TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

PRINCIPLES OF SOCIAL PERCEPTION

Investigating Perceptual and Neural Mechanisms with Cinematic Stimuli

Severi Santavirta

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1873 | MEDICA – ODONTOLOGICA | TURKU 2025

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

PRINCIPLES OF SOCIAL PERCEPTION

Investigating Perceptual and Neural Mechanisms with Cinematic Stimuli

Severi Santavirta

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1873 | MEDICA – ODONTOLOGICA | TURKU 2025

University of Turku

Faculty of Medicine
Department of Clinical Medicine
Clinical Neurosciences
Doctoral Programme in Clinical Research
Turku PET Centre

Supervised by

Professor Lauri Nummenmaa, PhD
Turku PET Centre
University of Turku
Turku, Finland

DSc Enrico Glerean
Department of Neuroscience and Biomedical Engineering
Aalto University School of Science
Espoo, Finland

Reviewed by

Professor Angelika Lingnau, Dr.
Chair of Cognitive Neuroscience
University of Regensburg
Regensburg, Germany

Docent Mikko Peltola, PhD
Faculty of Social Sciences
Psychology
Tampere University
Tampere, Finland

Opponent

Professor Carolyn Parkinson, PhD
Department of Psychology
University of California, Los Angeles
Los Angeles, United States

ChatGPT (GPT-4o) was used to improve the language quality of this thesis.

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-952-02-0130-2 (PRINT)
ISBN 978-952-02-0131-9 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama, Turku, Finland 2025

UNIVERSITY OF TURKU
Faculty of Medicine
Clinical Neurosciences
Turku PET Centre
SEVERI SANTAVIRTA: Principles of Social Perception: Investigating Perceptual and Neural Mechanisms with Cinematic Stimuli
Doctoral Dissertation, 240 pp.
Doctoral Programme in Clinical Research (DPCR)
April 2025

ABSTRACT

Sociability is central to humans. Every day, people engage in complex social interactions, perceiving others and the dynamics of these interactions to interpret situations accurately and respond appropriately. Despite the importance of social perception, the fundamental principles governing how individuals perceive the surrounding social world remain largely unresolved.

This thesis investigates the principles of social perception. Three independent studies were carried out to explore the social perceptual cascade, beginning with visual perception, progressing through neural processing, and culminating in social perceptual inference. Study I investigated how people rapidly infer social situations. Study II mapped the functional organization of social perception in the human brain. Study III analyzed how external perceptual features guide visual attention during social scenes.

In Study I, altogether 2,254 participants evaluated the presence of 138 social features in 234 movie clips and 468 images rich in social contents. Dimension reduction analyses were conducted to establish the basic dimensions underlying social scene evaluations. Study II involved functional magnetic resonance imaging (fMRI) of 97 participants as they viewed 96 short movie clips depicting social scenes, aiming to map the brain's functional organization for social perception. Study III investigated the relationship between perceptual features of movie stimuli and eye-tracking parameters (pupil size, gaze orientation, and blinking behavior) across three movie-viewing eye-tracking studies (166 participants, 193 minutes of movie stimuli), revealing how visual attention is guided by perceptual features in social scenes.

The results indicate that visual attention is predominantly guided by simple external features, such as human faces and visual motion, while high-level emotional arousal modulates pupillary responses. Subsequently, occipitotemporal brain network is involved in processing social perceptual information, and social situations are ultimately evaluated along *eight basic dimensions of social perception*.

KEYWORDS: social perception, social neuroscience, fMRI, eye tracking

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Kliiniset neurotieteet
Valtakunnallinen PET-keskus
SEVERI SANTAVIRTA: Sosiaalisen havaitsemisen periaatteet: Havainto- ja aivomekanismien selvittäminen elokuvien avulla
Väitöskirja, 240 s.
Turun Kliininen Tohtoriohjelma (TKT)
Huhtikuu 2025

TIIVISTELMÄ

Sosiaalisuus on keskeinen osa ihmisten elämää. Päivittäin osallistumme monimutkaisiin sosiaalisiin vuorovaikutustilanteisiin, joissa havainnoimme toisiamme ja vuorovaikutusta tilanteiden tulkitsemiseksi. Vaikka sosiaalinen havaitsemisen on tärkeää jokapäiväisessä elämässä, perusperiaatteet siitä, miten ihmiset havaitsevat sosiaalista ympäristöä, ovat suurelta osin tuntemattomia.

Tässä väitöskirjassa tutkittiin sosiaalisen havaitsemisen periaatteita. Toteutimme kolme itsenäistä osatyötä koko sosiaalisen havaintoketjun tutkimiseksi. Sosiaalinen tiedonkäsittely alkaa aistihavainnoista, joista aivoissa muokataan sosiaalisia havaintoja. Ensimmäinen osatyö selvitti millaisia sosiaalisia havaintoja ihmiset tekevät sosiaalisissa tilanteissa. Toinen osatyö selvitti sosiaaliseen havaitsemiseen liittyviä aivomekanismeja. Kolmannessa osatyössä tutkimme, miten ulkoiset ärsykkeet ohjaavat ihmisen katsetta ja laajempaa visuaalista tarkkaavaisuutta sosiaalisissa tilanteissa.

Osatyössä I yhteensä 2254 tutkittavaa arvioi 138 sosiaalisen piirteen esiintymistä sosiaalisia tilanteita esittävissä 234 elokuvaleikkeessä ja 468 kuvassa. Moniulotteisen havaintoaineiston avulla selvitimme sosiaalisen havaitsemisen pääulottuvuudet, ja niiden yleistyvyyden riippumattomissa havaintoaineistoissa. Osatyössä II tutkimme 97 henkilön aivotoimintaa funktionaalisella magneettikuvantamisella (fMRI) sosiaalisia tilanteita esittävien elokuvaleikkeiden aikana, kartoittaaksemme sosiaaliseen havaitsemiseen liittyvät aivoverkostot. Osatyössä III tutkimme ihmisten silmänliikkeitä elokuvien katselemisen aikana kolmessa riippumattomassa aineistossa (166 tutkittavaa, 193 minuuttia elokuvia) tavoitteenamme selvittää, miten ulkoiset ärsykkeet ja havainnot ohjaavat visuaalista tarkkaavaisuutta sosiaalisissa tilanteissa.

Tulokset osoittavat, että pääosin yksinkertaiset havainnot, kuten ihmisten kasvojen tai liikkeen havaitseminen, ohjaavat visuaalista tarkkaavaisuutta sosiaalisissa tilanteissa. Pupillin kokoon vaikuttaa myös tilanteiden aiheuttamat tunnereaktiot. Aivoissa sosiaalisten havaintojen käsittelyyn liittyy laaja aivojen takaosien aivoverkosto, ja lopulta sosiaalisia tilanteita arvioidaan *kahdeksan perusulottuvuuden kautta*.

AVAINSANAT: sosiaalinen havaitseminen, sosiaalinen neurotiede, fMRI, silmänliiketutkimus

Table of Contents

Abbreviations

AI	Artificial intelligence
ASD	Autism spectrum disorder
BOLD	Blood-oxygen-level-dependent
DI	The dynamic interactive theory of person construal
eISC	Intersubject correlation of gaze positions
EEG	Electroencephalography
GLM	General linear model
fMRI	Functional magnetic resonance imaging
HC	Consensus hierarchical clustering analysis
HRF	Hemodynamic response function
ICC	Intra-class correlation coefficient
ISC	Intersubject correlation of neural responses
MEG	Magnetoencephalography
MRI	Magnetic resonance imaging
MVPA	Multivariate pattern analysis
OLS	Ordinary least squares
PC	Principal component
PCoA	Principal coordinate analysis
PET	Positron emission tomography
RF	Radio frequency
RMS	Root mean square
ROI	Region-of-interest
TR	Repetition time
t-SNE	T-distributed stochastic neighbor embedding

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Santavirta S, Malén T, Erdemli A, Nummenmaa L. A taxonomy for human social perception: Data-driven modeling with cinematic stimuli. *Journal of Personality and Social Psychology*, 2024; 127(6): 1146–1171.

II Santavirta S, Karjalainen T, Nazari-Farsani S, Hudson M, Putkinen V, Seppälä K, Sun L, Glerean E, Hirvonen J, Karlsson HK, Nummenmaa L. Functional organization of social perception networks in the human brain. *NeuroImage*, 2023; 272: 120025.

III Santavirta S, Paranko B, Seppälä K, Hyönä J, Nummenmaa L. Modeling human social vision with cinematic stimuli. *[Manuscript]*.

1 Introduction

In one of the most powerful scenes in cinema, Forrest and Jenny Gump share a conversation on her deathbed (Zemeckis, 1994). He tells her stories about the beautiful moments he has experienced during his adventures. “I wish I could have been there with you.”, she says dreamingly. “But you were.”, he replies promptly and assuredly, as she touches his hand. Later, Forrest is visiting her grave, visibly shaken and in tears, reflecting on their life after her death. In these brief moments the author perceives a profound emotional connecting between Forrest and Jenny, along with inferring the peaceful acceptance that death will tear them apart. During Forrest’s monologue at her grave, he expresses deep longing for her while also conveying that he has managed to move on and take care of their family without her. “If there’s anything you need, I won’t be far away.”, he concludes before leaving the grave.

This scene highlights the humans’ astonishing ability to immediately parse complex social signals. Sociality is what has enabled humans to build advanced societies (Tomasello, 2014). Every day, people engage in social situations where individuals have differing motives, objectives, and viewpoints. Successful social interaction, from cooperation to competition, requires accurate perception and interpretation of the situation, the people, and the behavior (Funder, 2006). This social perception is the first step that is required for successful social interaction (Molapour et al., 2021). The cognitive processing from pure sensory information to complex social information about other people must be swift for predicting how the situation will unfold. Robust evidence shows that human faces and bodies are rapidly detected and prioritized in natural scenes (Fletcher-Watson et al., 2008; Ro et al., 2007) highlighting their importance as fast mediators of social information.

The example scene above also elucidates how humans can perceive instantly multiple simultaneously occurring social features ranging from other people’s identities, intentions, hopes, and desires to their actions and subtle affective characteristics of social interaction. All this is possible despite the complexity, high dimensionality, and fast unfolding of social processes (Adolphs et al., 2016). Considering the brain's computational constraints, it is unlikely that people attend to every perceivable social feature instantly and independently (Freeman et al., 2012).

In other domains, people use heuristics to ease the computations and to allow fast judgments with incomplete information (Gigerenzer & Brighton, 2009). Similar "short cuts" could be used to filter the most important social information based on easily recognizable social cues (e.g., facial expressions) for swift and accurate interpretation of the social situation with minimal processing effort (Freeman & Ambady, 2011). It is thus more likely that people infer social situations by parsing perceptual information from a limited number of basic social perceptual dimensions, but the basic dimensionality of social perception is currently not understood.

The processing cascade for social perception begins with (audio-visual) sensory input. This purely physical information is then processed in the brain to infer complex social information about other people and their interaction. To understand social perception, we need to investigate the entire process. This includes the principles guiding visual information sampling in social situations, subsequent neural processing, and the resulting social inference. In this thesis, I report results from three independent studies, one for studying social vision, one for investigating functional organization of social perception in the brain, and one for establishing a taxonomy for social perception based on perceptual data to establish the principles of human social perception.

2 Review of the Literature

2.1 Principles of social perception

2.1.1 Theoretical framework for social perception

This thesis focuses on social perception unfolding in the timescale of seconds. Traditionally, social psychology has conceptualized social situations as a triad consisting of the person, the situation, and the behavior, considering these components as separate entities. However, in real life, persons, situation, and behavior are in constant interaction (Funder, 2006). In this thesis, all available information about others and their interactions is considered relevant for interpreting social situations, rather than distinguishing between different aspects of social perception.

The dynamic interactive theory of person construal (DI) provides a conceptual framework for understanding social perception (Freeman & Ambady, 2011). This theory describes social perception as a dynamic interaction between low-level sensory information and higher-order cognitive processes, such as prior experiences, motives, goals, and the current affective state. These elements influence each other bidirectionally, shaping how we perceive the surrounding social world. The DI framework helps to explain how social information is processed so quickly despite its complexity.

According to the DI theory, social perception is a hierarchical processing stream beginning with the detection of social cues (e.g., detecting an angry face), followed by social categorization (e.g., the person is angry) and resulting in general stereotyping (e.g., inferring that the person is possibly hostile). This heuristic enables humans to automatically link easily detectable social cues to more complex inferences about others and the broad social context. While the DI framework provides a robust conceptual basis, it does not specify which social cues are extracted in different social situations or the types of perceptual inferences people make rapidly based on these cues.

2.1.2 Existing models for social perception

Previous studies have explored the dimensionality of cognitive phenomena closely related to social perception, beginning with the early observation that semantic judgments of English words primarily vary along three main dimensions: valence, potency, and activity (Osgood & Suci, 1955). Relatedly, the circumplex model of affect maps emotions within a two-dimensional space defined by valence and arousal (Russell et al., 1989). Subsequent models have addressed how people and groups are stereotyped. The stereotype content model and the dual perspective model of agency and communion both describe a two-dimensional framework for understanding these stereotypes (Abele & Wojciszke, 2014; Fiske, 2018). The warmth/communion dimension reflects social traits related to others' intentions, such as whether they are inclined to contribute to the community, while the competence/agency dimension assesses the capacity to successfully pursue these goals. The ABC of stereotypes extends these models by categorizing social groups based on their agency/socioeconomic success and conservative–progressive beliefs, with communion emerging as a product of these two dimensions rather than as an independent factor (Koch et al., 2016). Additionally, prior studies have established taxonomies for mental state categorization (Thornton & Tamir, 2020), personality traits (Goldberg, 1990; Lee & Ashton, 2004; McCrae & Costa, 1987; Simms, 2007), psychological situations (Parrigon et al., 2017; Rauthmann et al., 2014) and action understandings (Thornton & Tamir, 2022). These studies mainly focus on the semantic similarities of words or concepts and imagined scenarios, rather than on the actual perception of dynamic social situations.

Substantial evidence of the dimensionality in social perception comes from face perception studies, where people evaluate standardized images of faces (Sutherland & Young, 2022). Valence and dominance emerged as the two primary perceptual dimensions for face evaluation (Jones et al., 2021; Oosterhof & Todorov, 2008) and they extend to body perception (Tzschaschel et al., 2022). Youthfulness/attractiveness has been proposed as a third evaluative dimension in addition to valence and dominance (Sutherland et al., 2013; Vernon et al., 2014). Moreover, femininity-masculinity is traditionally viewed as a single evaluative continuum representing sex characteristics (O'Toole et al., 1998). Femininity is often associated with youthfulness/attractiveness (O'Toole et al., 1998; Vernon et al., 2014) and masculinity with dominance (Oosterhof & Todorov, 2008; Sutherland et al., 2013), raising the question of whether sex characteristics are truly independent from other evaluative dimensions (but see (Lin et al., 2021)).

Many previous taxonomies are interrelated, suggesting that they describe partially overlapping processes (Horstmann et al., 2021; Lin & Thornton, 2023; Stolier et al., 2020; Wilkowski et al., 2020). For example, learning the situational mental state of others influences how people evaluate their enduring psychosocial

traits, or vice versa, indicating that mental state and trait inferences depend on each other (Lin & Thornton, 2023). Additionally, social trait inferences demonstrate high structural similarity across face impressions, familiar person knowledge, and group stereotypes, pointing to their conceptual convergence which may be learned through experience (Stolier et al., 2020). These findings support the need for an integrated approach, as social perception is at the core of social cognition and likely shares similarities with many of the currently established taxonomies.

2.1.3 Motivation for a unified taxonomy for social perception

The first objective of this thesis was to establish a unified taxonomy for social perception, integrating previous dimensions of social cognition into a cohesive framework. Such a taxonomy is essential for multiple reasons.

First, no prior taxonomy specifically focuses on the perception of dynamic social situations. Existing taxonomies emphasize conceptual similarities between different situations, semantic similarities in language, or the perception of static images. It remains unresolved whether the findings from these approaches can be generalized to the perception of real-world, dynamic social interactions.

Second, a unified taxonomy would provide improved tools for investigating altered human behavior by providing a more accurate understanding of the preceding social perception. Many psychiatric and neurological conditions are characterized by difficulties in social interaction (Kennedy & Adolphs, 2012). For example, face perception and inferring others' mental states are altered in autism spectrum disorders (ASD) (Dalton et al., 2005; Moran et al., 2011), and sensory processing issues have been linked with social difficulties in children with ASD (Kojovic et al., 2019). Additionally, individuals with depression exhibit biases towards perceiving unpleasant characteristics in others (Liu et al., 2012), These findings suggest that impaired social perception may be central to social difficulties in these disorders. A taxonomy of social perception for the general population would provide a valuable reference point for studying these alterations in neurological and psychiatric conditions.

Finally, seamless interaction between humans and artificial agents, such as intelligent robots, requires that these agents understand how humans perceive the social world. A comprehensive understanding of the dimensions of social perception could be used to improve the social capabilities of AI systems. Preliminary findings have demonstrated that AI models are already capable of perceiving some abstract social perceptual information (Malik & Isik, 2023; Santavirta, Wu, et al., 2024).

2.2 Brain basis of social perception

2.2.1 Approaches for studying social perception in the brain

Since social processes unfold rapidly, studying social perception in the living human brain requires methods with high temporal resolution. Most functional brain research has been conducted using functional magnetic resonance imaging (fMRI) due to its wide availability and high spatial resolution (Poldrack et al., 2011). However, other techniques, such as electroencephalography (EEG) (Jackson & Bolger, 2014) and magnetoencephalography (MEG) (Hansen et al., 2022), offer even higher temporal resolution making them suitable for investigating rapid social processing. Positron emission tomography (PET) allows studying the molecular systems in the living brain (Heurling et al., 2017), but its temporal resolution with standard methods is insufficient for capturing immediate brain responses. Nevertheless, PET can be valuable for investigating socioaffective states with longer-lasting effects on the brain (Karjalainen et al., 2017, 2018; Manninen et al., 2017). In the future, advances in functional PET imaging may enable studying the molecular basis of rapid social processes by achieving the necessary temporal resolution (Li et al., 2020). Intracranial recordings offer high spatial specificity and millisecond-level temporal resolution for studying the living brain, but they are limited to rare clinical cohorts in which invasive procedures are clinically justified (Mukamel & Fried, 2012). In addition to *in vivo* techniques, brain lesion studies provide valuable insights into the dysfunction of specific brain regions (Vaidya et al., 2019).

2.2.2 Mapping the social brain

Since the invention of fMRI in the early 1990s (Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1990), researchers have been investigating the brain mechanisms underlying social cognition. Early studies using static image stimuli demonstrated that the fusiform gyrus (FG) is involved in face perception (Haxby et al., 2000), while the lateral occipitotemporal cortex (LOTC) plays a role in body perception (Downing et al., 2001). When participants read social stories during scanning, activity in the temporoparietal junction (TPJ) was found to reflect the processing of others' mental states (Saxe & Kanwisher, 2003), but the TPJ might also serve other functions, such as processing social context and attention (Carter & Huettel, 2013).

Superior temporal sulcus (STS) has been consistently identified as a central hub for processing multiple aspects of social perception (Deen et al., 2015; Isik et al., 2017; Nummenmaa & Calder, 2009; Pelphrey et al., 2005; Puce et al., 1996). Language processing has been associated with a network that includes superior

temporal gyrus (STG, which includes the primary auditory cortex), STS (with Wernicke's area in left posterior STS), TPJ, angular gyrus, middle temporal gyrus (MTG), and inferior frontal gyrus (IFG, which includes Broca's area in the left IFG) (Price, 2012). Finally, medial frontal cortex (MFC) has been extensively studied in theory of mind tasks and self-representation (Amodio & Frith, 2006).

2.2.3 Advanced ecological validity with complex dynamic stimuli

Early studies used highly controlled study designs and simple stimuli, which lack the complexity of real social interaction raising questions about their generalizability outside the laboratory (Nastase et al., 2020; Sonkusare et al., 2019). A major limitation of extensively used static image stimuli is the lack of the temporal aspect of social perception. For example, the first comparisons between brain responses to static images and dynamic videos of faces revealed that the face-selective region in STS responded to dynamic faces, while face-related areas in LOTC and FG showed similar responses to both static and dynamic faces (Pitcher, Dilks, et al., 2011).

Movies provide a rich source of naturalistic social content that can be controlled and presented during neuroimaging. They evoke strong emotions within the limits that can be used for research purposes, making them attractive stimuli for studying social cognition. Viewing movies also synchronizes neural responses across participants (Hasson et al., 2010, 2004) indicating their ability to capture attention and induce mental states (Nummenmaa et al., 2018). Consequently, fMRI studies using movie stimuli have shown that the posterior STS specifically responds to dynamic social stimuli (Lahnakoski et al., 2012). Not surprisingly, naturalistic stimuli, such as movies or spoken narratives, have become increasingly common in affective neuroimaging (Saarimäki, 2021). Studies utilizing movies have demonstrated, for example, that perspective-taking can synchronize brain activity across participants (Lahnakoski et al., 2014), and that the degree of neural synchronization can predict how closely participants are connected with each other in their social network (Parkinson et al., 2018).

2.2.4 Representational spaces for cognitive processes

Most studies investigating the neural basis of social perception, including those utilizing dynamic stimuli, have focused on mapping brain networks related to isolated social features such as faces, bodies, biological motion, or differentiating between social and non-social stimuli. However, a statistical combination of responses to a few simple stimulus features may not be able to predict the social brain network working in complex natural situations (Felsen & Dan, 2005). This has

led to the proposal of using data-driven methods with minimal prior hypotheses for studying social perception (Adolphs et al., 2016).

Consequently, data-driven approaches have been developed to define neural representational spaces for cognitive processes. Typically, researchers first create a multi-dimensional stimulus model by defining an extensive set of stimulus features. Subsequently, the neural representational space for the broader domain is mapped using dimension reduction techniques applied either to the stimulus feature set or to the resulting neural response patterns. Using these methods, neural representational spaces have been identified for observed actions and objects (Huth et al., 2012; Tarhan & Konkle, 2020; Tucciarelli et al., 2019), language (Huth et al., 2016), and emotions (Koide-Majima et al., 2020; Lettieri et al., 2019), but similar representational spaces for social perception have not been previously investigated.

2.2.5 Spatial specificity of brain response patterns

Traditionally, studies utilize univariate modeling of brain activation, which means that each brain region or voxel is analyzed independently from others. This artificial separation of neuronal populations into discrete voxels fails to account for the natural network structure of neurons. As a result, univariate analyses are suboptimal in revealing the spatial specificity of brain activation patterns associated with experimental conditions. More specifically, univariate analyses fail to reliably determine whether two conditions with overlapping response patterns still exhibit spatially distinct response profiles.

In contrast, multivariate pattern analysis (MVPA) provides a means to investigate these detailed spatial brain activation patterns (Brooks et al., 2020; Tong & Pratte, 2012). Pattern recognition studies using MVPA have identified unique spatial activation signatures for various social features. For instance, faces (Haxby et al., 2001) and their racial groups exhibit distinct spatial activation patterns in FG (Brosch et al., 2013). Different facial expressions can also be distinguished in FG, but also in STS (Harry et al., 2013; Said et al., 2010; Wegrzyn et al., 2015). Additionally, neural patterns in LOTC and inferior parietal lobe enable decoding of goal-oriented motor actions at varying levels of abstraction, indicating that these regions process conceptual aspects of actions rather than just their low-level properties (Wurm & Lingnau, 2015). Notably, decoding of goal-oriented actions in LOTC is achievable regardless of whether the actions are perceived from a first-person or third-person perspective, further indicating that the activation patterns represent higher-level conceptual information (Oosterhof et al., 2012). However, the spatial specificity of neural activations for different perceptual social features remains largely unexplored.

2.3 Investigating human social vision with eye tracking

Abundance of visual information is available during social situations. Prior to social perceptual processing in the brain, the (visual) attentional systems sample and extract the most important information from the current situation. This sampling primarily happens by adjusting the gaze position, fixation frequency, blinking behavior, and by controlling the amount of light entering the retina through pupillary control. Through these operations, we quickly recognize objects, evaluate affective contents from scenes (Nummenmaa et al., 2010) and facial expressions (Calvo & Nummenmaa, 2008), and prioritize human faces as the primary source of social signals over other information (Morrisey et al., 2019).

2.3.1 Complexity of the human social vision

Complex spatiotemporal dynamics of social situations challenge the investigation of real-life visual attention. Different people, objects and features are present simultaneously and more abstract social events unfold in different yet overlapping temporal scales. This complicates the traditional eye-tracking paradigms that use highly controlled and typically static stimuli. Highly controlled and static experimental designs, however, abolish the real dynamics and richness of social interaction. Therefore, it is debated how the findings from static stimuli transfer to dynamic social environments (Williams & Castelhano, 2019), particularly as the visual system responds differently to dynamic and static stimuli (Dorr et al., 2010). Furthermore, studies usually focus on modeling a single eye-tracking parameter (e.g., pupillary response, gaze direction, blinking etc.) with only a few external features. Yet, a more comprehensive understanding of the visual system in dynamic situations would require parallel investigation of the different parameters of the visual system with rich stimulus models to establish whether they are uniquely or similarly influenced by external factors.

