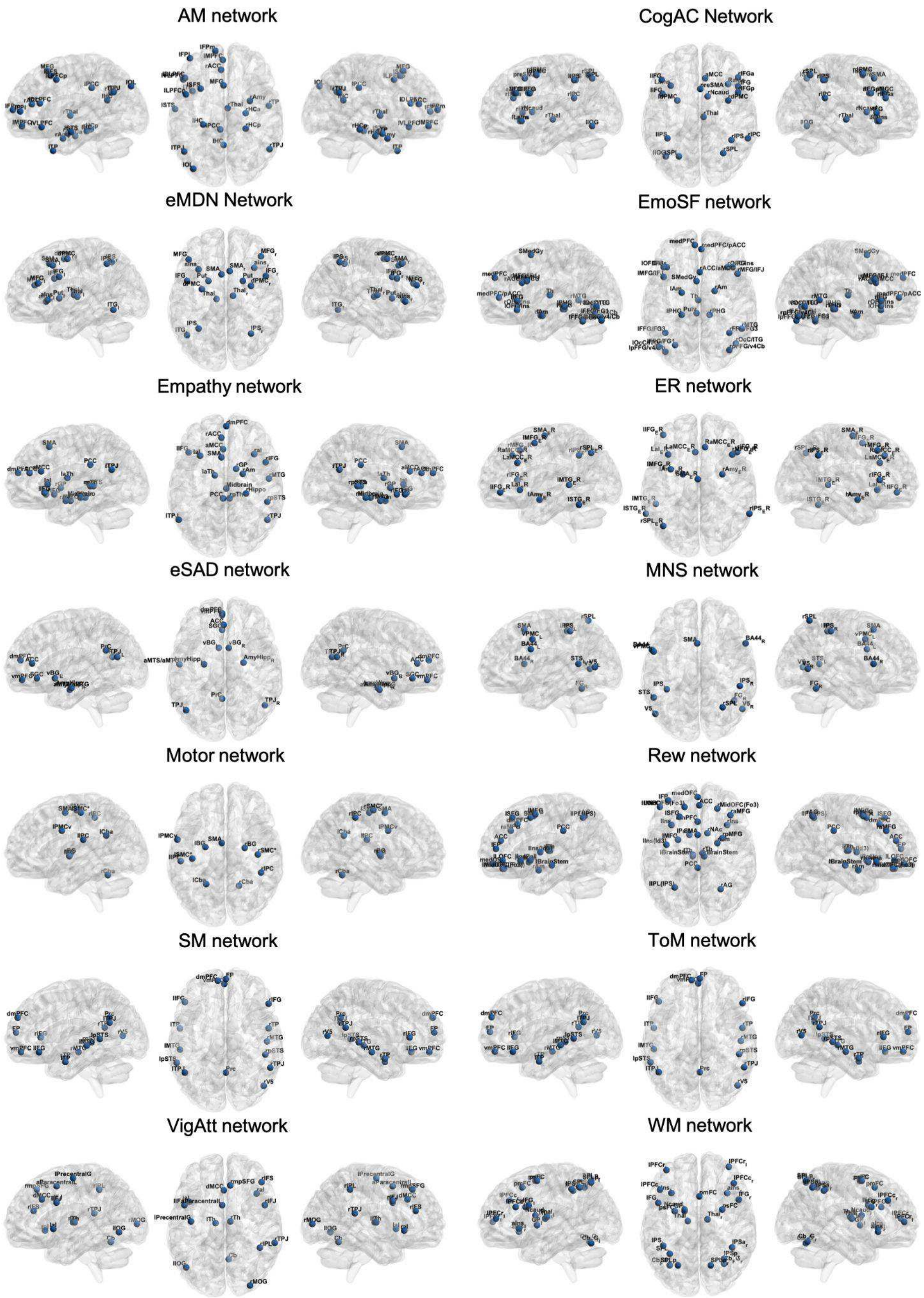
	18	40	32	32
	19	-28	-56	48
	20	28	-58	50
	21	0	-32	32
	22	-36	50	10
	23	-46	42	-4
	24	30	4	50
	25	-22	30	48
ToM	1	0	52	-12
	2	2	58	12
	3	-8	56	30
	4	2	-56	30
	5	56	-50	18
	6	-48	-56	24
	7	54	-2	-20
	8	-54	-2	-24
	9	52	-18	-12
	10	-54	-28	-4
	11	50	-34	0
	12	-58	-44	4
	13	54	28	6
	14	-48	30	-12
	15	48	-72	8
VigAtt	1	-2	8	50
	2	8	32	46
	3	0	26	34
	4	50	8	32