Generating rich stimulus models from dynamic stimuli requires data-driven approaches due to complex and overlapping temporospatial dynamics of the stimulus features. Previous research has shown that the pupil responds to luminance changes but is also indicative of emotional arousal (Bradley et al., 2008; Hess & Polt, 1960). Additionally, people recognize objects before they are able to evaluate the objects' affective properties (Nummenmaa et al., 2010). These exemplary findings indicate that the visual system can be influenced by features from different levels of the cognitive processing cascade. However, we do not yet know whether cognitively high-level social perceptual features (e.g., pleasantness of the situation) are mere end products inferred from the low- and mid-level perceptual features (e.g., luminance or faces). Alternatively, high-level social perception could influence the sampling of

visual information alongside low-level perceptual features. Thus, simultaneous investigation of the human visual system with low-level (e.g., luminance), mid-level (e.g., faces), and high-level perceptual features (e.g., perceived pleasantness) is required for more comprehensive understanding of the external drivers of visual attention.

2.3.2 Visual attention and gaze synchronization

Gaze position is the most studied variable in eye-tracking studies. Gaze patterns are remarkably consistent across individuals viewing the same dynamic stimulus (Dorr et al., 2010; Franchak et al., 2016; Smith & Mital, 2013), indicating that humans largely sample the same information in a time-locked fashion. This suggests that physical stimulus features are major modulators of gaze. Synchronization in gaze positions and concomitant information sampling could result in common understanding of the events. Gaze synchronization can be measured with (eye gaze) intersubject correlation (eISC), which can be as high as 0.4 – 0.6 during movie viewing (Hasson et al., 2008; Wang et al., 2012).

Previous research has established how gaze is directed by external or intrinsic features during scene perception using both top-down and bottom-up models (J. M. Henderson, 2003). Early bottom-up models can predict the gaze probabilities reasonably well by computing saliency maps from local color, intensity and orientation information (Itti & Koch, 2000), but these models tend to overestimate gaze probabilities on the object boundaries while people really gaze at the center of objects. Bottom-up models also fail to explain the strong preference for human faces and eyes in social scenes (Birmingham et al., 2008, 2009). Consequently, more emphasis on top-down models has been proposed since (Stoll et al., 2015), with the latest evidence suggesting that low-level visual saliency and object-information together yield the best performance (Nuthmann et al., 2020; Roth et al., 2023). Advances in deep learning models provide an alternative solution where gaze patterns can be predicted without feature engineering. They can, at least in theory, learn high-level object and semantic information (e.g., faces) in pure bottom-up fashion from the complex interactions between low-level features. Increasingly accurate models are currently being developed for gaze prediction for images (Cornia et al., 2018; Lou et al., 2022) and videos (Bellitto et al., 2021; Jain et al., 2021). Although these models can increase the prediction accuracy, indirect measures are needed to interpret how these models produce the predictions (Hayes & Henderson, 2021). Currently, the primary focus of this research has been to develop ever more accurate models, while advancing the understanding of the human visual system itself has been given lower importance.

2.3.3 Pupillary responses

Pupillary responses index perceptual and cognitive processing. The pupil controls the amount of light entering the retina but adjusts depending on internal states such as emotions and cognitive effort. Pupil dilates when perceiving pleasant and unpleasant stimuli (Babiker et al., 2013; Bradley et al., 2008; Hess & Polt, 1960; Kawai et al., 2013), as well as when presented in auditory format (Oliva & Anikin, 2018; Partala & Surakka, 2003), and when imagining emotion-evoking situations (R. R. Henderson et al., 2018). Additionally, pupil dilates as a function of cognitive load (Ayres et al., 2021; Hyönä et al., 1995; Kahneman & Beatty, 1966; van der Wel & van Steenbergen, 2018). Conversely, pupil constricts when observing attractive individuals or aesthetic visual objects such as natural scenes (Liao et al., 2021). Pupil constriction also indexes the novelty of a scene and indicates how well a scene is memorized (Naber et al., 2013). In fact, pupil is shown to constrict even during sudden changes of simple stimuli when luminance is held constant indicating a general response for rapidly changing visual input (Kimura et al., 2014). Adrenergic and cholinergic neurotransmitter systems engage during emotions and cognitive effort and these systems likely also mediate the luminance-independent pupillary responses (Joshi et al., 2016; Reimer et al., 2016). All in all, real-life pupillary response is a complex combination of the pupillary light reflex and effects reflecting different cognitive states, such as emotions (Cherng et al., 2020; Steinhauer et al., 2004), but it is not well established how the different factors simultaneously influence the pupillary responses during dynamic vision.

2.3.4 Blinking

Blinking can reveal higher-order cognitive processes. The main function of blinking is to clean and lubricate the eye surface, but sudden, intense tactile, visual, or auditory stimuli also modulate blinking behavior (Grillon & Baas, 2003). However, blinking is modulated by other cognitive and affective factors as well, and together with pupil size they are sensitive markers for many physiological measures and cognitive load (Ayres et al., 2021). A study investigating the blinking behavior of the contestants in the *Mastermind* TV-quiz showed nicely that blinks tend to occur during attention breakpoints during high stress (Wyly et al., 2024). Additionally, blink rates vary based on attention and the emotional contents of the stimuli (Maffei & Angrilli, 2019), spontaneous blinks synchronize between participants sharing the same stimuli (Nakano et al., 2009), and blink synchronization is stronger between participants that are interested in the stimuli (Nakano & Miyazaki, 2019). These findings indicate that blinking can reveal attentional disengagement which is supported by functional neuroimaging. Neural activity in the dorsal attention

network has been shown to decrease after blink onset, while activity in the default mode network simultaneously increases (Nakano et al., 2013).

2.4 Brigding the information gap in social perception research

The objective of this thesis is to establish the principles of social perception. Based on the literature review, several understudied areas and open questions remain: (1) Social perception research has mostly relied on static stimuli neglecting the temporal aspects of social perception. (2) While taxonomies have been established for other domains within social cognition, a detailed model for social perception in dynamic situations is currently lacking. (3) Neural representations for complex social features have not been established using naturalistic and dynamic stimuli. (4) An integrative analysis of how multiple perceptual features modulate human social vision remain unexplored and cannot be fully inferred from studies using simple features. (5) To understand the basic principles of social perception, research should investigate the entire social perceptual cascade from sensory input through neural processing to social perceptual inference.

This doctoral research project aims to bridge these gaps in the literature. Three independent studies were conducted, each focusing on a different part of the social perceptual cascade (Study I: social perceptual inference, Study II: neural processing, Study III: social vision). Each study utilizes movies as stimuli to study life-like dynamic social perception. Given the replicability crisis in psychological science (Open Science Collaboration, 2015), this research focuses on exploratory, data-driven approaches and replicability testing rather than narrowly focusing on theory-driven hypotheses. Specifically, data-driven models for social perceptual inference (Study I), neural representations for social perception (Study II), and social vision (Study III) were established by first collection multi-dimensional perceptual datasets from the movie stimuli and then refining the perceptual space using dimension reduction techniques. Final conclusions are drawn by integrating findings across these three studies of the social perceptual cascade, ultimately establishing the principles of social perception.

3 Aims

The aim of this thesis was to map functional, neural, and attentional mechanisms of social perception in uncontrolled dynamic settings. Three independent studies were conducted to investigate the social perceptual processing cascade from the audio-visual input to neural processing and the resulting perceptual inference. To this end, we used movies as naturalistic social stimuli and collected multimodal datasets containing social perceptual evaluations, functional neuroimaging, and eye-tracking. Furthermore, novel analytical methods were developed for each of these datasets and study questions.

The objectives of the specific studies were:

I. To establish a low-dimensional perceptual taxonomy for social perception.

II. To investigate the neural organization of social perception.

III. To establish how the external stimulus features guide the visual system during dynamic social scenes.

4 Materials and Methods

4.1 General methodology

4.1.1 Design and stimuli for perceptual study (Study I)

Study I examined how people perceive social information in complex social situations. A total of 1,140 participants were recruited to watch short, unrelated movie clips rich in social content (N = 234). The clips were selected primarily from Hollywood movies. Some of the same movie clips from this stimulation set were used in Studies II and III. The average duration of the movie clips was 10.5 s (range: 4.1 - 27.9 s) with a total duration of 41 min. The stimulus set was an extension of a previously validated set of socioemotional movies used in several previous neuroimaging studies (Karjalainen et al., 2017, 2018; Lahnakoski et al., 2012; Nummenmaa et al., 2021). The participants were instructed to evaluate the presence of 138 pre-defined social perceptual features from the movie clips to get a detailed description of their social content. The ratings were collected on a continuous visual scale to allow for detailed investigation of the perceptual rating distributions. The participants were asked to evaluate the magnitude of the presence of a given social feature between abstract endpoints “absent” and “a lot”. To investigate the generalizability of the results across dynamic and static perception, another set of participants (N = 1,109) was recruited to evaluate the same social features from 468 images that were captured from the primary stimulus movie clips (two images were captured from each clip). Generalization across different movie stimuli was tested using a retrospective dataset where five participants evaluated the presence of the same 78 social features when watching a full-length (70 min 14 s) Finnish historical movie (Louhimies, 2008) in short clips. The perceptual ratings were collected with online experiment platform Gorilla (Gorilla, 2024). The Ethics Committee of the Faculty of Social Sciences, University of Turku, waived the study from ethical review due to its minimal impact on human participants.

4.1.2 Design and stimuli for fMRI (Study II)

Study II investigated the neural processing of social perceptual information. A total of 104 participants were recruited to participate in fMRI brain imaging at Turku PET Centre. During the fMRI scan, the participants watched 96 short movie clips that were rich in social content. These stimuli were part of the previously validated neuroimaging stimulation set, also used in Study I. The ethics board of the Hospital District of Southwest Finland approved the protocol (ethical approval Dnro: 46/1801/2017), and the experiment was conducted in accordance with the Declaration of Helsinki.

Traditional fMRI studies use blocked or event-related designs, which assume that a condition is present or absent either for a prolonged or a transient period. Such designs do not fit well with an uncontrolled movie stimulus, in which stimulus features are present at varying intensities, and events occur simultaneously but with different temporal scales. Thus, a parametric stimulation model, in which the predictors dynamically reflect intensity changes in the stimulus features, is often used with movie stimulation (Hudson et al., 2020; Karjalainen et al., 2018). To allow parametric modeling of the functional brain imaging data, the movie clips were rated for 112 social perceptual features at four-second temporal resolution by five independent annotators, similar to those in Study I.

4.1.2.1 Functional magnetic resonance imaging

In magnetic resonance imaging (MRI), the contrast is based on protons' interaction with an external magnetic field. A thorough explanation of the process is beyond of the scope of this thesis and can be found elsewhere (McRobbie & Graves, 2007). Briefly, elementary particles and composite particles, such as protons, have a property called *spin,* which gives the particles a magnetic momentum. In the absence of external magnetic field, the magnetic moments of different protons can have any orientation. However, in an external magnetic field they (1) align with the external magnetic field and (2) precess around the direction of the field with an angular frequency, called the Larmor frequency, which is directly proportional to the magnitude of the external magnetic field (McRobbie & Graves, 2007).

In MRI scanning, the axis of precession is briefly altered by modifying the external magnetic field locally with radio frequency (RF) pulses. As a result of an RF pulse, the precessing axes of protons turn in relation to the static magnetic field. After the pulse, the precessing axes realign again with the static magnetic field in a process called *relaxation*. During relaxation, the protons return to a lower energy state emitting energy as a signal detectable by the MRI scanner. Tissues have different relaxation properties, most importantly differing relaxation durations,

which enable tissues to be differentiated in MR images as color contrasts (McRobbie & Graves, 2007).

Functional MR imaging (fMRI) quantifies dynamic changes in brain metabolism through changes in the blood-oxygen-level-dependent (BOLD) contrast (Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1990). The dynamic changes in BOLD contrast are based on the fact that the magnetic properties of hemoglobin are dependent on the amount of oxygen bound to it (Pauling & Coryell, 1936). Hemoglobin oxygenation in arterial blood depends on the blood flow and oxygen consumption, among other factors. Typically, an increase in the BOLD signal is observed after stimulation, which is then inferred to measure a reactive increase in blood flow, whereas the initial oxygen consumption is rarely observed in the BOLD signal (Hillman, 2014). In functional brain imaging, the BOLD signal is used as an indirect measure of neuronal activity, although the measure is a net effect including metabolic changes, such as respiratory and cardiac changes, that may not directly relate to specific neuronal activity (Keilholz et al., 2017).

Stimulus-evoked BOLD responses are relatively slow. Studies with simple and brief stimuli, such as single flashes of light, have established that the hemodynamic response takes a specific shape, which is typically modeled with the canonical hemodynamic response function (HRF) (Lindquist et al., 2009). The BOLD signal peaks (indicating a maximal increase in blood flow) approximately five seconds after a stimulus and then decreases to baseline or even briefly below it within the next twenty seconds completing the response. Thus, stimulation models need to be transformed to match the expected hemodynamic response using convolution prior to statistical modeling (Poldrack et al., 2011). In addition to the dynamics of the BOLD signal itself, the repetition time (TR) of the MRI scanner for BOLD acquisition restricts how dynamically the hemodynamic changes can be measured. TR is the “frame rate” of the MRI scanner, specifying how long it takes to collect one full volume, which is typically between two and three seconds.

4.1.2.2 FMRI acquisition and preprocessing

MR imaging was conducted at the Turku PET Centre. The functional MRI data were acquired using a Philips Ingenuity TF PET/MR 3-T whole-body scanner. High-resolution structural images were obtained with a T1-weighted (T1w) sequence (1.0 mm^3 resolution, TR 9.8 ms, TE 4.6 ms, flip angle 7°, 250 mm FOV, 256 × 256 reconstruction matrix). A total of 467 functional volumes were acquired for the experiment with a $T2^*$-weighted echo-planar imaging sequence sensitive to the BOLD signal contrast (TR 2,600 ms, TE 30 ms, 75° flip angle, 240 mm FOV, 80 × 80 reconstruction matrix, 62.5 kHz bandwidth, 3.0 mm slice thickness, 45 interleaved axial slices acquired in ascending order without gaps).

The acquired MR data were preprocessed using the standardized fMRIPrep preprocessing pipeline (Esteban et al., 2019). The following preprocessing was performed on the anatomical T1w image: correction for intensity non-uniformity, skull-stripping, brain surface reconstruction, spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009), and brain tissue segmentation. The following preprocessing was performed on the fMRI data: coregistration to the T1w reference and spatial normalization to the MNI152NL 2009c Asym template, slice-time correction, motion correction, and spatial smoothing with a 6 mm Gaussian kernel, followed by the non-aggressive automatic removal of motion artefacts (ICA-AROMA) (Pruim et al., 2015). The data were then detrended using a 240-s Savitzky–Golay filtering to remove scanner drift (Cukur et al., 2013) and demeaned to make the regression coefficients comparable across participants (G. Chen et al., 2017).

4.1.3 Design and stimuli for eye tracking (Study III)

Study III investigated the influences of perceptual features on gaze control during dynamic social perception. Three eye-tracking experiments were carried out, enabling generalizability testing across independent datasets. In Experiment 1, 110 participants watched 68 short movie clips with rich social content (14 min 26 s). These stimuli were part of the previously validated neuroimaging stimulation set, also used in Studies I and II. In Experiment 2, 28 participants watched a full-length (70 min 14 s) feature film (Louhimies, 2008), and other 28 participants watched a full-length (109 min 3 s) horror movie (Wan, 2016) in Experiment 3. Eye movements and other eye-tracking parameters were dynamically measured during the experiments with an eye-tracker. The Ethics Committee of the Faculty of Social Sciences, University of Turku, waived the study from ethical review due to its minimal impact on human participants.

4.1.3.1 Eye tracking

Optical eye tracking is commonly used for non-invasive recording of eye movements, pupil size changes, and blinks. Typically, the stimulus is presented on a computer screen while head movement is minimized with a head mount. A camera measuring visible or infrared frequencies records the reflections from the cornea to track the eye and measure pupil size (Klaib et al., 2021). Before and during the experiments, the eye-tracker is calibrated by instructing the participant to gaze at certain points on the presentation monitor so that the measured gaze positions can be transformed into screen coordinates. The recorded data are then processed using eye-tracking algorithms to divide the measured eye movements into fixations (moments

of stationary gaze position), saccades (moments when the focus shifts to another location), and blinks among other parameters of interest. Eye tracking is a useful tool for studying social perception and cognition because it has a high temporal resolution, up to 2,000 Hz.

In Experiment 1, the eye-tracking data were collected with an SR EyeLink 1000 Plus (SR Research, Ontario, Canada) eye tracker with the following setup: v5.15 Jan 24 2018, Eyes: Right, File filtering level: Extra, Pupil tracking algorithm: Centroid. The eye tracker was calibrated and validated using a five-point calibration, and a one-point validation was repeated before the experiment (validation error < 1°). Validation was repeated three times during the experiment. In Experiments 2 and 3, the eye-tracking data were collected with an EyeLink 1000 (SR Research, Ontario, Canada) eye tracker with the following setup: v4.594 Jul 6 2012, Eyes: Right, File filtering level: Extra, Pupil tracking algorithm: Ellipse. The full-movie stimuli were presented in ~3-4-minute-long segments, and the above-described calibration was performed between each segment.

Fixation and saccade reports were generated with EyeLink DataViewer 4.1.1 software (SR Research, 2025). Fixations shorter than 80 ms were considered unreliable, and the previous reliable fixation was extrapolated to continue until the next reliable fixation to create a continuous time series of fixation information.

4.2 Participants

All participants gave an informed consent prior to participation in the reported studies.

4.2.1 Perceptual evaluators (Study I)

For the primary movie clip experiment, English-speaking adults were recruited through the online platform Prolific (Prolific, 2025) until ten ratings were collected for each movie clip and evaluated social feature. A total of 1,140 fluent English-speaking adults completed the experiment. The final sample, after excluding 44 participants based on data quality control, included 1,096 participants from 60 nationalities and various ethnicities. Of these, 515 participants were females (47%), and the median age of the participants was 28 years (range 18 - 78 years).

A similar protocol and target sample size were selected when recruiting participants for evaluating social features in images for the generalizability analysis across dynamic and static stimuli. A total of 1,109 fluent English-speaking adults completed the experiment. The final sample, after excluding 15 participants based on data quality control, included 1,094 participants from 56 nationalities and various ethnicities. Of these, 448 participants were females (41%), and the median age of the

participants was 32 years (range 18–77 years). The final dataset contained ten ratings per image and social feature.

To allow generalizability analyses across different movie stimuli, retrospective social feature evaluations for the full-length movie were used. This dataset included evaluations from five Finnish participants.

4.2.2 Neuroimaging participants (Study II)

A total of 104 participants took part in the fMRI study. This sample size was considered sufficient for one-group analyses based on information from a previous simulation study about fMRI replicability with different sample sizes (Cremers et al., 2017). The study-specific exclusion criteria included a history of neurological or psychiatric disorders, alcohol or substance abuse, BMI under 20 or over 30, and the current use of medication affecting the central nervous system. Two participants were excluded due to a gradient coil malfunction and two others because of anatomical abnormalities in structural MRI. Additionally, three participants were excluded based on visible motion artifacts in the preprocessed fMRI data. The final sample included 97 participants (50 females, mean age 31 years, range 20 - 57 years).

4.2.3 Eye-tracking participants (Study III)

A total of 166 volunteers participated in one of three independent experiments (Exp. 1: 110, Exp. 2: 28, Exp. 3: 28), and 15 participants were excluded based on quantitative data quality control. The final sample included 151 participants (Exp. 1: total sample 106, 66 females, mean age 27.1, range 19 - 74; Exp. 2: total sample 21, 19 females, mean age 23.6, range 19 - 38; Exp. 3: total sample 24, 19 females, mean age 27.0, range 19 - 57).

4.3 Social perceptual features

All studies used specific sets of dynamically evaluated social perceptual features from the stimulus movies. Prior to this doctoral research, there was no clear understanding about which important features people perceive in dynamic social situations. The feature set for annotation should be comprehensive enough to capture the underlying dimensionality of social perception while being limited enough for data collection purposes. Some studies leave the feature selection for the participants (Koch et al., 2016; Nicolas et al., 2022; Osgood & Suci, 1955), but letting participants freely describe what they perceive could bias the research towards conscious reasoning, overlooking unconscious but important processing of social

features. Therefore, we decided to define the feature set based on previous taxonomies within social cognition.

First, we selected broad categories that would cover the perception of people, their actions, and qualities of social interaction. These broad categories were a person's traits, a person's physical characteristics, a person's internal situational states, somatic functions, sensory states, qualities of the social interactions, communicative signals, and a person's movement. Second, we searched through previous narrow taxonomies of social cognition (Abele & Wojciszke, 2014; Fiske, 2018; Goldberg, 1990; Lee & Ashton, 2004; Parrigon et al., 2017; Rauthmann et al., 2014; Russell et al., 1989; Schwartz et al., 2012; Wilkowski et al., 2020) to find several candidate features for the above-mentioned broad categories. The feature selection was also guided by previous work in social perception from a neuroscience perspective (Hudson et al., 2020; Karjalainen et al., 2017; Lahnakoski et al., 2012; Manninen et al., 2017; Nummenmaa et al., 2011).

A set of 112 candidate social perceptual features from the broad categories was initially defined for Study II, which focused on the brain basis of social perception. This feature set was further refined with additional theory-based features for Study I, which focused on mapping the low-dimensional structure of social perception. A more detailed explanation of the feature selection can be found in the original publication I (Page 3, section: "Evaluated Social Features"). Based on the evolving understanding of the social perceptual structure from Studies I and II, eight social features (pleasant feelings, unpleasant feelings, arousal, pain, talking, body movement, feeding, and playfulness) were selected for Study III to investigate the association between social perception and the human visual system. For Study I, the perceptual features were evaluated separately for each unrelated movie clip, resulting in an approximately ten-second temporal resolution, but to allow parametric modeling of fMRI and eye-tracking data, the evaluations were collected in a four-second temporal resolution in Studies II and III.

4.4 Statistical analyses

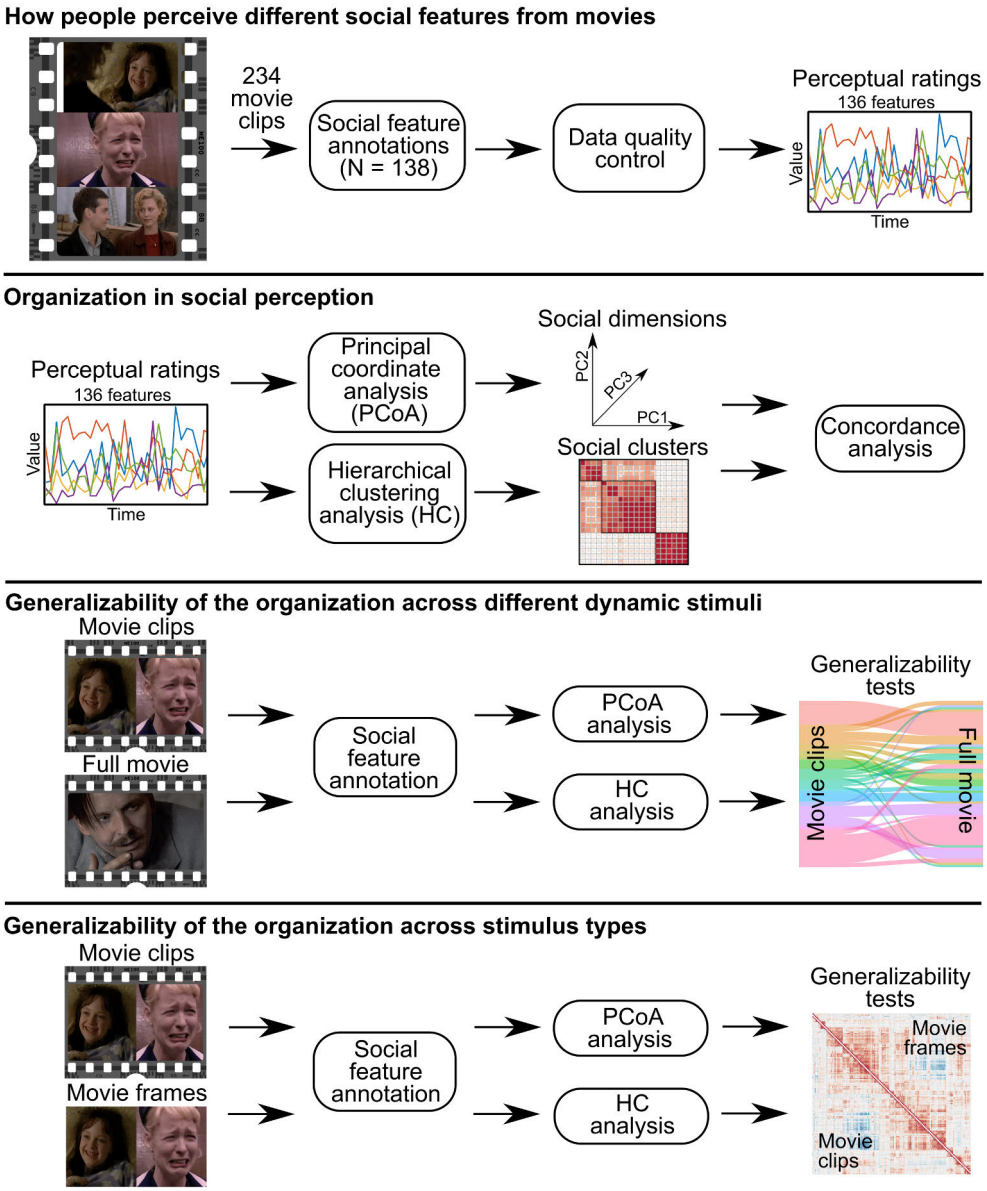

Figure 1. Analytical pipeline for Study I. First, an investigation into how people perceive, and rate individual social features was carried out. Second, a low-dimensional taxonomy for social perception was defined using principal coordinate analysis and hierarchical clustering. Third, the generalizability of the low-dimensional organization for social perception was tested across dynamic (movie) and static (image) stimuli, and across different movie stimuli. Modified from the original publication. Copyright © 2024 by American Psychological Association. Adapted with permission (Santavirta, Malén, et al., 2024).

4.4.1 Analyses of the perceptual ratings (Study I)

Figure 1 shows the analytical pipeline for Study 1.

4.4.1.1 Principal coordinate analysis

It is unlikely that the participants perceive each evaluated social feature independently of all other features. To investigate the low-dimensional space for social perception, we used principal coordinate analysis (PCoA) on the Pearson correlation matrix of social feature ratings to decompose the correlation structure into orthogonal principal components (PC) (Gower, 1966). PCoA is based on the eigenvalue decomposition of a symmetric diagonal matrix, where the N x N matrix is decomposed into N PCs with one scalar eigenvalue and one eigenvector of size N x 1 for each component. The sum of the eigenvalues represents the total variance explained by all components. Consequently, the variance explained can be calculated for each PC by dividing its eigenvalue by the sum of all eigenvalues. The eigenvector reflects the direction of the PC axis in relation to the original variables of the matrix. When a correlation matrix of social feature ratings is used as input for PCoA, the corresponding eigenvector value indicates the loading or "importance" of a given social feature for the given PC. Thus, the social perceptual information that the component conveys can be inferred by investigating the eigenvector values. The social perceptual label for each identified PC was formed by reaching a consensus among the original authors, local researchers (N=10), the general population (N=92), and ChatGPT 3.5 (OpenAI, 2024). A full explanation of the labeling process can be found from the original publication I (Page 9, section: "Naming the Identified Social Dimensions and Clusters").

After defining N orthogonal principal components with PCoA, it is necessary to identify how many PCs explain more variation in the data than would be expected by chance, since most PCs likely model just random noise in the data. If a given PC explains more variance than expected by chance, this will indicate that it describes some real social perceptual structure in the current data. Thus, a permutation test was implemented to generate null distributions for eigenvalues and eigenvectors. The columns of the social feature dataset (evaluations x features) were independently shuffled, and PCoA was conducted on this random data, repeating the process 1,000,000 times to produce null distributions. Finally, the true eigenvalues of the PCs and the eigenvector values of each PC were ranked within their corresponding null distributions to assess their statistical significance (exact p-value).

4.4.1.2 Consensus hierarchical clustering analysis

PCoA forces the social perceptual dimensions to be strictly orthogonal. However, finer-grained social semantic categories could emerge as specific combinations of orthogonal dimensions. To investigate this, we used an additional dimension-reduction method, consensus hierarchical clustering analysis (HC) (Chiu & Talhouk, 2018; Murtagh & Contreras, 2012) to generate social feature clusters that are not strictly orthogonal. A consensus approach was selected to achieve a stable clustering solution within subsets of the data and across different numbers of clusters. The final clustering solution was based on the consensus over 1,000 subsets of the data (randomly selected 80% of the rating data) and over different numbers of clusters (from 5 to 45). The final clustering solution was obtained by hierarchically ordering the resulting consensus matrix.

4.4.1.3 Concordance analysis

To further investigate how the fine-grained social semantic information contained in the social feature clusters arises from the basic evaluative dimensions (PCs from PCoA), we investigated whether HC-based social feature clusters can be explained as combinations of information from the identified PCoA dimensions. T-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten & Hinton, 2008) was used to map the overall association between the PCoA derived social dimensions and the HC based social clustering structure. The social features were mapped to a 2D projection based on their loadings for statistically significant PCs with t-SNE, and the HC cluster membership was plotted as a color-coded representation in the same plot. If the features form separable clusters in this t-SNE space, this will indicate that the PCoA and HC solutions for social perception are structurally similar. To study the associations between PCs and HC clusters in more detail, we also estimated the cluster-level PC loadings by averaging the feature-specific loadings over all features within each cluster.

4.4.1.4 Generalizability analyses

The generalizability of the PCoA and HC structures for social perception was tested with image and movie datasets where participants viewed images with social content or a full-length movie in short intervals instead of unrelated movie clips. PCoA and HC were conducted independently for these validation datasets prior to generalizability testing. To assess the similarity of PCoA components, we identified how many components were statistically significant in the validation datasets similarly to the main analysis (null distribution generation with 1,000,000 permutations) and correlated the feature loadings of the significant components

between the datasets. The structural similarity of the HC analysis was assessed with a non-parametric Mantel test with 1,000,000 permutations (Mantel, 1967) between the correlation and consensus matrices of the primary and validation datasets.

4.4.2 Neuroimaging data analyses (Study II)

Figure 2 shows the overview of Study II.

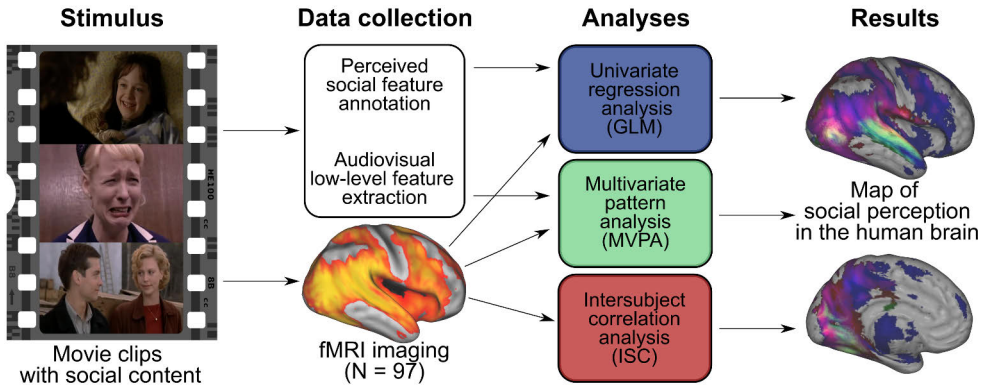

Figure 2. Overview of Study II. The neural responses to movie stimuli were modeled with annotated social perceptual features using univariate regression analysis to establish brain networks sensitive to social features. Additionally, the social perceptual context was predicted based on brain activation patterns using multivariate pattern analysis to reveal whether spatial brain activation patterns differentiate the social perceptual processing within the brain. The response patterns for social perception were compared to the neural synchronization patterns measured across participants with intersubject correlation analysis. Reprinted from the original publication (Santavirta et al., 2023).

4.4.2.1 Perceptual models for the fMRI data

112 social perceptual features were evaluated dynamically from the stimulus movie clips with the goal of developing a data-driven, low-dimensional model for predicting fMRI responses. Focusing on the most consistently evaluated social features across participants was justified, as the evaluators themselves did not participate in fMRI scanning. Based on the intra-class correlation coefficient (ICC; a two-way random model with absolute agreement), 45 social features were perceived with sufficient consistency (ICC > 0.5), and were also present in at least five rating points in the current stimulus set. Hierarchical clustering was used to further identify six meaningful clusters within these 45 reliable social features (Antisocial behavior, Sexual & affiliative behavior, Play, Feeding, Communication, and Body movement) while seven original features were not assigned to any cluster (Male, Female, Crying, Using an object, Running, Walking, and Searching). These

13 predictors (clusters + independent features) formed the social perceptual model. Cluster predictors were calculated as the mean of all feature ratings within the cluster. Clustering ensured that the social predictors are meaningful and relatively uncorrelated with other predictors allowing multiple regression. More detailed explanation of the clustering analysis can be found from the original publication II (Page 4, section: "2.7. Dimension reduction of the social perceptual space").

To control for confounding effects, we defined a low-level model that contained the first eight principal components computed from a set of 14 different audio-visual features that were extracted from the movie clips. The extracted low-level features included six visual features (luminance, first derivative of luminance, optic flow, differential energy, and spatial energy with two different frequency filters) and eight auditory features (RMS energy, first derivative of RMS energy, zero crossing, spectral centroid, spectral entropy, high frequency energy, and roughness). The low-level model was extended with mean signals from the cerebrospinal fluid and white matter and with a binary "non-social" regressor indicating time points when no people were present in the stimuli. The regressors in both models were convolved with a canonical HRF prior to the statistical modeling of fMRI data.

4.4.2.2 Cross-validated Ridge regression

Clustering does not yield strictly uncorrelated social features, and ordinary least squares regression (OLS) may overfit the model, reducing the generalizability of the results. Hence, Ridge regression was selected for fitting the social perceptual model to the BOLD signals (Hoerl & Kennard, 1970). Ridge regression introduces a penalty term λ that shrinks the regression coefficients towards zero. Consequently, the shrinkage of coefficients may yield more accurate out-of-sample predictions by reducing overfitting to the training data. The λ parameter was optimized with leave-one-participant-out cross-validation by minimizing the prediction error of the left-out participant's BOLD data.

A summary-statistics approach for mixed-effects modeling was used in fMRI analysis. The participants were treated as a random effect by first modeling each participant's fMRI data separately (first-level analysis) and then subjecting the parametric maps to a one sample t-test at the population level (second-level analysis) (Poldrack et al., 2011). A region-of-interest (ROI) analysis was used to summarize the results in anatomically segmented bilateral regions using the AAL2 atlas (Rolls et al., 2015).

To conservatively control for low-level confounds, we first fitted the low-level model to the preprocessed fMRI data and then modeled the residuals of this analysis with the social model, still including the low-level predictors as covariates. Adding the low-level features to the second model was motivated by the possibility of

interaction effects between confounds and social features, as well as potential correlations between social features and the confounds. As a *post hoc* analysis, we fitted the social and low-level models independently to the BOLD signals and compared where in the brain the social model predicted the BOLD signal better than the low-level model based on adjusted R^2.

4.4.2.3 Multivariate pattern analysis

Regression analysis can be used to identify the brain areas where the BOLD signal is associated with specific social features. If the activation maps for two features overlap, there are two competing interpretations about the function of the overlapping region: 1) The regional activity reflects a cognitive process (e.g., attention or working memory) that is similarly activated by perceiving different social features, or 2) the region is involved in the processing of specific social information. Spatial specificity of the neural activation patterns within such a region would favor the processing of specific social information while failure to identify spatially differing activation patterns would suggest a shared cognitive process activated by different social features. Multivariate pattern analysis (MVPA) (Hanke et al., 2009) was thus conducted to find evidence for specific social processing by investigating the spatial specificity of the neural activation patterns evoked by different social features. A successful prediction of the social context based on the neural activation patterns for different social contexts would indicate that the regional brain activation patterns are overlapping but spatially different supporting the interpretation of specific social processing.

For classification, we identified time periods of similar social context, which we call events. This was achieved by first labeling each fMRI time point with the main social context by selecting the social features with the highest Z-score among all 11 social features. Next, fMRI data were divided into events by identifying time points with the same social label within ~39-second-long time windows. For each event and participant, an OLS regression model without covariates was fitted to the BOLD data (see Figure 1 and sections "2.10.2. Discrete social labelling for each stimulus time point" and "2.10.3. Time window selection and general linear modelling before classification" of the original publication II for more information). Finally, a shallow neural network model (two hidden layers) was trained to predict the social label of each event, based on the neural activation patterns associated with that event using leave-one-participant-out cross-validation. The accuracy of the trained model was determined by calculating the percentage of correct classifications of the left-out-participant's events in each cross-validation round. Chance-level prediction accuracy was estimated with 500 permutations of the model training by shuffling the event labels before each iteration.

In the primary analysis, the classifier was trained with whole-brain data in the 3,000 most selective voxels (the voxels with the highest F-scores in ANOVA voxel selection). Separate classifiers were trained also for each ROI. To account for possible low-level confounding in the classification results, the confound-controlled residual fMRI time series (regressed with the low-level model) were used as input data instead of the original fMRI data.

4.4.2.4 Intersubject correlation analysis

Movies effectively synchronize neural activity between participants (Hasson et al., 2010, 2004) and the degree of synchronization can be measured with intersubject correlation analysis (ISC) (Kauppi et al., 2014). Synchronization is highest in the primary sensory regions, but robust synchronization occurs also in associative brain areas, such as LOTC and STS (Hasson et al., 2010). This indicates that higher-order processes, such as shared social perception, could also synchronize brain responses. To investigate this, we assessed whether the neural activation patterns of the regression and MVPA analyses are associated with neural synchronization. ISC was calculated over the experiment, and the spatial ISC distribution was compared to the cumulative activation map of social features and to the regional classification accuracies indicated by the MVPA analysis.

4.4.3 Eye-tracking data analyses (Study III)

Pupil size, gaze position, fixation rate, and blink rate were extracted dynamically as time series from the eye-tracker reports. Intersubject correlation analysis of eye gaze patterns (eISC) was used to identify gaze synchronization dynamically (Nummenmaa, Smirnov, et al., 2014). The primary temporal resolution for analyses was 500 milliseconds, which allows modeling swift changes in the eye-tracking parameters.

4.4.3.1 Stimulus model for eye tracking

The aim of the eye-tracking study was to investigate how stimulus features guide the visual system in dynamic social scenes. These features were identified at three different levels of cognitive processing: (1) low-level features that describe purely audio-visual properties of the stimulus (e.g., luminance), (2) mid-level features that require semantic categorization (e.g., faces), and (3) high-level social perceptual information (e.g., pleasantness of the situation). This categorization into low-, mid-, and high-level features reflects the cognitive complexity of each feature. Low-level physical information is processed early in the brain before socio-affective

information (Dima et al., 2022), and semantic categorization precedes affective evaluation (Nummenmaa et al., 2010), justifying this division.

Six visual and eight auditory features, along with their time derivatives, were extracted from the stimulus videos to describe the physical qualities of the audio-visual input. Visual features included luminance, visual entropy, optic flow, spatial energy for edge detection, and differential energy for measuring the total change between consecutive frames. Auditory features included audio intensity (RMS), properties of the frequency spectrum (geometric mean, standard deviation, entropy, and high-frequency energy), waveform sign change rate or "noisiness", and sensory dissonance or "roughness". Scene cuts are known to influence the eye movements (Bruckert et al., 2023), and these were identified with ffmpeg tool (ffmpeg, 2025).

Open-source computer vision models were used to segment each movie frame into the following mid-level semantic categories: bodies, objects, background, eyes, mouth, and face (Deng et al., 2020; Keles et al., 2022; Kirillov et al., 2019; Wu et al., 2019). To first segment the whole image into bodies, animals, objects, and background, we used a panoptic feature pyramid network (FPN) segmentation model from the Deceptron2 Python library (Wu et al., 2019). Next, we used the RetinaFace face detection model (Deng et al., 2020), following the implementation of a previous eye-tracking study on autism, to segment rectangular face, eye, and mouth areas from the videos (Keles et al., 2022). Areas not recognized by the models were tagged as unknown. Based on Studies I and II, we selected eight important social perceptual features (pleasant feelings, unpleasant feelings, arousal, pain, talking, body movement, feeding, and playfulness) that are also consistently evaluated across participants. These high-level social perceptual features were evaluated in the movie stimulus dynamically in a four-second temporal scale.

The final design matrix for modeling the eye-tracking parameters was generated by identifying feature clusters among all 39 extracted features using hierarchical clustering (Murtagh & Contreras, 2012) to limit the multicollinearity of the predictors. This clustering identified a total of 16 clusters: seven clusters for low-level features, five clusters for mid-level semantic categories, and four clusters for high-level social information.

4.4.3.2 Total gaze time analysis

To investigate attentional prioritization during dynamic perception, we calculated how long each participant gazed at each semantic category during the experiments. A long total gaze time can be observed just by chance if a category is present frequently (i.e., frequently present and spanning large areas of the screen). A long gaze time could also be due to a centrality bias (Dorr et al., 2010). Hence, a long gaze time does not itself indicate prioritization. The true gaze times should thus be

compared to the estimated chance-level gaze times for the features to reveal the degree of prioritization. A permutation test in which the gaze coordinate time series were circularly bootstrapped (Politis & Romano, 1992) was implemented to break the synchrony between the stimulus and gaze. Chance-level gaze times for semantic categories were calculated after bootstrapping the participants' gaze coordinate time series. This procedure was repeated 500 times to generate the null distribution for gaze times.

4.4.3.3 Multi-step regression analysis

Independent associations of the 16 perceptual predictors with pupil size, eISC, fixation rate, and blink rate were established with a multi-step regression analysis. Even after clustering, the predictors were not fully independent. Hence, a multi-step regression analysis was developed to control the feature-specific association with other predictors and to test the predictive power of different perceptual models. Leave-one-experiment-out cross-validation (three independent eye-tracking datasets) was used to test the prediction accuracy of the regression models in each analysis step.

First, simple regressions were run separately for each social feature. If the regression coefficient's signs for a given predictor were consistent across cross-validation rounds, the predictor was included in the second analysis phase. Otherwise, inconsistency between cross-validation rounds indicated that there was no evidence of significant association between the feature and the given eye-tracking parameter. In the following stepwise regression, consistent predictors were added to a multiple regression model one by one, and the out-of-sample prediction accuracy was tested for each model. If the prediction accuracy for left-out experiment data was higher than would be expected by chance, the feature was considered to be significantly associated with the eye-tracking parameter.

The chance-level prediction accuracy was estimated with a permutation test, in which the last column of the design matrix (newly added predictor) was circularly bootstrapped before fitting the regression model. This procedure was repeated 500 times to generate null prediction accuracies. If the predictor was not considered significant ($p < 0.05$) it was dropped from the next regression model so that the following regression models only included predictors that increased the model's predictive capabilities.

The order of the added predictors likely influences the results, such that the first added predictors are more likely to be considered significant. Thus, we added predictors based on their out-of-sample prediction accuracy in the original simple regression, to give more weight to the features that are expected to have a robust association with the eye-tracking parameter. Weaker predictors in the initial

regression still had a chance to be added to the model if they improved the model's predictions.

4.4.3.4 Gaze prediction analysis

The regression analysis described above was not designed to predict exact gaze locations at any given time. Next, we aimed to predict the population-level gaze probability distributions (gaze heatmaps) in short 200 ms time windows. We selected random forest regression as the analytical approach because it allows data-driven model generation and supports non-linear relationships while at the same time being computationally efficient compared to even more flexible deep neural networks. Random forest regression is based on decision trees in which the original dataset is resampled multiple times and a decision tree is generated for each sample (Breiman, 2001). The final predictions are based on the consensus of all individual decision trees. The method involves optimizable hyperparameters (number of trees, number of branches in each tree, and resampling parameters) Based on within-experiment optimization (80% train / 20% test split), 50 decision trees and 63 branches (6 branches from the tree trunk to the leaf) were selected for the final model fitting. Default resampling parameters of the fitrensemble function (The MathWorks Inc, 2024a) were used (selecting N out of N observations and ⅓ of the predictors).

Random forest models were trained separately for each experiment and then tested to the other two experiments. The primary performance metric was the linear correlation between the out-of-sample predictions and true gaze heatmaps. As a secondary measure, we calculated the Euclidean distance between the predicted and true peak gaze probability values for each heatmap to better understand how well the models can locate the most important areas. Relative predictor importance (The MathWorks Inc, 2024b) was used as a metric to identify which predictors were most influential for predictions. Relative importance does not indicate whether the predictor is positively or negatively associated with gaze probability. Hence, we used simulation to reveal the sign and shape of the association between predictors and gaze probability. The simulation was repeated for each trained model on 200 000 randomly selected pixels from the training dataset by keeping other predictor values constant and randomly testing how the predictions change when the value of a given predictor is altered.

4.4.3.5 Scene cut effect analysis

Scene transitions artificially influence the eye-tracking parameters in cinematic experiments (Bruckert et al., 2023). To quantify the dynamics of the visual system after a scene transition, we extracted pupil size, eISC, and blink rates dynamically

for a 3,600 ms time window around each identified scene cut (from cut – 600 ms to cut + 3,000 ms) for each participant. The population level dynamic was identified by taking the mean response over participants separately for each Experiment. To estimate whether the eye-tracking parameters deviate from the baseline after a scene cut, 3,600 ms time periods were sampled from 100 random time points and then averaged to get a random response pattern. This procedure was repeated 500 times to generate null distributions for eye-tracking parameter dynamics for a random 3,600 ms period. Significant deviation from baseline was identified for time points with $p < 0.05$, based on the permutation test.

4.5 Data and code availability

Copyrights preclude public redistribution of the stimulus movies. The anonymized perceptual data and the analysis scripts for Study I are publicly available in the project's GitHub repository (Santavirta, 2024b). In accordance with Finnish legislation, the original (even anonymized) neuroimaging data used for Study II cannot be released for public use. The voxel-wise (unthresholded) result maps are available in NeuroVault (Santavirta, 2023b), and the analysis scripts are available in the project's GitHub repository (Santavirta, 2023a). The subjects' consent for public distribution of the subject-level eye-tracking data for Study III was not collected, and hence the eye-tracking data cannot be distributed, but the analysis scripts are available in the project's GitHub repository (Santavirta, 2024a).

5 Results

5.1 How humans perceive social environments (Study I)

5.1.1 Low-dimensional model for social perception

Principal coordinate analysis with permutation testing indicated that eight orthogonal principal components explained more variation than expected by chance ($p < 10^{-7}$ or $p < 0.05$) for the correlation structure of 136 social features (Figure 3), accounting for 78% of the total variation. PC1 (*Pleasant - Unpleasant*) extracted the overall emotional valence of the presented social situation by ordering features from pleasant (e.g., "Pleasant", "Interacting positively", and "Feeling calm") to unpleasant ones (e.g., "Feeling displeasure", "Feeling disappointed", and "Feeling pain"). PC2 (*Empathetic - Dominant*) revealed the perceived dominance structure of the social interaction by ordering features from dominant ("Dominance", "Authority", and "Stubborn") to empathetic characteristics ("Intimate", "Affectionate", and "Crying"). PC3 (*Physical - Cognitive*) ordered features from physical engagement, energy consumption and impulsive behavior (e.g., "Interacting physically", "Nude", "Feeling energetic", "Moaning", and "Jumping") to cognitive reasoning and controlled behavior (e.g., "Thinking or reasoning", "Formal", and "Intelligent"). PC4 (*Disengaged - Loyal*) captured the perceptual distinction between inactive self-related behavior, where people are disengaged from others (e.g., "Lazy", "Superficial", "Daydreaming", "Eating something", and "Hungry/thirsty") and proactive behaviors (e.g., "Conscientious", "Loyal", and "Brave"). PC5 (*Introvert - Extravert*) related to the perception of broad social engagement and personality types, introversion (e.g., "Alone", "Pursuing a goal", and "Introvert") and extraversion (e.g., "Talking", "Making gaze contact", and "Extravert"). PC6 (*Playful - Sexual*) described the evaluation of social interactions based on their affiliative (e.g., "Joking", and "Laughing") versus sexual nature (e.g., "Sexual", "Sexually aroused", and "Nude") of the interaction. PC7 (*Alone - Together*) simply described whether the scene involved social interaction between people or not, and PC8 (*Feminine - Masculine*) described a dimension for perceived sex characteristics.

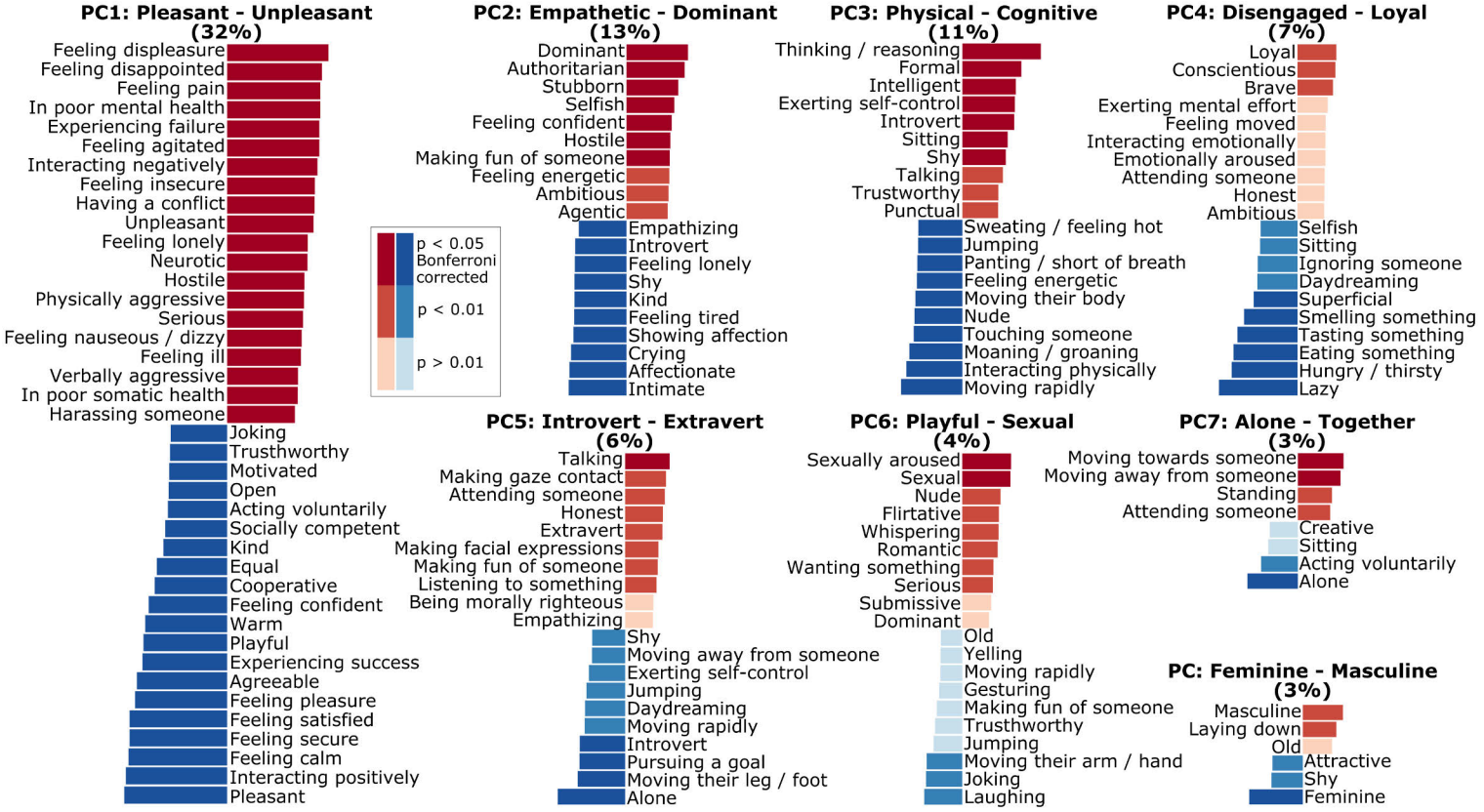

Figure 3. Social perceptual dimensionality based on the PCoA analysis. Eight PCs explained statistically significant amounts of variation (78% in total). The explained variance of each PC is shown in parentheses. Descriptive names for the perceptual dimensions were inferred from the feature loadings (the importance of the original feature for the PC). The barplots show these original feature loadings. A permutation test was used to assess whether the loadings deviate statistically from zero to ease the interpretation. Modified from the original publication. Copyright © 2024 by American Psychological Association. Adapted with permission (Santavirta, Malén, et al., 2024).