	5	40	22	-4
	6	46	36	20
	7	-40	-12	60
	8	-46	-68	-6
	9	-48	8	30
	10	62	-38	17
	11	8	-12	6
	12	32	-90	4
	13	-42	12	-2
	14	-10	-14	6
	15	6	-58	-18
	16	44	-44	46
WM	1	-32	22	-2
	2	-48	10	26
	3	-46	26	24
	4	-38	50	10
	5	36	22	-6
	6	50	14	24
	7	44	34	32
	8	38	54	6
	9	2	18	48
	10	-28	0	56
	11	30	2	56
	12	-42	-42	46
	13	-34	-52	48
	14	-24	-66	54
	15	42	-44	44

	16	32	-58	48
	17	16	-66	56
	18	-12	-12	12
	19	-16	2	14
	20	-16	0	2
	21	12	-10	10
	22	-34	-66	-20
	23	32	-64	-18
SM	1	-46	-69	28
	2	-50	-56	31
	3	-64	-44	-4
	4	-47	-24	-17
	5	-40	-12	-30
	6	-8	-57	17
	7	-20	36	44
	8	-53	27	-4
	9	54	-59	30
	10	43	-72	31
	11	-1	51	-7
	12	-5	56	24
	13	-31	-34	-16
	14	-8	29	-10
	15	-46	25	23
	16	64	-41	-2
	17	-43	-53	55
	18	-1	-18	40
	19	-2	-56	46

20	51	20	26
21	64	-38	32
22	-23	26	-16
23	-5	-39	40

S4 Nodes of all networks



Supp. Figure S4. **Visualization of networks nodes of all 14 meta-analytic networks.** For peak coordinates of each node, see S3.

Acknowledgements

This thesis really was a lot of work, and I could not have done it alone.

I want to thank first the people that gave me the opportunity to pursue a PhD in the first place: the medRSD, represented by Dr. Jessica Gätjens, who was readily available for answering questions, helping organize the PhD, and giving friendly and useful support.

Thank you, PD Dr. Julian Caspers, for agreeing to co-supervise me, your encouraging curiosity about my project and the kind watch you kept over my progress.

Thank you, Prof. Dr. Simon Eickhoff, for allowing me to work in your institutes, where collaboration, trust and curiosity are lived values that bring together many amazing colleagues, inspire to probe and stretch and overcome the limitations of our work and our minds, and invite deep and open discussions of our ideas.

Thank you, PD Dr. Suanne Weis, for your endless trust, patience, encouragement, empathy, and humour. You were available at all times - and that is no overstatement - and supported me through all obstacles, set-backs, lows and difficulties. More than that, you added some really good memories to those times! It was a pleasure to work with you and I appreciate our time together deeply.

Great thanks goes to my working group and colleagues: Dr. Patrick Friedrich, Dr. Xuan Li, Dr. Julia Camilleri, Lisa Wiersch, Gianna Kuhles, Jean Kröll, Natalie Schlothauer, and everyone else who accompanied me on this journey. Thank you for all the laughs, the discussions, the practical and emotional support, the few travels Corona couldn't take from us, the social gatherings and office talks that made work a warm place. Work was always fun with you (even when it was annoying)!

Thank you to all of my friends: Gesa and Juliana, for always letting us pick up where we left off, no matter how much time has passed. For the steady connection and many years shared with me. Conny and Mona, for inspiring me and showing me new perspectives, for inviting me in and always making me feel welcome. Anne, for all the times when just sitting together and knitting is enough, and all your interesting stories. Camellia, for more than I can put into words, all that we share, all that you give and let me give back.

Many, many thanks go to my family: Mama, Emilio, Jannik and Niño, Papa and Delia, my grandparents, aunt, uncles, and cousins. Thank you for your understanding of all the times I wanted to be there, but couldn't, and how much time and energy were bound to this project. Thank you for always being there for me, at any time and in any way I needed you to be. Knowing that you've got my back gives me relief and safety.

Last but not least I want to thank myself for pushing through, never giving up and being too stubborn to stop believing in myself, despite it all.