5.1.2 Social perceptual clusters and their concordance with PCoA components

Hierarchical clustering confirmed that social perception organizes most saliently around emotion valence, indicated by the negative correlation between pleasant and unpleasant features (Figure 4**)**. The main difference between the PCoA components and HC clusters was that clustering revealed a more fine-grained representation of social information. The concordance analysis, based on t-SNE plotting and the estimation of PC loadings for each HC cluster, indicated that the information conveyed by the clusters can be constructed as unique combinations of PC information. This was indicated by separability of clusters in the t-SNE space (Figure 5). More specifically, PC cluster loadings described how each cluster could be described with the PCs (see Figure SI-3 of the original publication I). For example, features that were assigned to the "Antisocial behavior" cluster were perceived as unpleasant (PC1) and dominant (PC2) based on their PCoA component loadings. This distinguished them from features in the "Unpleasant feelings" cluster, which described unpleasant (PC1) but not dominant (PC2) characteristics. Similarly, features in the cluster "Emotional affection" loaded as pleasant (PC1) and empathetic (PC2), while features in the cluster "Extraversion & playfulness" loaded as pleasant (PC1) and dominant (PC2), distinguishing these pleasantly perceived characteristics from the solely pleasant features in the cluster "Pleasant feelings & prosociality".

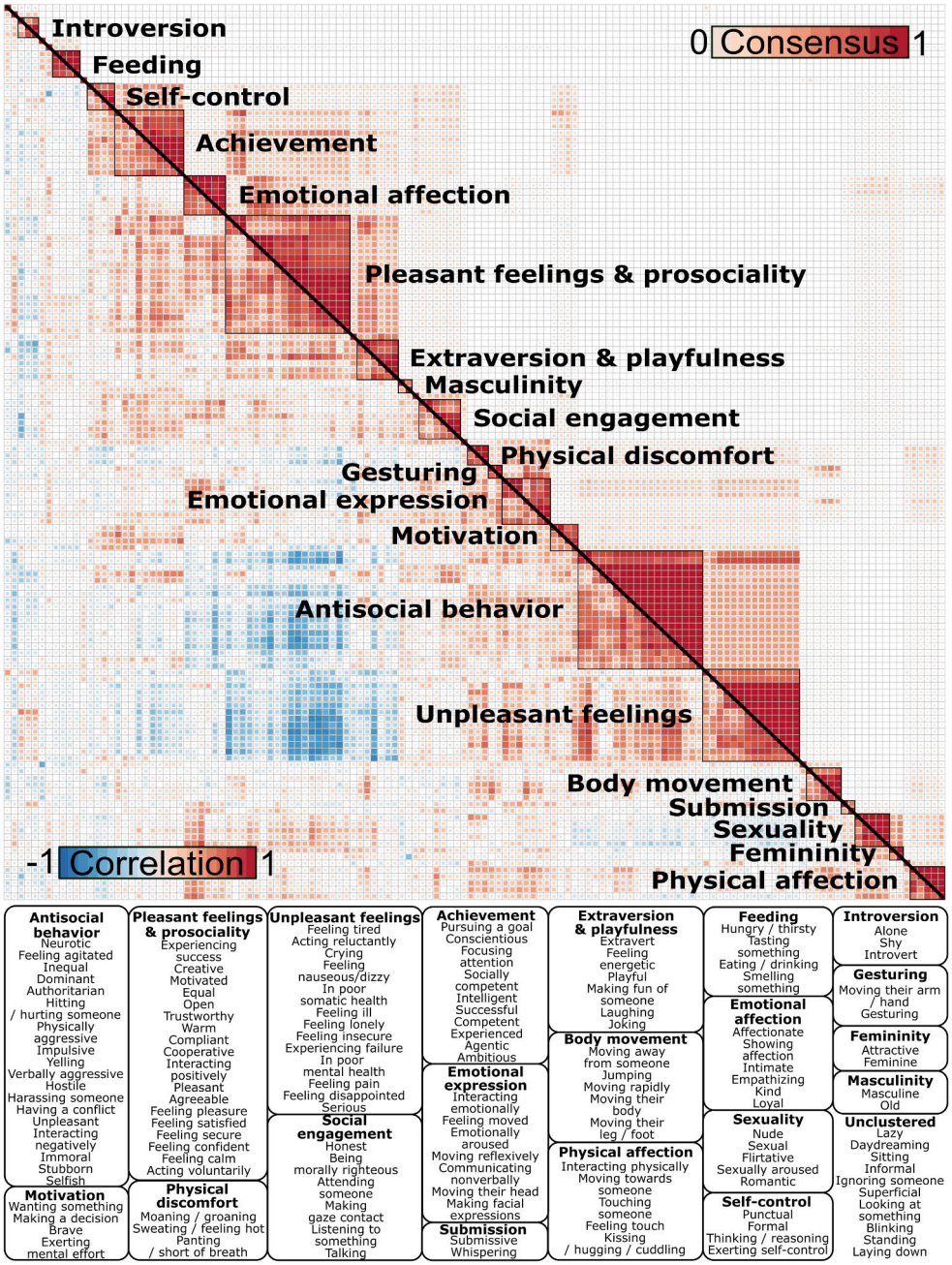

Figure 4. Results of the clustering analysis. The upper triangle shows the consensus matrix, and the lower triangle the correlation matrix of social features. The consensus matrix indicates how many times, out of all subsamples, the pair of features were clustered together. The boxes at the bottom show which social features belonged to each cluster. Reprinted from the original publication. Copyright © 2024 by American Psychological Association. Reproduced with permission (Santavirta, Malén, et al., 2024).

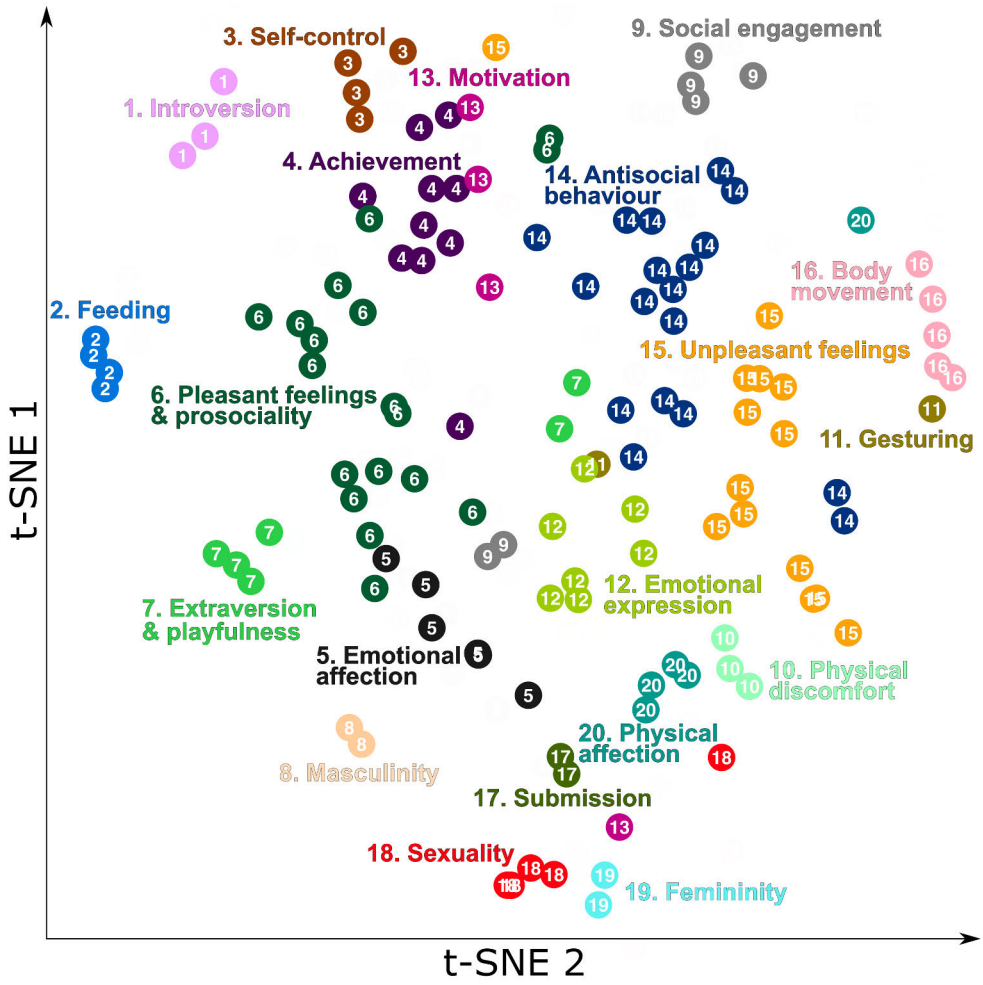

Figure 5. The relationship between HC clusters and PCoA components. T-SNE projects the original 136 social features into a two-dimensional plane based on their loadings to the eight significant PCs. The color and number of each point indicate which cluster the feature was assigned to in the HC analysis, and the cluster labels are embedded in the figure. The clusters are well separated in the t-SNE space indicating that the PCoA components and HC clusters have a structural relationship. Modified from the original publication. Copyright © 2024 by American Psychological Association. Adapted with permission (Santavirta, Malén, et al., 2024).

5.1.3 Generalizability of the social perceptual structure

After establishing the social perceptual structure for the primary movie clip dataset, we tested the generalizability of the social perceptual structure with two independent datasets by using PCoA and HC analysis. Generalizability across different movie stimuli was tested with a retrospective full-movie dataset with similar perceptual ratings for a subset of the features. Each identified PC in the primary dataset

correlated significantly with a corresponding PC in the validation full-movie dataset ($p < 0.05$, Figure 6a). The structure of social perception based on HC analysis also generalized well (similarity of correlation matrices: $r = 0.68$, $p < 10^{-6}$, similarity of consensus matrices: $r = 0.54$, $p < 10^{-6}$, Figure 6b). Figure 7 shows how the clustering result changed between the independent datasets.

Generalizability across stimulus types, from videos to images, was tested prospectively with a dataset of images captured from the primary movie clip dataset. Each PC in the primary movie clip dataset showed significant and high correlation with a corresponding PC in the movie frame dataset ($p < 0.05$) and the structural similarity based on clustered correlation matrices was high ($r = 0.92$, $p < 10^{-6}$).

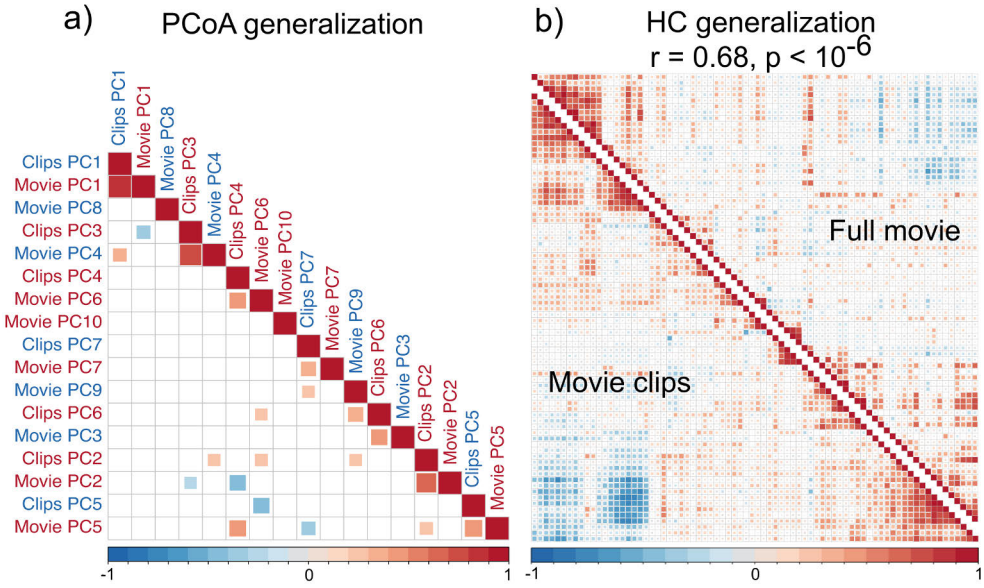

Figure 6. Generalizability of social perceptual structure across cinematic stimuli. **A)** Correlations between the PC feature loadings of corresponding PCs between primary movie clip dataset and validation full-movie dataset (only significant $p < 0.05$ correlations are shown). **B)** Independently clustered correlation matrices between primary and validation datasets were highly similar ($r = 0.68$, $p < 10^{-6}$). Modified from the original publication. Copyright © 2024 by American Psychological Association. Adapted with permission (Santavirta, Malén, et al., 2024).

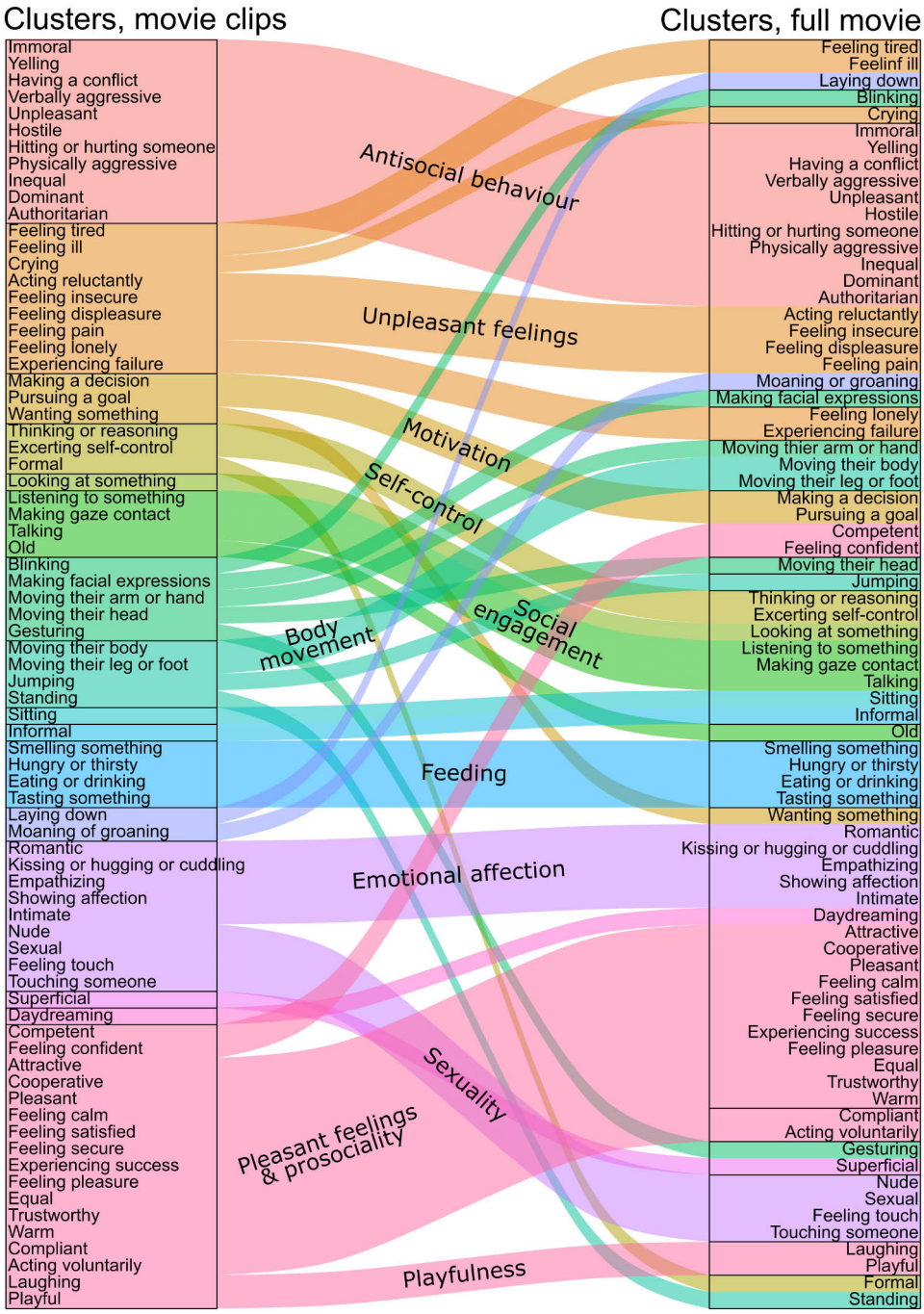

Figure 7. Generalizability of clustering structure across movie stimuli. The two columns are ordered based on the independent clustering result of the datasets. The alluvial diagram thus shows how the clustering structures align between the datasets. Modified from the original publication. Copyright © 2024 by American Psychological Association. Adapted with permission (Santavirta, Malén, et al., 2024).

5.2 How the brain processes social information in dynamic scenes (Study II)

5.2.1 Neural responses for social perceptual features

Based on data-driven hierarchical clustering of 45 most consistently perceived social features (ICC > 0.5), we defined a social perceptual model that included 13 distinct social predictors. Multiple regression with Ridge regularization established the neural activation patterns for social features (Figure 8). Social perceptual processing engaged areas in both hemispheres and in all brain lobes. Most features were associated with the brain responses in occipital, temporal, and parietal regions, while frontal and subcortical regions responded more selectively to only a few features. The Heschl's gyrus and superior temporal gyrus in temporal lobe, and superior occipital gyrus and calcarine sulcus in occipital lobe, were significantly associated with all social features. A few social features, mainly those that are arousing or pleasurable (Antisocial behavior, Sexual & affiliative behavior, Feeding), were associated with brain responses in frontal midline and subcortical areas.

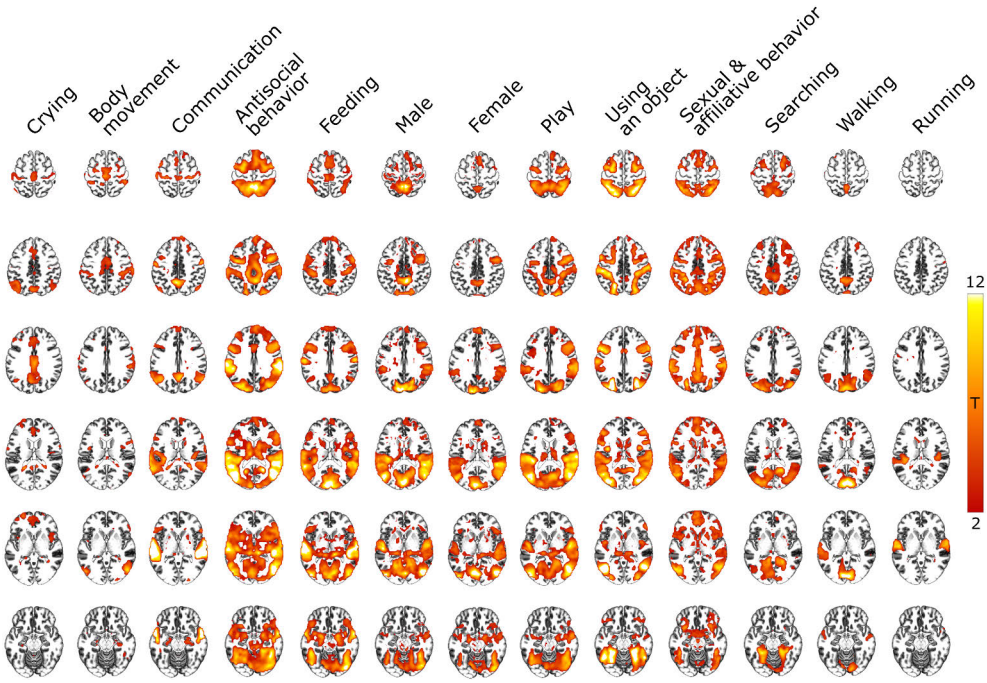

Figure 8. Brain regions that were positively associated with the social features in multiple regression analysis. Results show the statistically significant positive T-values (voxel-level FDR-corrected, q = 0.05). Reprinted from the original publication (Santavirta et al., 2023).

5.2.2 Cerebral gradient in social perception

Figure 9a shows the cumulative activation map (a binarized sum of the results visualized in Figure 8) over the 13 social features. This map highlights a cerebral gradient in social perception, where temporal and occipital areas were associated with most of the social features. The regional selectivity increased towards frontal and subcortical areas, where associations were observed with only a few social features. The functional network, including superior temporal sulcus (STS), lateral occipitotemporal cortex (LOTC), temporoparietal junction (TPJ), fusiform gyrus (FG), as well as precuneus and auditory and visual cortices, were broadly tuned to different social features. Figure 9b shows the voxel-specific ISCs over the cortex, indicating the brain areas that were highly synchronized across participants during the experiment. The association between the cumulative neural activations for social features and ISC was high (r = 0.86), indicating that response generality vs. selectivity for social features was positively associated with the neural synchronization.

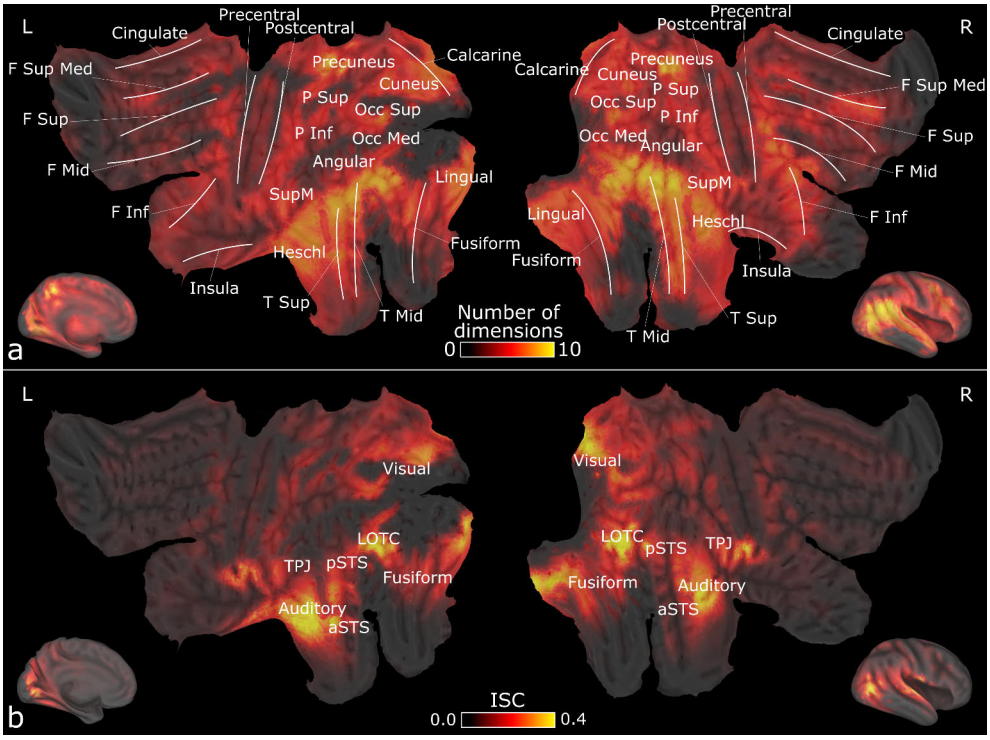

Figure 9. The cumulative activations for social features. **A)** A cumulative activation map showing how many social features were positively associated (voxel-level FDR-corrected, q = 0.05) with the neural responses at each voxel. Major gyri are localized with white lines. **B)** Neural synchronization based on intersubject correlation (FDR-corrected, q = 0.05). Selected functional regions are highlighted. Modified from the original publication (Santavirta et al., 2023).

5.2.3 Classifying social context from the neural responses

To further investigate how specific the spatial activation patterns were for each social feature, a multivariate pattern analysis (MVPA) was conducted to predict the social context (i.e., the most prevalent social feature) from the spatial hemodynamic activation patterns of the fMRI data. The whole-brain classifier achieved a 52% accuracy in predicting the correct social context from 11 possible choices, which was significantly above the permuted chance level ($p < 0.01$, $acc_{chance} = 0.128$). Prediction accuracies/positive predictive values for each social feature in the whole-brain classification were: walking: 0.49/0.51, using an object: 0.53/0.50, searching: 0.70/0.69, running: 0.56/0.62, sexual & affiliative behavior: 0.45/0.48, play: 0.53/0.51, feeding: 0.46/0.48, crying: 0.46/0.51, communication: 0.55/0.55, body movement: 0.52/0.50, and antisocial behavior: 0.55/0.53.

The whole-brain classifier achieved higher prediction accuracy than any of the ROI classifiers. The 3,000 ANOVA-selected voxels for the whole-brain classification localized in the social perceptual areas in STS, LOTC, TPJ, FG, and in the occipital cortex (Figure 10a). Regional classifiers yielded lower classification accuracies than the whole-brain classifier and revealed a cerebral gradient from high classification accuracies in the occipital and temporal cortices to low and near chance-level classification accuracies in the frontal and subcortical areas (Figure 10b). The association between regional prediction accuracies and the ISC of neural responses was high ($r = 0.85$), indicating that neural responses are spatially more specific for social context in highly synchronizing brain regions.

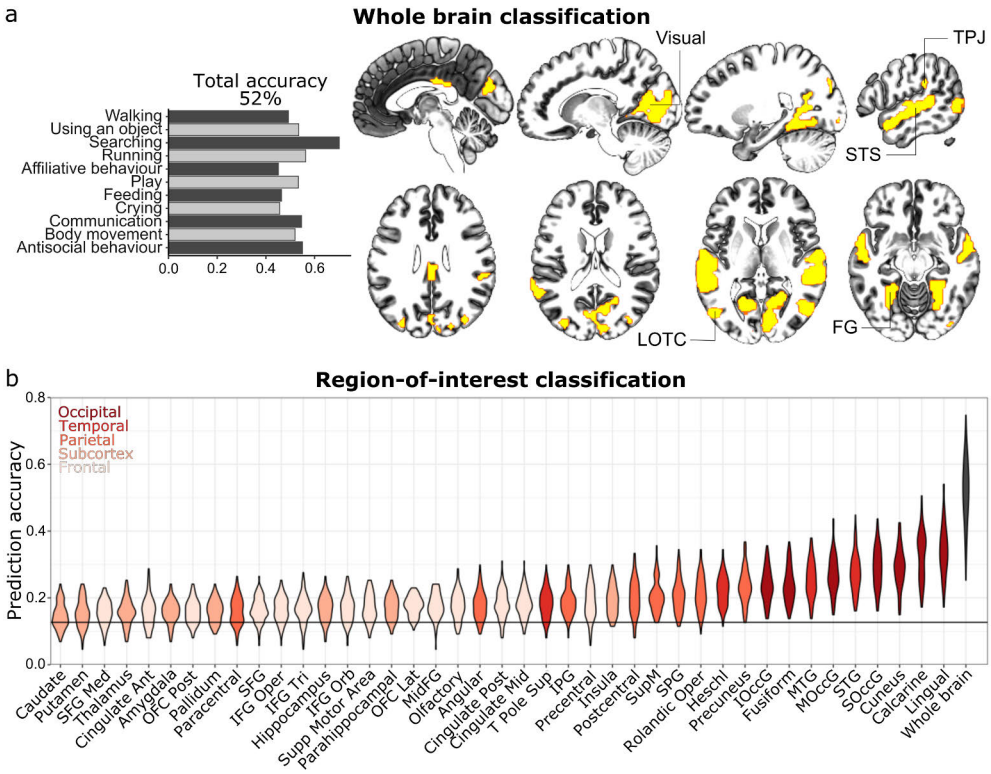

Figure 10. Results of the multivariate pattern analysis. **A)** Bar plots show the classification accuracies of the whole-brain classifier for each social feature, alongside the localization of the voxels that provided information for the classifier. **B)** Violin plots show the prediction accuracies of ROI classifiers compared to the whole-brain classification (shown on the right). The lobar localization of each ROI is indicated with color-coding, and the permuted chance-level classification accuracy (acc = 0.128) is plotted as a horizontal line. Modified from the original publication (Santavirta et al., 2023).

5.2.4 Comparison between the social and low-level models

As a *post hoc* analysis, we investigated where in the brain the social model predicted the BOLD signals more accurately compared to the low-level model (Figure 11). The social and low-level models were separately fitted to the BOLD data, and their fits were estimated with adjusted R^2. Based on the adjusted R^2-values, the social model predicted neural responses more accurately in the identified social perceptual network (STS, LOTC, TPJ, FG, and IFG), while the low-level model was significantly more accurate, particularly in the primary visual and auditory areas (as well as outside grey matter). MVPA, conducted within voxels where the social model predicted neural responses significantly more accurately than the low-level model (hot areas in Figure 11), achieved 35% ($p < 0.01$) classification accuracy,

which was slightly higher than the highest ROI classification accuracy (34% in lingual gyrus).

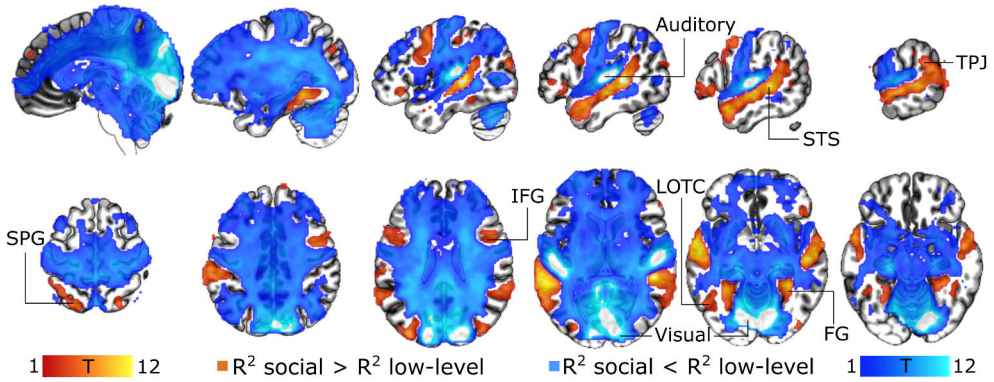

Figure 11. Comparison between the social and low-level models. Hot colors indicate areas where the social model predicted the BOLD signal more accurately (FDR-corrected, q = 0.05) than the low-level model, while blue colors indicate regions favoring the low-level model. Modified from a supplementary figure in the original publication (Santavirta et al., 2023).

5.3 How the visual system is externally modulated by dynamic social scenes (Study III)

5.3.1 Attentional prioritization of social cues

The gaze time analysis revealed how much time participants allocated to viewing specific semantic categories of the stimulus movies (Figure 12). In all three experiments, the participants showed an attentional preference for the eyes and mouth areas. Between 21% and 33% of the total viewing time was allocated to the eyes, which was significantly more than expected if people had watched the scenes randomly ($p < 0.005$, estimated chance gaze time = 4% - 9%). Between 10% and 11% of the viewing time was allocated to the mouth area ($p < 0.005$, estimated chance gaze time = 2% - 4%). The difference between viewing times for objects and face areas (excluding eyes and mouth) was small and inconsistent between Experiments. A relatively high proportion of viewing time was allocated to bodies (18% - 28%) and the background (15% - 23%), but these areas were still given low priority, since more viewing time would have been expected by chance ($p < 0.005$).

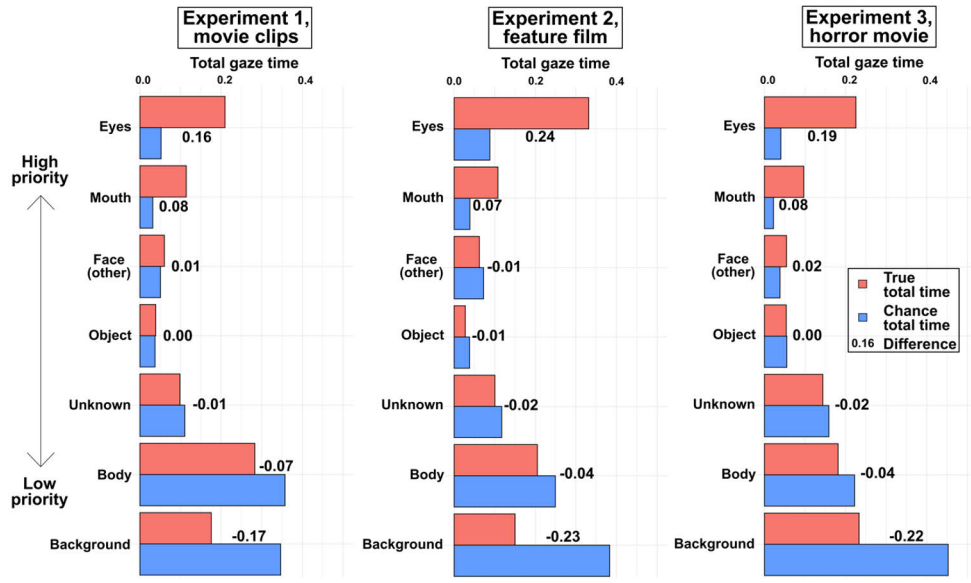

Figure 12. Results of the gaze time analysis. Total gaze times for the semantic categories are plotted separately for each Experiment. The red bars show how long (proportional to the total stimulus duration) the participants on average gazed at each given category, while the blue bars indicate the expected total gaze times if people had watched the scenes randomly. All differences between true and chance gaze times were statistically significant ($p < 0.005$). Reprinted from the original publication (Santavirta, Paranko, et al., 2024).

5.3.2 Multi-step regression results

Multi-step regression established reliable associations between stimulus features and the dynamic eye-tracking parameters of interest. An association was considered reliable if the feature (1) showed consistent association with the eye-tracking parameter across cross-validation rounds in simple regression, and (2) significantly increased the stepwise multiple regression model's out-of-sample prediction performance ($p < 0.05$). Reliable associations are marked with asterisks in Figure 13.

Pupil size was reliably associated with low-level features and perceived unpleasantness of the social scenes. The final model with the five reliable predictors yielded high performance in predicting dynamic pupil size changes in the left-out experiment data (Exp. 1 as the test set: $r = 0.36$; Exp. 2 as the test set: $r = 0.49$; Exp. 3 as the test set: $r = 0.50$).

EISC was reliably modeled with mid-level features and overall scene motion. The final model with the four reliable predictors yielded high performance in predicting dynamic eISC (Exp. 1 as the test set: $r = 0.43$; Exp. 2 as the test set: $r = 0.33$; Exp. 3 as the test set: $r = 0.40$).

Fixation rate was reliably associated with mid-level features and audio intensity/roughness. The final model with the seven reliable predictors yielded moderate performance in predicting changes in fixation rate (Exp. 1 as the test set: r = 0.26; Exp. 2 as the test set: r = 0.21; Exp. 3 as the test set: r = 0.24).

Blink rate reliably associated with three predictors. However, the final model with these predictors was unable to predict substantial variation in blink rate changes (Exp. 1 as the test set: r = 0.03; Exp. 2 as the test set: r = 0.02; Exp. 3 as the test set: r = 0.03).

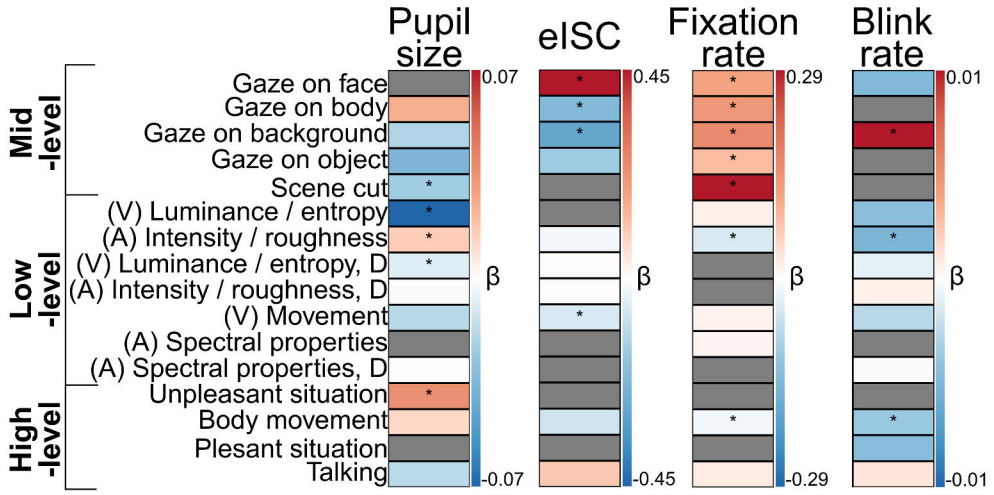

Figure 13. Independent associations between pupil size, eISC, fixations rate, blink rate and the stimulus features in the multi-step regression analysis. The regression coefficients from simple regressions are plotted with a blue (negative) - red (positive) color gradient to indicate the features that had a consistent association with the eye-tracking parameters (not tested for statistical significance). Gray color indicates that the coefficient estimate was inconsistent between the cross-validation rounds. An asterisk indicates that adding the feature increased the stepwise regression model's out-of-sample prediction performance ($p < 0.05$), signifying a reliable association. (V) denotes low-level visual predictors, (A) denotes auditory low-level predictors, and D denotes the time derivative of the feature. Reprinted from the original publication (Santavirta, Paranko, et al., 2024).

5.3.3 Gaze probability prediction

Random forest regression models were trained separately on each Experiment's data to predict the out-of-sample gaze probability distributions from the other Experiments' data in 200 ms temporal resolution. The trained models achieved robust out-of-sample prediction performance, as indicated by the high correlations (0.41 - 0.47) between the true and predicted gaze probability distributions (Figure 14). On average, the predicted peak gaze probability was located within 10% - 16%

of the image width from its true location, highlighting the models' ability to consistently capture the most gazed areas.

Based on their relative importance, the eyes, mouth, visual motion, and visual luminance/entropy were the most influential predictors of gaze location (Figure 14, bar plots) consistently across the independently trained models. These four predictors were all positively associated with gaze probability, based on the subsequent simulations using the trained models. High-level social information does not localize to any specific screen position, which prevents a direct association between momentary gaze locations. Instead, social information could have influenced predictions by interacting with a pixel-wise predictor (e.g., the eyes are viewed more closely in unpleasant situations). Exploratory simulations of interactions between the perceived social predictors and the four most important predictors, however, did not indicate any clear interaction effect.

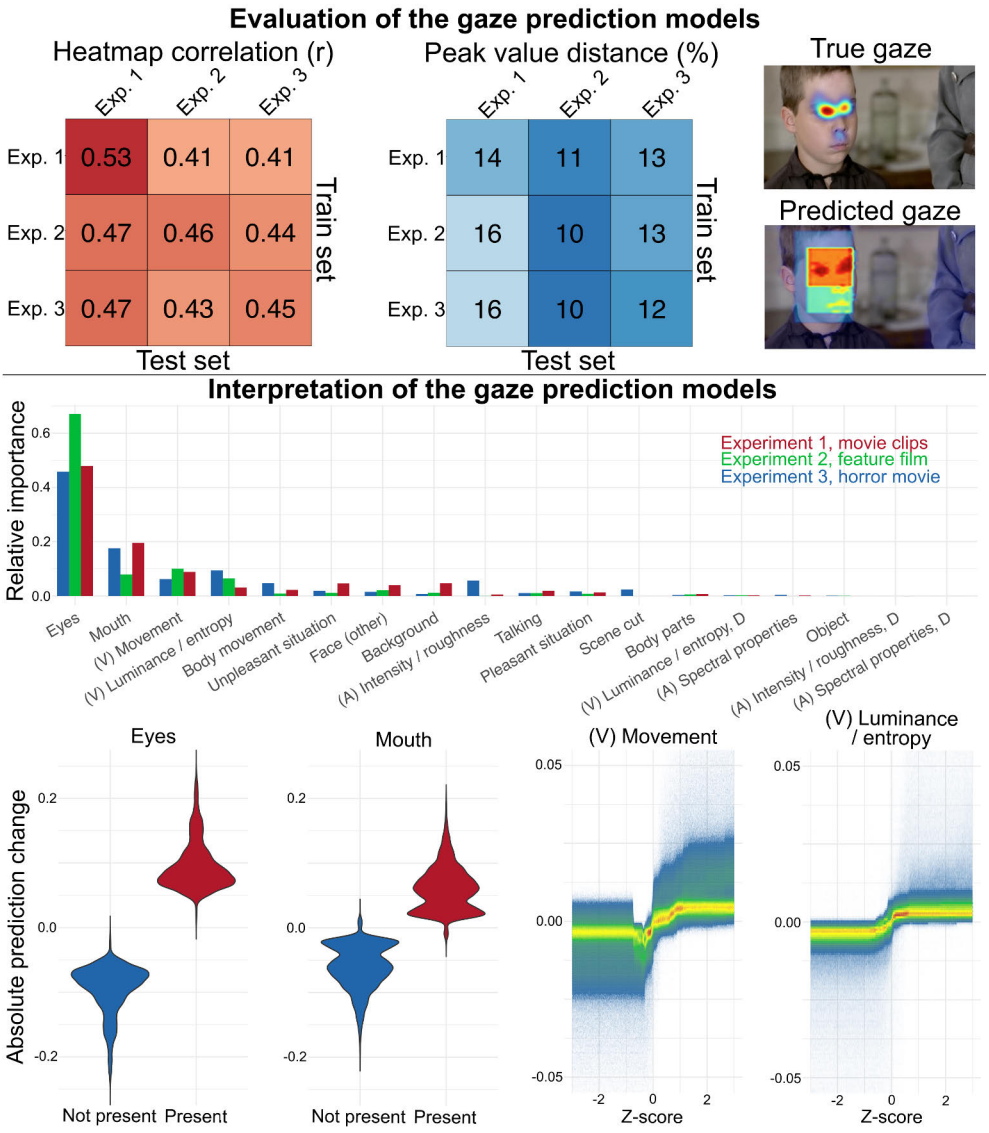

Figure 14. Performance and interpretation of the gaze prediction models. Top: The left confusion matrix shows the correlation between the true and predicted gaze probability distributions. The right matrix shows the average distance between the true and predicted locations of the peak gaze probabilities. Middle: Relative importance bar plots show how influential each predictor was in gaze prediction. Bottom: Violin and density plots indicate the simulated associations between the most important predictors and gaze probability. (V) denotes low-level visual predictors, (A) denotes auditory low-level predictors, and D denotes the time derivative of the feature. Modified from the original publication (Santavirta, Paranko, et al., 2024).

5.3.4 Scene cut effects

Figure 15 shows the average temporal dynamics of pupil size, eISC, and blinking behavior after a scene cut in the stimulus movies. The results were consistent between Experiments. Pupil diameter decreased shortly after the transition and remained smaller compared to baseline from 350 ms to 1,150 ms in all datasets ($p < 0.05$). In contrast, eISC increased after a scene cut briefly, lasting between 200 ms and 800 ms ($p < 0.05$) in Experiments 2 and 3, where the stimuli were continuous movies. In Experiment 1, with unrelated short movie clips, eISC remained elevated after a scene cut for a longer period (from 400 ms to 1,400 ms, $p < 0.05$). Fewer participants blinked during the first moments of a new scene compared to baseline (Exp. 1: 0 ms - 400ms after scene cut; Exp. 2: 200 ms - 400 ms after scene cut; Exp. 3: 0 ms - 200 ms after scene cut, $p < 0.05$). In Experiment 1, with the unrelated movie clip stimulus, blink suppression was followed by a short period of increased blinking (from 600 ms to 1000ms, $p < 0.05$), which, however, did not replicate in the other Experiments.

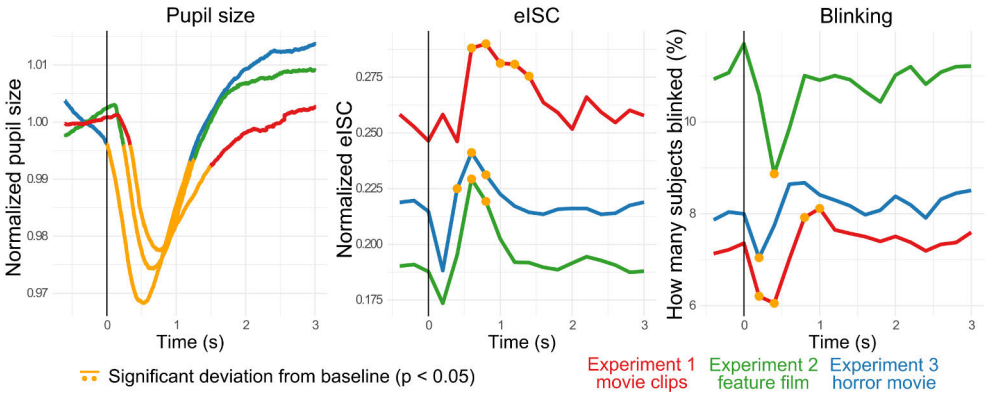

Figure 15. Temporal dynamics of the pupil size, eISC, and blinking behavior after a scene cut. The line plots show the average eye-tracking dynamics around scene transitions separately for each Experiment. The scene cut time is marked with a vertical line. The yellow period (for pupil) and yellow dots (for eISC and blinking) indicate statistically significant ($p < 0.05$) deviations from baseline. Reprinted from the original publication (Santavirta, Paranko, et al., 2024).

6 Discussion

This doctoral thesis investigated social perception in humans by studying the entire perceptual processing stream. The studies explored how people visually sample the social environment, how social perceptual information is processed in the brain, and how people make perceptual inferences from available social cues. To model life-like social contexts, movies were selected as naturalistic stimuli due to their rich social content and dynamic nature.

The results revealed that human social vision is predominantly guided by low-level physical features (e.g., luminance, motion), and mid-level social features (e.g., eyes, mouths, faces), rather than high-level social information (e.g., perceived pleasantness). Furthermore, blinking was suppressed during intense moments (during scene changes, intense/rough sounds, and perceived body movement) indicating attentional engagement. Pupillary responses varied as a function of luminance, emotional arousal, and rapid visual change: constriction occurred due to increased luminance and scene transitions, whereas dilation occurred during emotional arousal.

Functional brain imaging revealed that the network supporting social perception spans both hemispheres, localizing primarily in the occipitotemporal cortices. STS, LOTC, TPJ, and FG were established as the main hubs for social perception exhibiting neural responses to multiple social features with spatially unique activation patterns.

Finally, this work established that individuals make rapid inferences about social situations by evaluating them across a limited number of orthogonal evaluative dimensions. Based on the findings, we propose *eight basic dimensions of social perception* as a model for the organization of human social perceptual processing. These dimensions suggest that humans assess the social environment initially by inferring emotional valence, the balance between empathy and dominance, and the cognitive versus physical characteristics of behavior along with five additional dimensions.

6.1 Modeling human social vision

Figure 16 summarizes how the human visual system is modulated by external stimulus features during movie viewing.

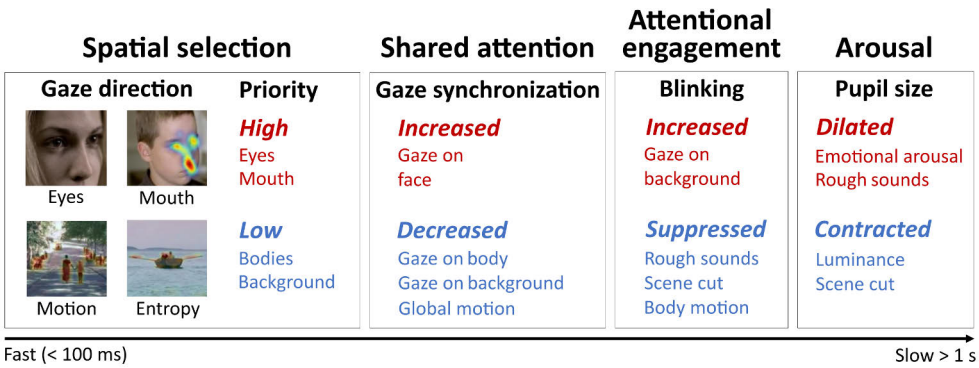

Figure 16. Bottom-up modulation of the human visual system during naturalistic movies. Human faces, especially the eyes, in parallel with low-level local information, drive the immediate visual sampling of the social environment. Intense scenes and scene transitions inhibit eye blinks, indicating attentional engagement with the stimuli. Pupillary responses are modulated by emotional arousal, scene cuts, and luminance changes. Reprinted from the original publication (Santavirta, Paranko, et al., 2024).

6.1.1 Visual attention during dynamic social scenes

Physical stimulus features are robust modulators of visual attention, as demonstrated by the high out-of-sample prediction accuracy for gaze synchronization ($r_{eISC} = 0.43$) and gaze probability distribution ($r_{gaze} = 0.47$). The results consistently showed that human eyes and mouths are the best predictors of gaze behavior at any moment (Figure 14). Human faces receive attentional priority over other elements, indicated by participants looking at faces far more frequently than would be expected by chance (Figure 12). Specifically, participants gazed at faces from 38% to 50% of the total stimulus duration (depending on the Experiment) compared to the expected 10% to 20% if they inspected the scenes randomly. This face priority led to fewer fixations on bodies and the background than would be expected by chance.

Regression analysis further confirmed this face priority, revealing that gaze synchronization increased when fixating on faces and decreased when gazing at bodies or backgrounds (Figure 13). The high attention to human faces has been previously established in controlled studies with simple stimuli (Bindemann et al., 2005; Morrisey et al., 2019; Theeuwes & Van der Stigchel, 2006). The current results expand the understanding of the face priority to the perception of dynamic social scenes by indicating the biological salience of faces and their ability to capture shared attention (Morrisey et al., 2019). Faces are a rich source of social cues whose

rapid identification is important for inferring more abstract social properties, such as affection or hostility (Freeman & Ambady, 2011). Previous controlled studies have emphasized that gaze behavior is simultaneously modulated by low-level visual features and social information so that social information is prioritized when it is available (End & Gamer, 2017). Social information prioritization is probably reflexive rather than voluntary (Rösler et al., 2017) and also task-independent (Smith & Mital, 2013) suggesting that orienting to social cues is highly automatic.

Low-level cluster of features related to luminance and entropy also modulated the gaze direction in social scenes (Figure 14). This low-level information cluster reflects how clearly an area in the video stands out, by combining local luminosity, visual entropy (a measure of randomness) and spatial energy (detecting local forms) thus relating to the density of low-level visual information. High visual contrast and the pixel intensity randomness have been shown to capture attention in static images (Krieger et al., 2000; Reinagel & Zador, 1999), supporting our finding that visually dense areas that stand out from the background capture attention.

Motion also modulated gaze direction. *Global* motion led to decreased gaze synchronization (Figure 13), likely due to increased need for fixation adjustments in highly dynamic scenes. However, the gaze prediction analysis indicated that *local* motion predicted where people look in social scenes (Figure 14), consistent with previous findings in humans (Abrams & Christ, 2003; Bruckert et al., 2023) and macaques (Mahapatra et al., 2008). Moreover, motion captures attention automatically, supporting its prioritization in visual processing (Smith & Mital, 2013).

The scene transitions cause abrupt discontinuities in cinema, which motivated further analysis of vision dynamics post-cut (Figure 15). Results showed that gaze synchronization increased temporarily (up to 800 ms in continuous movies and up to 1400 ms in uncorrelated clips) before returning to baseline, aligning with findings on scene transitions and gaze synchronization (Mital et al., 2011). This suggests that a consistent, time-locked orientation response takes place when new scene content is introduced. However, regression analysis did not find significant associations between gaze synchronization and scene transitions, likely because synchronization was better explained by other predictors, such as new face locations and changes in low-level features.

6.1.2 Dynamic modulation of the pupillary responses

Cross-validated regression models successfully predicted (up to $r = 0.5$) out-of-sample pupillary responses during movie viewing, indicating that the pupil size is influenced by bottom-up stimulus features. An expected negative association was found between scene luminosity and pupil size, as shown by the significant relationship between the “luminance/entropy” predictor and pupil size (Figure 13).

In turn, pupil dilated in response to perceived unpleasantness in the scenes. This cluster of unpleasant characteristics included perceived unpleasantness, arousal, aggression, and pain, indicating that the unpleasant scenes were also highly arousing. This implies that pupil dilation was driven by the emotional arousal or negative valence (or both) evoked by the scenes.

No association was found between perceived pleasantness and pupillary response, suggesting that the pupillary responses are more likely linked to emotional arousal than emotional valence. Moreover, the current findings indicated a positive relationship between sound intensity/roughness and pupil size. Intense sounds are arousing and alerting (Dean et al., 2011; Di Stefano & Spence, 2022; Ilie & Thompson, 2006; Trevor et al., 2020), which was also supported by a moderate correlation between perceived arousal and audio intensity ($r = 0.39$) in the present data. Some controlled studies have reported that both unpleasant and pleasant stimuli can induce pupil dilation across different stimulus modalities (Bradley et al., 2008; R. R. Henderson et al., 2018; Partala & Surakka, 2003), while others have found that unpleasant stimuli have a stronger effect (Babiker et al., 2013; Kawai et al., 2013), aligning with the present results.

The regression analysis (Figure 13) and scene transition dynamics (Figure 15) demonstrated a negative association between pupil size and scene transitions. Specifically, the pupil begins to constrict rapidly following a scene transition, with peak constriction occurring approximately 500 - 800 ms after the cut. The pupil then dilates back to baseline between 1,150 – 1,500 ms after transition. This response occurred on the same temporal scale as the ISC response to scene transitions. Pupil constriction after scene transition is likely a general response to sudden changes in visual input, consistent with findings from studies using controlled, simple stimuli under isoluminous conditions (Kimura et al., 2014).

Altogether, these results suggest that pupil size changes during naturalistic movie viewing are a combination of luminance-driven dilation, constriction during scene changes, and constriction during emotional arousal. This shows how low-level sensory processing and higher-order emotional arousal together dynamically shape the pupillary dynamics.

6.1.3 Blinking indicates attentional disengagement

Blinks occurred approximately once in every five seconds (0.2 Hz) in the present data, accounting for only a fraction of the total viewing time (~ 2%). Despite this, important information can be missed if blinks happen at critical moments. Analyses of blinking revealed that blinking is briefly suppressed after scene transitions (Figure 15) and that blink rate is negatively associated with sound intensity/roughness and perceived body movement (Figure 13). Additionally, blink rate increased when participants gazed at

the background. However, this finding should be interpreted with caution, as the initiation of a blink can confound the last recorded fixation location shifting it downwards from the center to areas that are more likely background.

These findings suggest that attentional engagement inhibits blinking during moments of high behavioral intensity. Previous research supports this interpretation (Ranti et al., 2020; Shin et al., 2015), showing that blink frequency decreases as attentional demands rise, and that blinks often occur at attentional breakpoints (Nakano et al., 2009; Wyly et al., 2024). Shared movie stimuli synchronizes blinking across viewers and narrative movie clips lead to lower blink rates compared to nature documentaries lacking a clear storyline (Nakano et al., 2009; Shin et al., 2015). Blink synchronization is also higher among participants who are more engaged with the topic (Nakano & Miyazaki, 2019). These findings support the idea that blinking is indeed inhibited during attentional engagement and that increased interest results in stronger attentional engagement and fewer blinks.

Functional neuroimaging findings show that brain activity shifts from the dorsal attention network to the default mode network after blink onset, adding evidence for attentional disengagement during blinking (Nakano et al., 2013). Thus, the present results indicate blinking as an indicator of attentional disengagement, and that higher-order social perceptual information has minimal impact on blinking. The regression models using external stimulus features were not able to predict much out-of-sample variation in the blink rates, which suggests that blink rates are more intrinsically regulated than pupil size and gaze direction.

6.2 Functional organization of social perception networks in the human brain

Figure 17 summarizes the results of Study II.

6.2.1 Social perceptual model for fMRI analysis

Data-driven hierarchical clustering indicated that the perceived social information from the movie stimuli can be modeled with 13 relatively independent social predictors. This dimensionality should be considered preliminary, as it depends on lesser data and simplified analytical tools compared to the dimensionality established in Study I. The neuroimaging experiment design also necessitated that we focus on the most consistently evaluated social features (ICC > 0.5), as the neuroimaging participants did not evaluate the social features themselves. This shifted the focus towards more observable social features compared to Study I. Despite these constraints, the identified perceptual clusters align well with the clusters defined in Study I, which indicates that pleasant versus unpleasant situations, playfulness,

sexuality, feeding, masculinity & femininity, and body movement, among others, are important social perceptual features.

6.2.2 Cerebral gradient in social perception

Multiple regression analysis with 13 social predictors and low-level covariates indicated that a distributed cortical network encodes the social content of the video stimuli (Figures 8 and 9). Most social features were associated with an increased BOLD response in STS, LOTC, TPJ, and FG, as well as other occipitotemporal and parietal areas. While these regions responded to a broad range of social features, more selective responses emerged in the frontal and subcortical areas, where only a few social features were significantly linked to neural activity. This finding suggests that social perceptual information is primarily processed in the caudal and lateral brain regions adjacent to primary auditory and visual areas.

Anatomically, the most consistent activations for social features were observed across all occipital regions and in temporal regions FG, STG, MTG, (STS is located between STG and MTG) and the Heschl's gyrus. In the parietal cortex, the most consistent responses were found in supramarginal gyrus (part of TPJ), SPG, and precuneus. Responses in the frontal cortex and subcortical areas were more selective and mostly limited to social features of high emotional impact (antisocial behavior, sexual & affective behavior, and feeding).

The analysis carefully controlled for low-level audiovisual confounds, but without the possibility of cross-validating the results across different stimuli, complete separation of social perception from the low-level perceptual processes is impossible. However, model comparisons between separate low-level and social models indicated that the social model predicted the BOLD responses more accurately than the low-level model in voxels within STS, LOTC, TPJ, FG, and IFG, while the low-level model was more accurate for predicting the BOLD responses in the visual and auditory cortex (Figure 11). This social perceptual network closely resembles the cortical areas where a social-affective model has been previously shown to explain unique variance not explainable by low-level features (Lee Masson & Isik, 2021). Based on these findings it is unlikely that the results outside primary auditory and visual cortex would be confounded by low-level processing.

These results align with prior univariate studies that attribute social functions to specific brain regions. STS has been repeatedly identified as a central hub for social perception (Deen et al., 2015; Lahnakoski et al., 2012; Nummenmaa & Calder, 2009; Pelphrey et al., 2005; Puce et al., 1996) especially in dynamic contexts (Isik et al., 2017; Lahnakoski et al., 2012; Lee Masson & Isik, 2021; Pitcher, Dilks, et al., 2011). LOTC is known to be involved in the perception of features relevant to social cognition, such as body representation, visual motion, and action processing

(Downing et al., 2001; Lingnau & Downing, 2015). FG plays a critical role in face and body perception (Haxby et al., 2000; Peelen & Downing, 2005). Neural activity in the TPJ has been associated with language processing, the judgment of others' mental states and subsequent decision-making, processing social context, perceived actions, and attention (Bitsch et al., 2018; Carter et al., 2012; Carter & Huettel, 2013; Price, 2012; Saxe & Kanwisher, 2003; Wurm & Schubotz, 2018).

6.2.3 Spatial specificity of the neural representations for social perception

Univariate regression analysis revealed the general topography of cortical areas involved in processing social signals. However, the analysis cannot differentiate whether a brain area whose activity is associated with most social features, such as STS, reflects a general process that is shared across all social perceptual features, (e.g., working memory or object recognition), or specific, context-dependent social processing. Results from the multivariate pattern analysis supported the latter explanation. Unique spatial activation patterns were observed for different social features, as evidenced by high whole-brain classification accuracy (52%, well above the 13% chance level). The whole-brain classification relied on a network of regions including the STS, LOTC, TPJ, FG, and visual areas in the occipital cortex, identified through ANOVA feature selection. This network outperformed any single anatomical region in classification accuracy, emphasizing the distributed nature of social information processing.

Region-specific classification analyses showed that temporal, parietal, and occipital regions exhibited moderate to high classification accuracies, whereas accuracies in frontal and subcortical regions approached the chance level (Figure 10). When classification was restricted to the voxels where the social model outperformed the low-level model in the regression analysis (the warm-colored areas of Figure 11), the accuracy was 35%, which was higher than the accuracy in any single anatomical region but lower compared to the whole-brain classification accuracy. Notably, this analysis excluded most of the occipital cortex, which was the main difference compared to the whole-brain classification. These results suggest that including occipital areas enhances classification accuracy. Although the classification was performed with confound-controlled data, it remains unclear whether this improvement reflects true social context-dependent processing in the visual cortex or residual confounds arising from low-level perceptual differences between social contexts. Higher-level information, such as body parts and actions, are shown to be more strongly associated with BOLD signals than low-level visual features in the occipital cortex outside of V1 (Tarhan & Konkle, 2020) supporting social perceptual processing in the occipital cortex.

The classification results indicate that while regional univariate responses to social features overlap, the spatial activation patterns exhibit feature specificity. Interestingly, voxels selected by ANOVA for discriminating social features in the whole-brain classification closely resemble the network proposed for processing social aspects of human actions (Tarhan & Konkle, 2020), although our findings demonstrated a more bilateral representation. Previous pattern recognition studies have established spatial specificity for individual social features in regions within the proposed social perceptual network. Early work revealed that representations of faces and objects in FG are distributed and overlapping yet feature-specific (Haxby et al., 2001). Subsequent studies have decoded facial expressions from spatial activation patterns in FG and STS (Said et al., 2010; Wegrzyn et al., 2015). STS and TPJ have been shown to represent multiple distinct social perceptual features, such as social versus non-social interaction, cooperation versus competition, mentalizing, and action judgments (Lee Masson et al., 2024; Dmitry Smirnov et al., 2017; Walbrin et al., 2018). LOTC has been shown to represent unique activation patterns for different actions (Tucciarelli et al., 2015; Wurm & Lingnau, 2015) and to distinguish between social and non-social actions (Wurm et al., 2017). However, LOTC representations may primarily reflect perceptual components necessary for interpreting social actions rather than more abstract representations of sociality (Wurm & Caramazza, 2019).

Our pattern recognition results integrate the hubs (STS, LOTC, FG, and TPJ) shown to decode individual aspects of social perception into a common network, suggesting that the abstract social context of uncontrolled dynamic scenes is processes within this distributed network. The findings also advance the field by moving beyond classifying pre-defined social categories from balanced stimuli to decoding data-driven, complex social information from spatial brain activation patterns elicited during dynamic movie viewing.

6.2.4 Neural synchronization during social perception

Intersubject correlation analysis revealed synchronized brain responses across participants in temporal and occipital regions during movie viewing (Figure 9b). Importantly, regional ISC was correlated with the number of social features associated with brain activity in the univariate regression analysis ($r = 0.86$) and classification accuracy in the multivariate pattern analysis ($r = 0.85$). These findings suggest that regions responding to multiple social signals also exhibit time-locked neural activity across participants.

Movies effectively synchronize neural responses across viewers (Hasson et al., 2004), and this effect is stronger for movies with a coherent plot compared to unstructured videos lacking rich socioemotional content and scene transitions

(Hasson et al., 2010). Neural synchronization is also influenced by emotional content. Both valence and arousal dynamically modulate the degree of synchronization while viewing movies (Nummenmaa et al., 2012) and listening to narratives (Nummenmaa, Saarimaki, et al., 2014; D. Smirnov et al., 2019). Consequently, recent work has focused on quantifying the principles of neural synchronization during social interaction (Levy et al., 2021; Nummenmaa et al., 2018). For example, conversation promotes speaker-listener neural coupling, where the degree of coupling predicts communication success (Stephens et al., 2010). Eye contact also synchronizes brain activity between participants, with stronger effects between friends than strangers (Luft et al., 2022). In turn, reduced neural synchronization has been associated with social difficulties in autism spectrum disorders (Quiñones-Camacho et al., 2021; Salmi et al., 2013; Suda et al., 2011).

The present results suggest that social perception synchronizes brain activity across participants within the social perceptual network in temporal and occipital regions. A similar interpretation of synchronization in STS was proposed in a recent study investigating social perception (Lee Masson et al., 2024). Neural synchronization during social perception could indicate "mental resonance", which is critical for the mutual understanding of social environments, also highlighting the centrality of social interaction in human brain function (Hari et al., 2015).

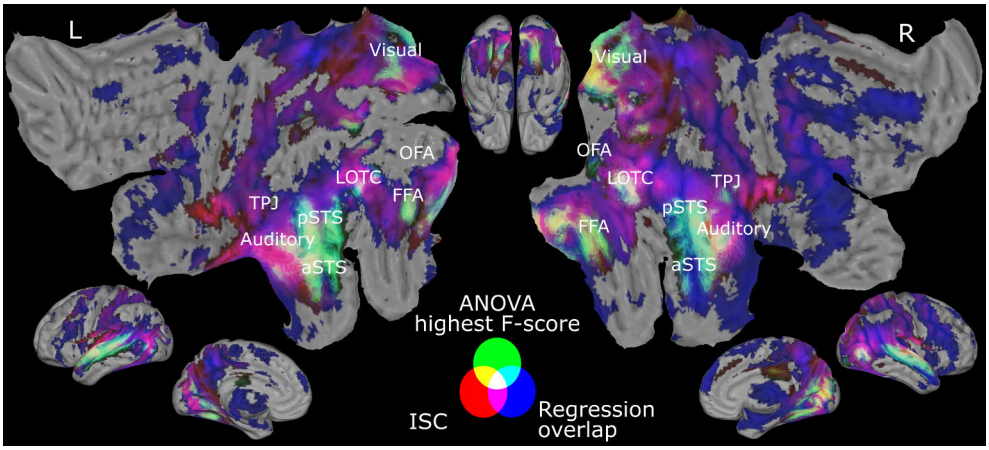

Figure 17. Social perception network in the human brain. The main findings from the three analyses are shown as additive RGB colors. Multiple regression results are shown in blue (regions activated by at least three social features, FDR-corrected, q = 0.05). The voxels included in the whole-brain classification are shown in green (the 3,000 highest F-scores from ANOVA feature selection). Significant ISC (FDR-corrected, q = 0.05) is shown in red. Overlapping colors highlight the most sensitive and specific areas for social perception that also synchronize across participants during movie viewing. Modified from the original publication (Santavirta et al., 2023).

6.2.5 The functional network for social perception

A lateral visual pathway specialized for social perception has been proposed, projecting from early visual cortex via LOTC to STS, with a particular emphasis on processing social information conveyed by moving faces and bodies (Pitcher, Dilks, et al., 2011; Pitcher & Ungerleider, 2021). Previously, the LOTC and STS regions were considered part of the ventral visual pathway, which was first defined for macaques (Kravitz et al., 2013). The ventral visual pathway also includes the occipital face area (OFA) and fusiform face area (FFA) (Pitcher & Ungerleider, 2021; Pitcher, Walsh, et al., 2011). It is argued that the lateral pathway is distinct from the ventral visual pathway because FFA and OFA show a visual field bias for faces, while the face area in pSTS does not (Pitcher & Ungerleider, 2021).

The current findings indicated that areas from both pathways are essential for social perception (Figure 17). LOTC and STS were identified as core regions of the social perceptual network, which aligns with prior studies linking these areas to various aspects of social perception (McMahon & Isik, 2023). The lateral visual pathway appears to follow a hierarchical organization: early visual areas process low-level features, LOTC handles mid-level social primitives and object detection, and STS processes increasingly abstract and communicative social information (McMahon et al., 2023). Although addressing the hierarchical organization of social processing was not the aim of the current research, our results generally support this hierarchy, as activity in the early visual areas was more accurately explained by the low-level model than the social model. On the contrary, neural activity in LOTC and especially in the STS was more accurately predicted by the social perceptual model (Figure 11).

Areas in the ventral pathway also contributed to social perception. Both (right) OFA and FFA in the ventral pathway were part of the most spatially specific whole-brain network for social context classification. This suggests that social perception relies on both pathways, but further research is needed to clarify how these pathways coordinate the social perceptual process together. Additionally, the results indicated that TPJ is involved in social perception. As part of the mentalizing network, TPJ has been consistently found to activate during mental state inference (McMahon & Isik, 2023; Saxe & Kanwisher, 2003). Social perception is tightly linked to inferring how others feel and think, and TPJ may thus have a role in this primary social inference.

Parietal regions, especially precuneus, supramarginal gyrus, and superior parietal gyrus, exhibited consistent responses to multiple social dimensions, though ISC and classification accuracies were only moderate. Previous research has linked precuneus to attention and memory retrieval (Cavanna & Trimble, 2006), and its neural activity also synchronizes across participants during the recollection of shared memories (J. Chen et al., 2017). Neural activity in supramarginal gyrus has been

associated with phonological (Hartwigsen et al., 2010) and visual processing of words (Stoeckel et al., 2009), and activity in superior parietal gyrus has been linked to visuospatial processing and working memory (Koenigs et al., 2009). These findings suggest that parietal regions are involved in general cognitive functions, such as recollection, visuospatial integration, or linguistic processing during social perception.

Frontal and subcortical regions exhibited limited associations with social features, weak spatial specificity, and low neural synchronization. These regions were mainly associated with emotionally charged social features (Antisocial behavior, Sexual & affective behavior, and Feeding), suggesting that they are involved in emotional processing. Limbic regions, including the amygdala and thalamus, have well-documented roles in emotional processing (Hudson et al., 2020; Karjalainen et al., 2018). Medial frontal cortex (MFC) activity is known for its interindividual variability, and MFC likely contributes to attributing affective meaning to ongoing experiences (Chang et al., 2021). Recent findings also indicate that neural activations related to felt but not perceived emotions generalize across stimuli in the MFC and thalamus, underlining their roles in processing felt emotions (Saarimäki et al., 2023).

The frontal cortex is also extensively studied in the context of mentalizing, decision-making, and social cognition (Amodio & Frith, 2006; de la Vega et al., 2016). Recent work has shown that TPJ, STS, and orbitofrontal cortex (OFC) represent others' traits, but only OFC predicts subsequent social decision-making (Kobayashi et al., 2022). Furthermore, previous classification studies have not identified spatially specific responses for social perception in the frontal cortex (Haxby et al., 2001; Oosterhof et al., 2012; Wegrzyn et al., 2015; Wurm & Lingnau, 2015). This suggests that frontal areas may mediate higher-order social processes, such as integrating social perception with abstract cognitive tasks like predicting others' actions, making social decisions, or linking perception to the affective system.

6.3 A taxonomy for social perception

Study I of this thesis examined the low-dimensional organization underlying social perception. Using dimension reduction techniques and generalizability testing on a high-dimensional dataset of social perceptual ratings, we identified a low-dimensional model for social perception.

According to the dynamic interactive theory of person construal individuals form rapid judgments about their social environment by detecting simple social cues (e.g., shared eye contact or mutual smiling) and projecting these rapidly processed cues onto broader situational inferences (e.g., cooperation or attachment) (Freeman &

Ambady, 2011). The theory suggests that social perception emerges as a bidirectional interaction between external sensory inputs and internal higher-order processes, such as emotional states, prior experiences, goals, and motives.

However, the time constraints inherent in social interactions and the computational limitations of the human brain do not allow for the perception of all available social information. We hypothesized that some social information is prioritized, and that this prioritization is captured by a limited set of basic evaluative dimensions.

Based on the findings of Study I, we propose a model in which social environments are evaluated along eight fundamental evaluative dimensions identified through the primary principal coordinate analysis (Figure 18: Social perceptual dimensions). Linear and nonlinear combinations of these basic dimensions can be used to construct a more fine-grained representation of the social environment, as shown by the hierarchical clustering analysis (Figure 18: social semantic categories). This evaluative process is dynamically influenced by external sensory input and individual higher-order processes, which continuously update the perception with new information to guide subsequent inference and action. Importantly, the proposed dimensions represent the population-level average perceptual framework rather than the specific ways in which individuals perceive social situations, acknowledging variability in personal and contextual factors that shape social perception.

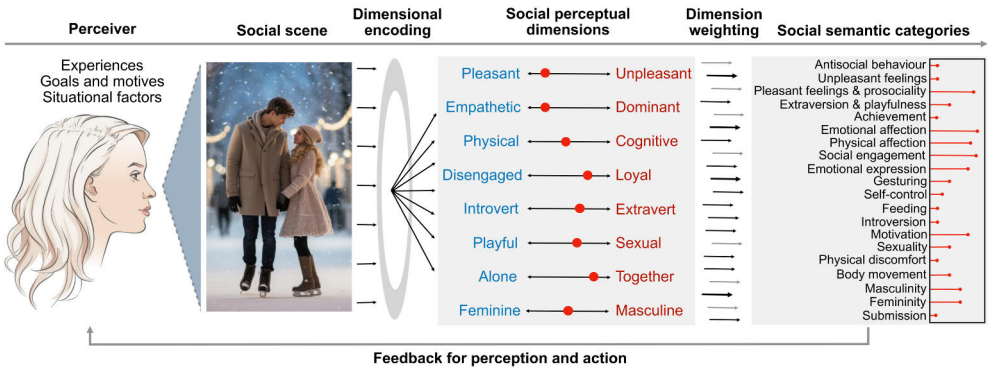

Figure 18. Framework for social perception. The eight basic dimensions of social perception are encoded from low-level information in the social scene. These dimensions are subsequently integrated to establish fine-tuned social semantic categories that may guide action and subsequent perception of the evolving social scene. Reprinted from the original publication. Copyright © 2024 by American Psychological Association. Reproduced with permission (Santavirta, Malén, et al., 2024).

6.3.1 Eight basic dimensions of social perception

The principal coordinate analysis found statistical evidence for eight orthogonal dimensions as a reduced model capable of capturing most of the variation in social perceptual rating data. The first three dimensions, valence, dominance, and cognitive vs. physical functions, explained 55% of the variation, and all eight dimensions explained 78% of the total variation in the perceptual ratings of 136 social features.

The first dimension, *Pleasant-Unpleasant*, explained 32% of the total variation and describes the overall valence of social environments. Valence is a critical evaluative dimension, since pleasant social situations are associated with cooperation and affective behaviors, while unpleasant features typically signal threat or harm, prompting defensive behaviors or avoidance. Valence has been previously identified as a major organizing principle across cognition (Oosterhof & Todorov, 2008; Russell et al., 1989; Zajonc, 1980), which supports its relevance in social perception.

The second dimension, *Empathetic-Dominant,* explained 13% of the total variation and describes social hierarchy and competition. Dominant characteristics include power, agency, ambition, and abuse, while empathetic characteristics encompass intimate and compassionate behaviors towards others. Dominance is an evolutionary strategy to maintain social rank (Maner, 2017) and is characterized as a basic human motivation (Schwartz et al., 2012). Identifying dominant individuals is important because it helps individuals navigate social hierarchies by avoiding conflicts with higher-ranking members or forming coalitions against them.

The third dimension, *Physical–Cognitive,* characterizes human behavior from physical actions to cognitive reasoning. This distinction contrasts fast, impulsive, and reflexive actions with slower, more cognitive, and controlled behaviors. The dimension closely parallels the “Type 1” and “Type 2” processes, or automated versus deliberate reasoning, described in dual-process theories of cognition (Evans & Stanovich, 2013). Both systems are necessary and engaged depending on the situation. Impulsive actions are crucial in situations requiring immediate responses (e.g., whether to fight of flight in a hostile situation), while a slower analytical approach is advantageous in complex situations where rapid actions are unnecessary (e.g., when making a major financial decision). The present findings show that perceiving others’ cognitive approaches in social situations is important for understanding and predicting their actions, subsequently allowing individuals to adapt their own behavior.

In addition to the three primary dimensions explaining > 50% of total variation, five additional dimensions were identified. The *Disengaged–Loyal* dimension describes whether individuals are perceived as actively engaging and contributing to social situations (e.g., conscientious, loyal, brave) or as disengaged and self-focused (e.g., lazy, superficial, selfish) paralleling the conscientiousness trait in personality

theories (Goldberg, 1990; Lee & Ashton, 2004; McCrae & Costa, 1987). The *Introvert–Extravert* dimension captures different social interaction styles, also well-established in personality theories (Goldberg, 1990; Lee & Ashton, 2004; McCrae & Costa, 1987). Prior research has demonstrated that the core valence-dominance model for face perception does not adequately explain judgments of conscientiousness and extraversion (Walker & Vetter, 2016). Additionally, conscientiousness and extraversion are often inferred from the body or require whole-person perception (Hu & O'Toole, 2023). These findings suggest that *Disengaged–Loyal* and *Introvert–Extravert* represent independent dimensions of whole-person perception that may not be accurately captured by static face images, while dynamic video stimuli capture them more accurately.

The *Playful–Sexual* dimension distinguished the perception of playful and friendly characteristics from sexual interactions. Recognizing playfulness and friendly behavior may help people to identify possibilities for friendships. Social laughter, which is often induced by humor, enhances relationships and non-reproductive alliances (Dunbar, 2012; Manninen et al., 2017; Scott et al., 2014). In contrast, perceiving sexual features allows individuals to automatically and rapidly evaluate potential mating partners (Hietanen & Nummenmaa, 2011; Putkinen et al., 2023). Integrating playfulness and sexuality into a single dimension suggests that people may automatically categorize affective interactions as either sexual or non-sexual. However, stereotypical movie content may also lack a more nuanced relationship between sexuality and playfulness.

The *Alone–Together* dimension, in turn, captured whether individuals were alone or interacting with others. This fundamental distinction highlights the social versus solitary nature of human behavior, which is central to understanding social dynamics.

Finally, *Feminine–Masculine* aligned with the perceived (fe)maleness of the individuals. Femininity and masculinity represent a well-defined perceptual axis (Hu et al., 2018; Little & Hancock, 2002), traditionally conceptualized as a single bipolar dimension linked to reproductive traits used to evaluate potential mates (Little et al., 2011; Mitteroecker et al., 2015). Identifying the sex of an interaction partner is crucial for various purposes, ranging from sexual preference and mate competition to the establishment of sex-specific social alliances. *Femininity–Masculinity* aligned well with the youthful/attractive dimension of face perception models (Sutherland et al., 2020). It distinguishes masculinity from dominance (Oosterhof & Todorov, 2008; Sutherland et al., 2013) but suggests that sex characteristics may not be perceived independent of youthfulness and attractiveness (Lin et al., 2021).

The developed permutation testing indicated that the rest of the identified dimensions did not explain more variation than would be expected by chance, suggesting that the unexplained variation primarily consists of irrelevant noise.

Generalizability tests further demonstrated that the identified *eight basic dimensions of social perception* generalize across dynamic movie stimuli and across different data types, including the perception of both videos and images.

6.3.2 How our model relates to existing models for social signals

The first three dimensions, *Pleasant–Unpleasant, Empathetic–Dominant* and *Physical–Cognitive* are supported by multiple models. Osgood's semantic differential, a pioneering model from the 1950s, stated that English words in general can be summarized in three dimensions, valence, potency, and activity (Osgood & Suci, 1955), that resemble, to some extent, the first three social perceptual dimensions. Additionally, the stereotype content model and the parallel dual perspective model of agency and communion are extensively used models for studying group stereotypes, first impressions and social perception (Abele & Wojciszke, 2014; Fiske, 2018). The warmth/communion dimension resembles the *Pleasant–Unpleasant* dimension and competence/agency closely relates to the *Empathetic–Dominant* dimension. Although the first two dimensions align with the previously established warmth and competence dimensions, the currently proposed naming of these two dimensions captures the fine-grained semantics in the context of social perception. Similarly, the first two dimensions valence and dominance have been identified as basic evaluative dimensions of faces (Jones et al., 2021; Morrison et al., 2017; Oosterhof & Todorov, 2008; Sutherland et al., 2013), extending to the perception of bodies (Hu et al., 2018; Morrison et al., 2017; Tzschaschel et al., 2022) and people's voices (McAleer et al., 2014). The present taxonomy is the first to show that these dimensions are elementary evaluative dimensions in dynamic audio–visual social perception, and they likely play a major role in real-life social perception as well.

The linguistic taxonomies of psychological situations provide a useful comparison to the present taxonomy, although they are based on semantic similarities between words rather than actual perceptual ratings. The DIAMONDS taxonomy organizes social situations into eight dimensions: positivity, negativity, adversity, intellect, duty, mating, sociality, and deception (Rauthmann et al., 2014). Similarly, the CAPTION taxonomy identifies seven dimensions: positive valence, negative valence, adversity, complexity, importance, humor, and typicality (Parrigon et al., 2017). Several dimensions from these taxonomies align with the current model of social perception. Valence is a common dimension in both DIAMONDS and CAPTION and corresponds to the present *Pleasant–Unpleasant* dimension. The intellect dimension (DIAMONDS) closely parallels the cognitive side of the current *Physical–Cognitive* dimension, while duty (DIAMONDS) and importance

(CAPTION) dimensions reflect the socially proactive qualities captured in the current *Disengaged–Loyal* dimension. The *Playful–Sexual* dimension relates to mating (DIAMONDS) and humor (CAPTION) dimensions, while sociality (DIAMONDS) aligns with the distinction between social and non-social contexts reflected in the *Alone–Together* dimension. Both taxonomies also describe adversity, albeit with slightly different interpretations. DIAMONDS frames adversity as situations involving conflict, competition, and victimization, whereas CAPTION characterizes it more broadly as depleting situations. Despite these differences, the adversity dimension corresponds to the dominant traits in the *Empathetic–Dominant* dimension of the present taxonomy. To summarize, many of the main components in these lexically derived taxonomies of situations can also explain the perception of dynamic social environments, but they cannot completely cover the complex perceptual and inferential space of rapid social scenes.

The 3d mind model describes variations in mental state inferences in social settings along three dimensions: rationality, social impact and valence (Thornton & Tamir, 2020). The valence dimension is shared with the current taxonomy, while rationality aligns with cognitive behaviors represented in the *Physical–Cognitive* dimension. Social impact, characterized as highly arousing and social states (e.g., lust and dominance) versus low-arousal and non-social states (e.g., fatigue and drowsiness), partially overlaps with the current *Empathetic–Dominant* dimension. However, emotional arousal was more strongly associated with empathetic rather than dominant traits within the *Empathetic–Dominant* dimension suggesting that it is better understood as contrasting "cold" dominant characteristics with empathetic and intimate ones. Meanwhile, the *Alone–Together* dimension captures the distinction between social and non-social situations, diverging from the social impact dimension. Thus, while the 3d mind model offers insights into mental state inferences, it may not fully generalize to social perception, which encompasses broader aspects beyond mental states. The interdependence of mental state and trait inferences (Lin & Thornton, 2023) further emphasizes the need to study social perception as a unified construct.

The current model also diverges from some existing data-driven taxonomies. The ACT-FAST taxonomy for action understanding categorizes complex human actions in six dimensions: abstraction, creation, tradition, food, animacy, and spiritualism (Thornton & Tamir, 2022). While feeding-related features were associated with self-focused end of the current *Disengaged–Loyal* dimension, the ACT-FAST taxonomy focuses on detailed and often abstract actions that are not central to social perception. Similarly, taxonomies of human goals (Wilkowski et al., 2020) and basic values (Schwartz, 2012) are challenging to connect with the current taxonomy, suggesting that perceptions of others' goals and values are likely integrated with other social information rather than perceived as distinct dimensions in social situations.

The *eight basic dimensions of social perception* affirm and refine dimensions of several established models while integrating them into a cohesive framework that is specific to the perception of dynamic social environments. The primary basic dimensions *Pleasant–Unpleasant, Empathetic–Dominant* and *Physical–Cognitive* align with multiple previous models, including circumplex emotion theory, Osgood's semantic differential, warmth/communion, and competence/agency models, 3d mind model, psychological situation models, and face perception models. However, social perception is not limited to these three dimensions. The current model introduces five additional dimensions (*Disengaged–Loyal, Introvert–Extravert, Playful–Sexual, Alone–Together, and Feminine–Masculine*), each supported by some related models. These additional dimensions highlight the complexity of social perception, providing a more comprehensive framework for understanding how humans navigate dynamic social environments.

6.3.3 Fine-grained social information emerges from the basic dimensions

Hierarchical clustering analysis revealed semantically distinct clusters of social features, some of which did not align strictly with any single basic dimension (Figure 4 & 18). However, concordance analysis between the HC clusters and PCoA dimensions demonstrated their convergence (Figure 5). A detailed examination of the associations between clusters and dimensions highlighted how social perceptual clusters emerge as specific combinations of the basic dimensions.

For example, the clusters labeled unpleasant feelings and antisocial behavior were explained as composites of *Pleasant–Unpleasant, Empathetic–Dominant* dimensions. Both clusters included unpleasant features, but unpleasant feelings were associated with empathetic, while antisocial behavior with dominant characteristics revealing their fine-grained distinction. Similarly, variations in dominance structures provided nuanced differentiation among pleasant feature clusters: pleasant feelings and prosociality (solely pleasant), emotional affection (pleasant + no competition), and extraversion and playfulness (pleasant + competition). These findings demonstrate how valence and dominance jointly shape the organization of social features, with both pleasant and unpleasant features forming distinct clusters based on dominance structures.

Additionally, the HC analysis grouped some social features into clusters that integrated information from more than two basic dimensions. For example, distinct communication types were identified in specific clusters labelled as social engagement and emotional expression, gesturing, and physical affection. These clusters indicate how multiple basic dimensions can interact to form compound social categories.

The HC analysis was conducted as an alternative dimension reduction method to validate the usefulness of the eight basic dimensions. It does not force the clusters to be strictly orthogonal, which enabled us to investigate how the clusters can be constructed based on the basic dimensions. The observed convergence with the PCoA dimensions supports the idea that specific perceptual semantic categories can be systematically derived from these basic dimensions. This indicates that *basic eight dimensions of social perception* is a capable model for organizing fine-grained and semantically meaningful social information.

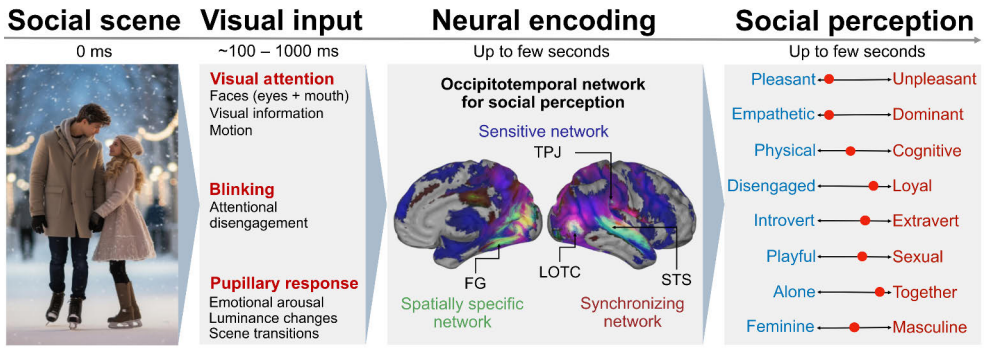

Figure 19. The social perceptual processing cascade. Visual prioritization in sub-second temporal resolution samples the sensory visual information, which is subsequently processed in the social perceptual network in the brain. As a result, humans evaluate social environments along eight basic perceptual dimensions to make inferences about the social situations within seconds. The temporal timeline of processing is abstract and reflects the temporal scales used in the original studies I-III.

6.4 The social perceptual processing cascade

Summary of the social perceptual processing cascade is shown in Figure 19. Social perception begins with extracting sensory information from the environment. The eye-tracking results indicated that the visual system is primarily guided by low-, (e.g., luminance, motion), and mid-level (especially, human eyes and faces) perceptual features. In turn, high-level information has little influence on the dynamic gaze orienting and blinking behavior, while pupillary responses are modulated by high-level emotional arousal. These findings suggest that the human visual system is predominantly controlled by cognitively simple processes, enabling sub-second temporal resolution of visual sampling, which is necessary for rapid perception of social information.

Sensory information is subsequently processed in the human brain to transform pure physical information into complex social inferences. The neuroimaging results indicated that an occipitotemporal network covering both hemispheres is responsible for the processing of social perceptual information. The most important hubs for

social perception are the superior temporal sulcus (STS), lateral occipitotemporal cortex (LOTC), fusiform gyrus (FG), and temporoparietal junction (TPJ). This functional brain network is activated during social perception, and the spatial activation patterns of these regions are context-specific, suggesting that abstract concepts of complex social situations are processed within the network.

Finally, social inferences about others and social interactions emerge as a result. The perceptual findings revealed that people rapidly evaluate social situations along eight basic dimensions. Evaluating social environments along this limited set of basic dimensions simplifies cognitive processing, which may explain why people can perceive complex social information fast enough to make useful predictions of others' behavior. This entire social perceptual cascade unfolds within seconds, allowing swift reactions in the ever-changing social environments.

6.5 Limitations and future directions

This research has some notable limitations and opens directions for future research. The studies employed naturalistic movie stimuli as a proxy for real-life social situations providing a better approximation of real social situations than more often-used static images. Movies may portray stereotypical or amplified versions of social contexts that do not fully align with everyday reality, which is both a limitation and a strength. Certain important aspects of social interactions, such as violence or sexual behavior, are not ethical to study in the wild, and fully natural stimuli may lack the needed variability for reliable investigation. Nevertheless, future research should strive for even more realistic stimulus models. For example, wearable cameras and eye-trackers could be employed to collect data during participants' daily lives, capturing scenarios where participants actively engage in social interactions rather than passively observe them.

The proposed model for social perception is based on ratings of 138 social features that were preselected based on prior relevant taxonomies (see 4.3. Social perceptual features within the Methods section). The identified dimensionality is influenced by the selection of these features, which also impacts the estimated importance (variance explained) of the PCoA dimensions. The current feature set was able to uncover a generalizable structure for social perception, but it cannot capture all imaginable social situations. Hence, some social perceptual dimensions may not have been established yet.

An alternative approach to feature selection is letting participants define the features (Koch et al., 2016; Nicolas et al., 2022; Osgood & Suci, 1955), which would minimize researcher-dependent choices. However, this method has its own limitations. Social perception likely involves unconscious prioritization of information and participants without expertise may overlook important but

unconsciously perceived information. While this approach would provide new insights into the stability of social perceptual dimensions, it does not fully resolve the challenges of feature selection.

Regardless of how the features are selected, there is an inherent trade-off between the quantity of data and the labor required, which is a significant bottleneck in collecting high-dimensional datasets. Expanding the feature or stimulus sets necessitates increased participant effort. For example, approximately 1,100 hours of participant labor was required for data collection in Study I. One intriguing way to reduce human labor would be to simulate or augment human responses using large language models (LLMs). This is currently under extensive investigation, and preliminary evidence is promising (Demszky et al., 2023). LLMs may even prove capable of perceiving complex social information from dynamic stimuli (Santavirta, Wu, et al., 2024). Future research should investigate the LLMs further in annotating social or emotional features and for aiding in stimulus and feature selection.

The audio-visual features used to control the neural and eye-tracking analyses were also researcher-defined, raising the possibility that some critical low-level information may have been overlooked. Cognitive neuroscience increasingly relies on visual deep learning models to extract the low-level features (Kriegeskorte, 2015). This approach enables extracting low-level information without conscious feature selection, but it introduces new challenges. First, interpreting how specific low-level information influences the results is challenging, since it is difficult to know which information different layers of the visual models convey without comparison with manually extracted features. Second, identifying the border between low-level feature processing and social information representation within deep visual models is challenging. In deeper layers, models may already represent social semantic categories, such as faces, or even abstract social information. This could lead to false negative findings for social perception if such layers are erroneously interpreted as representing purely low-level information. Despite these challenges, one possibility would be to train a deep visual model to specifically classify abstract social dimensions from naturalistic stimuli. Then modeling neural activity with layers representing different information (low-level information, social primitives, more abstract social information) could potentially increase the understanding of hierarchical processing in the human brain.

The present results focus on social perception occurring over short timescales of up to ten seconds. While social perception can occur in mere hundreds of milliseconds (Dima et al., 2022; Isik et al., 2020; Nummenmaa et al., 2010; Willis & Todorov, 2006), other types of social inference may require extended time periods for accurate evaluation, such as assessing another person's trustworthiness. Investigating how social inferential evaluations evolve over varying timescales, from

milliseconds to minutes, hours, or even years, would provide a more comprehensive understanding of the temporal dynamics of social perception and inference.

Finally, this thesis focuses on population-average social perception, providing a reference for future studies. However, social perceptual ratings show notable inter-individual differences, calling for future research on the participant-specific attributes that drive these differences. For example, similar methodologies could be used for studying detailed alterations of social perception in neuropsychiatric conditions, such as autism spectrum disorders or depression. This information could provide novel insights into the characteristics of these disorders, potentially advancing the development of diagnostic tools and treatments. Ultimately, future studies should aim for modeling individual social perceptual evaluations with a rich array of participant-specific attributes. This would increase our understanding of individual-level social perception, paving the way for more accurate predictions of human behavior.

7 Conclusions

This thesis aimed to investigate social perception in dynamic social environments. The whole cognitive processing stream, from the audio-visual input, through neural processing, to the perceptual inference, was investigated in three separate studies. The main findings of this thesis were as follows:

I. Social perception follows a limited set of evaluative dimensions, which enables people to rapidly infer and react in dynamic social environments. Based on the findings, we propose *eight basic dimensions of social perception* as a detailed model for social perception.

II. Social perceptual information is processed in a wide occipitotemporal network spanning both hemispheres of the human brain. The most important hubs for social perception include superior temporal sulcus (STS), lateral occipitotemporal cortex (LOTC), fusiform gyrus (FG), and temporoparietal junction (TPJ). This brain network is activated during social perception, and the spatial activation patterns of these regions are specific for the perceived social context.

III. The human visual system is primarily guided by simple stimulus features, including low-level audio-visual information (e.g., luminance and motion) and mid-level information (especially human eyes and faces). While emotional arousal modulates the pupillary response, high-level social information does not effectively predict dynamic gaze patterns or blinking behavior.

Acknowledgments

I am profoundly grateful to so many wonderful people around me. Without your support the completion, or even the start, of this thesis would not have been possible.

First and foremost, I am deeply thankful to my incredible supervisors Lauri Nummenmaa and Enrico Glerean. Lauri, thank you for being always available for me with clear guidance on how to move forward. Your extraordinary ability to filter out unnecessary noise and focus on what truly matters has been invaluable. Your enthusiasm for science is contagious, and your excitement has been an important motivator throughout the years. I could not have wished for a better supervisor — Thank you, truly. Enrico, I am especially grateful for your outstanding technical support. In most statistical programming tasks, I have relied on your code or expertise in some way. Although I mainly worked with Lauri on a day-to-day basis, I have always known that I could turn to you whenever I need help.

This thesis would not have been possible without the financial support that allowed me to work as a full-time doctoral researcher between 2022 and 2025. I was personally funded by the Doctoral Programme in Clinical Research at the University of Turku, Governmental Research Funding for Turku University Hospital and for the Western Finland collaborative area, Turku University Foundation and Alfred Kordelin Foundation. Thank you for believing in me.

Tomi Karjalainen, thank you for introducing me to the fascinating world of statistical programming. You have a remarkable talent for explaining complex matters in such a simple way that even I (with next to zero statistical and programming experience in 2016) felt like I understood something. I could always trust your advice — if you had made an analytical decision, I knew there was always a well-thought-out reason behind it.

Tuulia Malén, thank you for our countless conversations about statistics, uncertainties, and life in general. I fondly recall our (sometimes even heated) debates at the *Friends of Linear Associations*. It has been a privilege to learn all these exciting things about mixed-effects modelling and Bayesian statistics — topics I might never have explored had I only focused on my own research.

This thesis is powered by extremely high-quality data. Massive kudos to everyone involved in data collection and designing of the experiments. Without

access to such great datasets, I would not have been able to focus on developing the appropriate analytical methods for these highly data-driven studies. A special thank you to my co-authors (in no particular order): Vesa Putkinen, Kerttu Seppälä, Matthew Hudson, Sanaz Nazari-Farsani, Lihua Sun, Jussi Hirvonen, Henry Karlsson, Birgitta Paranko, Asli Erdemli and Jukka Hyönä, as well as all the research assistants who worked hard to conduct the fMRI and eye-tracking experiments. With your help, this thesis summarizes results from data collected from 2,519 volunteers. I sincerely thank each and every volunteer who participated in these studies.

It has been exciting to witness how The Human Emotions Systems Laboratory (Emotion Lab) has expanded into a buzzing hub of excellent scientists working on multidisciplinary projects during my years (2016 - 2025) in the lab. There has always been someone to consult about any problem, making me feel that I was never alone when facing obstacles. More importantly, you are such wonderful people, and I feel privileged to have got to know you also on a personal level. In addition to those already mentioned I have had the pleasure of working with Jouni Tuisku, Janne Isojärvi, Tatu Kantonen, Jinglu Chen, Santeri Palonen, Harri Harju, Heidi Laine, Rui Watanabe, Sandra Manninen, Tiina Saanijoki, Tuomo Noppari, Lasse Lukkarinen, Kyoungjune "Arto" Pak, Lili Järvinen, Yuhang Wu, Juha Lahnakoski, Heini Saarimäki, Laura Pekkarinen, Timo Heikkilä, Marco Bucci, Jarkko Johansson, Tuomas Knuuti, and Lauri Suominen. A warm welcome as well to Qingying Ye, Ksenia Egorova, Maya Rassouli, Zainab Yusuf, Petri Kaurola and Taru Garthwaite, our newest bright minds starting their (post-)doctoral journeys in the Emotion Lab.

Thank you, director Juhani Knuuti, and everyone at the Turku PET Centre for creating such a warm and enthusiastic atmosphere for pursuing doctoral studies. Thank you, Minna Kangasperko and Lenita Saloranta, for your help with administrative matters and thank you Rami Mikkola, Marko Tättäläinen, Jani Lehtilä, and Timo Laitinen for your help with IT-related issues.

I also want to express my gratitude to Valtteri Kaasinen for the opportunity to conduct doctoral research within Clinical Neurosciences at the University of Turku. Thank you, Juha Rinne and Juho Joutsa, for your support as members of my doctoral follow-up group. Lastly, I am grateful to Angelika Lingnau and Mikko Peltola for accepting to review this thesis. Your constructive feedback significantly improved its quality.

Pursuing and succeeding in a PhD is not just about the people who you work with. It is equally about the people that matter to you outside the lab. They are the ones who let you be yourself, who share laughter and experiences, and who support you unconditionally, even when you stumble.

Thank you to my lifelong friends Artturi Hannula and Jesse Salonen and to the whole *Olutpiakomitea*: Artturi Hannula, Jesse Salonen, Eetu Nyman, Jan Reipas, and Jussi Klimoff. I could not have asked for a better team to grow up with. I am

incredibly happy that we still stick together and that our original group has evolved into an even larger circle of amazing people.

I am also deeply grateful for the friendships formed in medical school. Thank you, Santeri Lehtonen, Petri Peng, Samuel Söderqvist, and Kasperi Alakylä, for your companionship during numerous adventures — whether far away (California, Florida, Malta, the Alps, and Japan) or closer to home (the Turku archipelago, Himos, Tervalampi, Juankoski, and Kauhajoki). Santeri, I am honored to be the godfather of your son, Verneri.

I also want to thank Julia Salo, Johanna Kevo, Tommi Heikkinen, Sofia Kanasuo, Jaana Mella and Antti Raaska, as well as the other members of *Ravut*. I always look forward to seeing you and hope that our triannual gatherings continue. Salute to all my fellow skiers, our shared interest for skiing shows how friendships can evolve, and new groups can emerge. Thank you, Lauri Mattila and Arttu Laisi, for sharing my passion for fly fishing, and to all the members of *KuuLa* for the years of sweating on the football field. Here I describe the important people in my life through shared activities and interests, but what truly matters is the unique connections and memories I have made with each of you. Thank you to all my friends — both mentioned and unmentioned — for being part of my life.

Sirkku Jyrkkiö, thank you for being my godmother. You took me under your wings during the intense weeks of preparing for the medical school entrance exam. Without your incredible support at that crucial time, this thesis would have never been written. Your mentorship throughout medical school and my doctoral studies has meant a lot to me. I always look forward to our irregular but meaningful meetings.

Lastly, I am most grateful for the unconditional love and support of my family. On top of everything else, thank you for always being so curious about my PhD journey and well-being. It has been incredibly freeing to share both the highs and lows with you.

Kiitos Mummu ja Mumma. Kiitos Isä, Kiitos Äiti. I love you with all my heart.

Turku, April 2025

Severi Santavirta

References

Abele, A. E., & Wojciszke, B. (2014). Chapter Four - Communal and Agentic Content in Social Cognition: A Dual Perspective Model. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 50, pp. 195–255). Academic Press. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, *14*(5), 427–432. https://doi.org/10.1111/1467-9280.01458

Adolphs, R., Nummenmaa, L., Todorov, A., & Haxby, J. V. (2016). Data-driven approaches in the investigation of social perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1693). https://doi.org/10.1098/rstb.2015.0367

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews. Neuroscience*, *7*(4), 268–277. https://doi.org/10.1038/nrn1884

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. G. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology*, *12*, 702538. https://doi.org/10.3389/fpsyg.2021.702538

Babiker, A., Faye, I., & Malik, A. (2013). Pupillary behavior in positive and negative emotions. *2013 IEEE International Conference on Signal and Image Processing Applications*, 379–383. https://doi.org/10.1109/ICSIPA.2013.6708037

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., & Hyde, J. S. (1992). Time course EPI of human brain function during task activation. *Magnetic Resonance in Medicine*, *25*(2), 390–397. https://doi.org/10.1002/mrm.1910250220

Bellitto, G., Proietto Salanitri, F., Palazzo, S., Rundo, F., Giordano, D., & Spampinato, C. (2021). Hierarchical Domain-Adapted Feature Learning for Video Saliency Prediction. *International Journal of Computer Vision*, *129*(12), 3216–3232. https://doi.org/10.1007/s11263-021-01519-y

Bindemann, M., Burton, A. M., Hooge, I. T. C., Jenkins, R., & de Haan, E. H. F. (2005). Faces retain attention. *Psychonomic Bulletin & Review*, *12*(6), 1048–1053. https://doi.org/10.3758/BF03206442

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition*, *16*(2–3), 341–355. https://doi.org/10.1080/13506280701434532

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research*, *49*(24), 2992–3000. https://doi.org/10.1016/j.visres.2009.09.014

Bitsch, F., Berger, P., Nagels, A., Falkenberg, I., & Straube, B. (2018). The role of the right temporo-parietal junction in social decision-making. *Human Brain Mapping*, *39*(7), 3072–3085. https://doi.org/10.1002/hbm.24061

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Brooks, J. A., Stolier, R. M., & Freeman, J. B. (2020). Computational approaches to the neuroscience of social perception. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsaa127

Brosch, T., Bar-David, E., & Phelps, E. A. (2013). Implicit race bias decreases the similarity of neural representations of black and white faces. *Psychological Science*, *24*(2), 160–166. https://doi.org/10.1177/0956797612451465

Bruckert, A., Christie, M., & Le Meur, O. (2023). Where to look at the movies: Analyzing visual attention to understand movie editing. *Behavior Research Methods*, *55*(6), 2940–2959. https://doi.org/10.3758/s13428-022-01949-7

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: salient physical features guide effective visual search. *Journal of Experimental Psychology. General*, *137*(3), 471–494. https://doi.org/10.1037/a0012771

Carter, R. M., Bowling, D. L., Reeck, C., & Huettel, S. A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science (New York, N.Y.)*, *337*(6090), 109–111. https://doi.org/10.1126/science.1219681

Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal-parietal junction. *Trends in Cognitive Sciences*, *17*(7), 328–336. https://doi.org/10.1016/j.tics.2013.05.007

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain: A Journal of Neurology*, *129*(Pt 3), 564–583. https://doi.org/10.1093/brain/awl004

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

Chen, G., Taylor, P. A., & Cox, R. W. (2017). Is the statistic value all we should care about in neuroimaging? *NeuroImage*, *147*, 952–959. https://doi.org/10.1016/j.neuroimage.2016.09.066

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125. https://doi.org/10.1038/nn.4450

Cherng, Y.-G., Baird, T., Chen, J.-T., & Wang, C.-A. (2020). Background luminance effects on pupil size associated with emotion and saccade preparation. *Scientific Reports*, *10*(1), 15718. https://doi.org/10.1038/s41598-020-72954-z

Chiu, D. S., & Talhouk, A. (2018). diceR: an R package for class discovery using an ensemble driven approach. *BMC Bioinformatics*, *19*(1), 11. https://doi.org/10.1186/s12859-017-1996-y

Cornia, M., Baraldi, L., Serra, G., & Cucchiara, R. (2018). Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*. https://doi.org/10.1109/TIP.2018.2851672

Cremers, H. R., Wager, T. D., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PloS One*, *12*(11), e0184923. https://doi.org/10.1371/journal.pone.0184923

Cukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770. https://doi.org/10.1038/nn.3381

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., Alexander, A. L., & Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, *8*(4), 519–526. https://doi.org/10.1038/nn1421

de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D., & Yarkoni, T. (2016). Large-Scale Meta-Analysis of Human Medial Frontal Cortex Reveals Tripartite Functional Organization. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(24), 6553–6562. https://doi.org/10.1523/JNEUROSCI.4402-15.2016

Dean, R. T., Bailes, F., & Schubert, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music: an experimental investigation. *PloS One*, *6*(4), e18591. https://doi.org/10.1371/journal.pone.0018591

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus. *Cerebral Cortex* , *25*(11), 4596–4609. https://doi.org/10.1093/cercor/bhv111

Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., Eichstaedt, J. C., Hecht, C., Jamieson, J., Johnson, M., Jones, M., Krettek-Cobb, D., Lai, L., JonesMitchell, N., Ong, D. C., Dweck, C. S., Gross, J. J., & Pennebaker, J. W. (2023). Using large language models in psychology. *Nature Reviews Psychology*, *2*(11), 688–701. https://doi.org/10.1038/s44159-023-00241-5

Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5203–5212. https://doi.org/10.1109/cvpr42600.2020.00525

Di Stefano, N., & Spence, C. (2022). Roughness perception: A multisensory/crossmodal perspective. *Attention, Perception & Psychophysics*, *84*(7), 2087–2114. https://doi.org/10.3758/s13414-022-02550-y

Dima, D. C., Tomita, T. M., Honey, C. J., & Isik, L. (2022). Social-affective features drive human representations of observed actions. *ELife*, *11*. https://doi.org/10.7554/eLife.75027

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28. https://doi.org/10.1167/10.10.28

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Dunbar, R. I. M. (2012). Bridging the bonding gap: the transition from primates to humans. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1597), 1837–1846. https://doi.org/10.1098/rstb.2011.0217

End, A., & Gamer, M. (2017). Preferential Processing of Social Features and Their Interplay with Physical Saliency in Complex Naturalistic Scenes. *Frontiers in Psychology*, *8*, 418. https://doi.org/10.3389/fpsyg.2017.00418

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. 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, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience*, *8*(12), 1643–1646. https://doi.org/10.1038/nn1608

ffmpeg. (2025). *Select function*. https://ffmpeg.org/ffmpeg-filters.html#select_002c-aselect

Fiske, S. T. (2018). Stereotype Content: Warmth and Competence Endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*(4), 571–583. https://doi.org/10.1068/p5705

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

Franchak, J. M., Heeger, D. J., Hasson, U., & Adolph, K. E. (2016). Free Viewing Gaze Behavior in Infants and Adults. In *Infancy* (Vol. 21, Issue 3, pp. 262–287). https://doi.org/10.1111/infa.12119

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–279. https://doi.org/10.1037/a0022327

Freeman, J. B., Johnson, K. L., Adams, R. B., Jr, & Ambady, N. (2012). The social-sensory interface: category interactions in person perception. *Frontiers in Integrative Neuroscience*, *6*, 81. https://doi.org/10.3389/fnint.2012.00081

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, *40*(1), 21–34. https://doi.org/10.1016/j.jrp.2005.08.003

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Gorilla. (2024). *Gorilla*. https://gorilla.sc/

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, *53*(3–4), 325–338. https://doi.org/10.1093/biomet/53.3-4.325

Grillon, C., & Baas, J. (2003). A review of the modulation of the startle reflex by affective states and its application in psychiatry. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *114*(9), 1557–1579. https://doi.org/10.1016/S1388-2457(03)00202-5

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. *Neuroinformatics*, *7*(1), 37–53. https://doi.org/10.1007/s12021-008-9041-y

Hansen, J. Y., Shafiei, G., Markello, R. D., Smart, K., Cox, S. M. L., Nørgaard, M., Beliveau, V., Wu, Y., Gallezot, J.-D., Aumont, É., Servaes, S., Scala, S. G., DuBois, J. M., Wainstein, G., Bezgin, G., Funck, T., Schmitz, T. W., Spreng, R. N., Galovic, M., … Misic, B. (2022). Mapping neurotransmitter systems to the structural and functional organization of the human neocortex. *Nature Neuroscience*. https://doi.org/10.1038/s41593-022-01186-3

Hari, R., Henriksson, L., Malinen, S., & Parkkonen, L. (2015). Centrality of social interaction in human brain function. *Neuron*, *88*(1), 181–193. https://doi.org/10.1016/j.neuron.2015.09.022

Harry, B., Williams, M. A., Davis, C., & Kim, J. (2013). Emotional expressions evoke a differential response in the fusiform face area. *Frontiers in Human Neuroscience*, *7*, 692. https://doi.org/10.3389/fnhum.2013.00692

Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., & Siebner, H. R. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(38), 16494–16499. https://doi.org/10.1073/pnas.1008121107

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*(5664), 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. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(10), 2539–2550. https://doi.org/10.1523/JNEUROSCI.5487-07.2008

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/s1364-6613(00)01482-0

Hayes, T. R., & Henderson, J. M. (2021). Deep saliency models learn low-, mid-, and high-level features to predict scene attention. *Scientific Reports*, *11*(1), 18434. https://doi.org/10.1038/s41598-021-97879-z

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. https://doi.org/10.1016/j.tics.2003.09.006

Henderson, R. R., Bradley, M. M., & Lang, P. J. (2018). Emotional imagery and pupil diameter. *Psychophysiology*, *55*(6), e13050. https://doi.org/10.1111/psyp.13050

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*(3423), 349–350. https://doi.org/10.1126/science.132.3423.349

Heurling, K., Leuzy, A., Jonasson, M., Frick, A., Zimmer, E. R., Nordberg, A., & Lubberink, M. (2017). Quantitative positron emission tomography in brain research. *Brain Research*, *1670*, 220–234. https://doi.org/10.1016/j.brainres.2017.06.022

Hietanen, J. K., & Nummenmaa, L. (2011). The naked truth: the face and body sensitive N170 response is enhanced for nude bodies. *PloS One*, *6*(11), e24408. https://doi.org/10.1371/journal.pone.0024408

Hillman, E. M. C. (2014). Coupling mechanism and significance of the BOLD signal: a status report. *Annual Review of Neuroscience*, *37*(1), 161–181. https://doi.org/10.1146/annurev-neuro-071013-014111

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression - Biased Estimation for Nonorthogonal Problems. *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, *12*(1), 55-. https://doi.org/Doi 10.1080/00401706.1970.10488634

Horstmann, K. T., Rauthmann, J. F., Sherman, R. A., & Ziegler, M. (2021). Unveiling an exclusive link: Predicting behavior with personality, situation perception, and affect in a preregistered experience sampling study. *Journal of Personality and Social Psychology*, *120*(5), 1317–1343. https://doi.org/10.1037/pspp0000357

Hu, Y., & O'Toole, A. J. (2023). First impressions: Integrating faces and bodies in personality trait perception. *Cognition*, *231*, 105309. https://doi.org/10.1016/j.cognition.2022.105309

Hu, Y., Parde, C. J., Hill, M. Q., Mahmood, N., & O'Toole, A. J. (2018). First Impressions of Personality Traits From Body Shapes. *Psychological Science*, *29*(12), 1969–1983. https://doi.org/10.1177/0956797618799300

Hudson, M., Seppala, K., Putkinen, V., Sun, L., Glerean, E., Karjalainen, T., Karlsson, H. K., Hirvonen, J., & Nummenmaa, L. (2020). Dissociable neural systems for unconditioned acute and sustained fear. *NeuroImage*, 116522. https://doi.org/10.1016/j.neuroimage.2020.116522

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014

Hyönä, J., Tommola, J., & Alaja, A. M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *48*(3), 598–612. https://doi.org/10.1080/14640749508401407

Ilie, G., & Thompson, W. F. (2006). A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect. *Music Perception*, *23*(4), 319–329. https://doi.org/10.1525/mp.2006.23.4.319

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(43), E9145–E9152. https://doi.org/10.1073/pnas.1714471114

Isik, L., Mynick, A., Pantazis, D., & Kanwisher, N. (2020). The speed of human social interaction perception. *NeuroImage*, *215*(116844), 116844. https://doi.org/10.1016/j.neuroimage.2020.116844

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506. https://doi.org/10.1016/s0042-6989(99)00163-7

Jackson, A. F., & Bolger, D. J. (2014). The neurophysiological bases of EEG and EEG measurement: a review for the rest of us. *Psychophysiology*, *51*(11), 1061–1071. https://doi.org/10.1111/psyp.12283

Jain, S., Yarlagadda, P., Jyoti, S., Karthik, S., Subramanian, R., & Gandhi, V. (2021). ViNet: Pushing the limits of Visual Modality for Audio-Visual Saliency Prediction. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3520–3527. https://doi.org/10.1109/IROS51168.2021.9635989

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxsom, N. G., Lewis, S. C., Foroni, F., Willis, M. L., Cubillas, C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., … Coles, N. A. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, *5*(1), 159–169. https://doi.org/10.1038/s41562-020-01007-2

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*, *89*(1), 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Karjalainen, T., Karlsson, H. K., Lahnakoski, J. M., Glerean, E., Nuutila, P., Jaaskelainen, I. P., Hari, R., Sams, M., & Nummenmaa, L. (2017). Dissociable Roles of Cerebral mu-Opioid and Type 2 Dopamine Receptors in Vicarious Pain: A Combined PET-fMRI Study. *Cerebral Cortex* , *27*(8), 4257–4266. https://doi.org/10.1093/cercor/bhx129

Karjalainen, T., Seppala, K., Glerean, E., Karlsson, H. K., Lahnakoski, J. M., Nuutila, P., Jaaskelainen, I. P., Hari, R., Sams, M., & Nummenmaa, L. (2018). Opioidergic Regulation of Emotional Arousal: A Combined PET-fMRI Study. *Cerebral Cortex* . https://doi.org/10.1093/cercor/bhy281

Kauppi, J. P., Pajula, J., & Tohka, J. (2014). A versatile software package for inter-subject correlation based analyses of fMRI. *Frontiers in Neuroinformatics*, *8*, 2. https://doi.org/10.3389/fninf.2014.00002

Kawai, S., Takano, H., & Nakamura, K. (2013). Pupil Diameter Variation in Positive and Negative Emotions with Visual Stimulus. *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 4179–4183. https://doi.org/10.1109/SMC.2013.712

Keilholz, S. D., Pan, W.-J., Billings, J., Nezafati, M., & Shakil, S. (2017). Noise and non-neuronal contributions to the BOLD signal: applications to and insights from animal studies. *NeuroImage*, *154*, 267–281. https://doi.org/10.1016/j.neuroimage.2016.12.019

Keles, U., Kliemann, D., Byrge, L., Saarimäki, H., Paul, L. K., Kennedy, D. P., & Adolphs, R. (2022). Atypical gaze patterns in autistic adults are heterogeneous across but reliable within individuals. *Molecular Autism*, *13*(1), 39. https://doi.org/10.1186/s13229-022-00517-2

Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends in Cognitive Sciences*, *16*(11), 559–572. https://doi.org/10.1016/j.tics.2012.09.006

Kimura, E., Abe, S., & Goryo, K. (2014). Attenuation of the pupillary response to luminance and color changes during interocular suppression. *Journal of Vision*, *14*(5), 14. https://doi.org/10.1167/14.5.14

Kirillov, A., Girshick, R., He, K., & Dollar, P. (2019). Panoptic Feature Pyramid Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6399–6408. https://doi.org/10.1109/cvpr.2019.00656

Klaib, A. F., Alsrehin, N. O., Melhem, W. Y., Bashtawi, H. O., & Magableh, A. A. (2021). Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies. *Expert Systems with Applications*, *166*(114037), 114037. https://doi.org/10.1016/j.eswa.2020.114037

Kobayashi, K., Kable, J. W., Hsu, M., & Jenkins, A. C. (2022). Neural representations of others' traits predict social decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(22), e2116944119. https://doi.org/10.1073/pnas.2116944119

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative-progressive beliefs, and communion. *Journal of Personality and Social Psychology*, *110*(5), 675–709. https://doi.org/10.1037/pspa0000046

Koenigs, M., Barbey, A. K., Postle, B. R., & Grafman, J. (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(47), 14980–14986. https://doi.org/10.1523/jneurosci.3706-09.2009

Koide-Majima, N., Nakai, T., & Nishimoto, S. (2020). Distinct dimensions of emotion in the human brain and their representation on the cortical surface. *NeuroImage*, *222*, 117258. https://doi.org/10.1016/j.neuroimage.2020.117258

Kojovic, N., Ben Hadid, L., Franchini, M., & Schaer, M. (2019). Sensory processing issues and their association with social difficulties in children with Autism Spectrum Disorders. *Journal of Clinical Medicine*, *8*(10), 1508. https://doi.org/10.3390/jcm8101508

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spatial Vision*, *13*(2–3), 201–214. https://doi.org/10.1163/156856800741216

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, *1*, 417–446. https://doi.org/10.1146/annurev-vision-082114-035447

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., & Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(12), 5675–5679. https://doi.org/10.1073/pnas.89.12.5675

Lahnakoski, J. M., Glerean, E., Jaaskelainen, I. P., Hyona, 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., Jaaskelainen, 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*, 233. https://doi.org/10.3389/fnhum.2012.00233

Lee, K., & Ashton, M. C. (2004). Psychometric Properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Lee Masson, H., Chang, L., & Isik, L. (2024). Multidimensional neural representations of social features during movie viewing. *Social Cognitive and Affective Neuroscience*, *19*(1). https://doi.org/10.1093/scan/nsae030

Lee Masson, H., & Isik, L. (2021). Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. *NeuroImage*, *245*, 118741. https://doi.org/10.1016/j.neuroimage.2021.118741

Lettieri, G., Handjaras, G., Ricciardi, E., Leo, A., Papale, P., Betta, M., Pietrini, P., & Cecchetti, L. (2019). Emotionotopy in the human right temporo-parietal cortex. *Nature Communications*, *10*(1), 5568. https://doi.org/10.1038/s41467-019-13599-z

Levy, J., Lankinen, K., Hakonen, M., & Feldman, R. (2021). The integration of social and neural synchrony: a case for ecologically valid research using MEG neuroimaging. *Social Cognitive and Affective Neuroscience*, *16*(1–2), 143–152. https://doi.org/10.1093/scan/nsaa061

Li, S., Jamadar, S. D., Ward, P. G. D., Premaratne, M., Egan, G. F., & Chen, Z. (2020). Analysis of continuous infusion functional PET (fPET) in the human brain. *NeuroImage*, *213*(116720), 116720. https://doi.org/10.1016/j.neuroimage.2020.116720

Liao, H.-I., Kashino, M., & Shimojo, S. (2021). Attractiveness in the eyes: A possibility of positive loop between transient pupil constriction and facial attraction. *Journal of Cognitive Neuroscience*, *33*(2), 315–340. https://doi.org/10.1162/jocn_a_01649

Lin, C., Keles, U., & Adolphs, R. (2021). Four dimensions characterize attributions from faces using a representative set of English trait words. *Nature Communications*, *12*(1), 5168. https://doi.org/10.1038/s41467-021-25500-y

Lin, C., & Thornton, M. (2023). Evidence for bidirectional causation between trait and mental state inferences. *Journal of Experimental Social Psychology*, *108*, 104495. https://doi.org/10.1016/j.jesp.2023.104495

Lindquist, M. A., Meng Loh, J., Atlas, L. Y., & Wager, T. D. (2009). Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. *NeuroImage*, *45*(1 Suppl), S187-98. https://doi.org/10.1016/j.neuroimage.2008.10.065

Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*, *19*(5), 268–277. https://doi.org/10.1016/j.tics.2015.03.006

Little, A. C., & Hancock, P. J. B. (2002). The role of masculinity and distinctiveness in judgments of human male facial attractiveness. *British Journal of Psychology* , *93*(Pt 4), 451–464. https://doi.org/10.1348/000712602761381349

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1571), 1638–1659. https://doi.org/10.1098/rstb.2010.0404

Liu, W.-H., Huang, J., Wang, L.-Z., Gong, Q.-Y., & Chan, R. C. K. (2012). Facial perception bias in patients with major depression. *Psychiatry Research*, *197*(3), 217–220. https://doi.org/10.1016/j.psychres.2011.09.021

Lou, J., Lin, H., Marshall, D., Saupe, D., & Liu, H. (2022). TranSalNet: Towards perceptually relevant visual saliency prediction. *Neurocomputing*, *494*, 455–467. https://doi.org/10.1016/j.neucom.2022.04.080

Louhimies, A. (2008). *Käsky*. Helsinki-Filmi, Two Thirty Five, Mogador Film.

Luft, C. D. B., Zioga, I., Giannopoulos, A., Di Bona, G., Binetti, N., Civilini, A., Latora, V., & Mareschal, I. (2022). Social synchronization of brain activity increases during eye-contact. *Communications Biology*, *5*(1), 412. https://doi.org/10.1038/s42003-022-03352-6

Maffei, A., & Angrilli, A. (2019). Spontaneous blink rate as an index of attention and emotion during film clips viewing. *Physiology & Behavior*, *204*, 256–263. https://doi.org/10.1016/j.physbeh.2019.02.037

Mahapatra, D., Winkler, S., & Yen, S.-C. (2008). Motion saliency outweighs other low-level features while watching videos. *Human Vision and Electronic Imaging XIII*, *6806*, 246–255. https://doi.org/10.1117/12.766243

Malik, M., & Isik, L. (2023). Relational visual representations underlie human social interaction recognition. *Nature Communications*, *14*(1), 7317. https://doi.org/10.1038/s41467-023-43156-8

Maner, J. K. (2017). Dominance and Prestige: A Tale of Two Hierarchies. *Current Directions in Psychological Science*, *26*(6), 526–531. https://doi.org/10.1177/0963721417714323

Manninen, S., Tuominen, L., Dunbar, R. I., Karjalainen, T., Hirvonen, J., Arponen, E., Hari, R., Jaaskelainen, I. P., Sams, M., & Nummenmaa, L. (2017). Social Laughter Triggers Endogenous Opioid Release in Humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(25), 6125–6131. https://doi.org/10.1523/JNEUROSCI.0688-16.2017

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220. https://www.ncbi.nlm.nih.gov/pubmed/6018555

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say “hello”? Personality impressions from brief novel voices. *PloS One*, *9*(3), e90779. https://doi.org/10.1371/journal.pone.0090779

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. https://doi.org/10.1037//0022-3514.52.1.81

McMahon, E., Bonner, M. F., & Isik, L. (2023). Hierarchical organization of social action features along the lateral visual pathway. *Current Biology: CB*, *33*(23), 5035-5047.e8. https://doi.org/10.1016/j.cub.2023.10.015

McMahon, E., & Isik, L. (2023). Seeing social interactions. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2023.09.001

McRobbie, D. W., & Graves, M. J. (2007). *MRI from picture to proton* (2nd ed.). Cambridge University Press.

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

Mitteroecker, P., Windhager, S., Müller, G. B., & Schaefer, K. (2015). The morphometrics of "masculinity" in human faces. *PloS One*, *10*(2), e0118374. https://doi.org/10.1371/journal.pone.0118374

Molapour, T., Hagan, C. C., Silston, B., Wu, H., Ramstead, M., Friston, K., & Mobbs, D. (2021). Seven computations of the social brain. *Social Cognitive and Affective Neuroscience*, *16*(8), 745–760. https://doi.org/10.1093/scan/nsab024

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 2688–2692. https://doi.org/10.1073/pnas.1011734108

Morrisey, M. N., Hofrichter, R., & Rutherford, M. D. (2019). Human faces capture attention and attract first saccades without longer fixation. *Visual Cognition*, *27*(2), 158–170. https://doi.org/10.1080/13506285.2019.1631925

Morrison, D., Wang, H., Hahn, A. C., Jones, B. C., & DeBruine, L. M. (2017). Predicting the reward value of faces and bodies from social perception. *PloS One*, *12*(9), e0185093. https://doi.org/10.1371/journal.pone.0185093

Mukamel, R., & Fried, I. (2012). Human intracranial recordings and cognitive neuroscience. *Annual Review of Psychology*, *63*(1), 511–537. https://doi.org/10.1146/annurev-psych-120709-145401

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, *2*(1), 86–97. https://doi.org/10.1002/widm.53

Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision*, *13*(2), 11. https://doi.org/10.1167/13.2.11

Nakano, T., Kato, M., Morito, Y., Itoi, S., & Kitazawa, S. (2013). Blink-related momentary activation of the default mode network while viewing videos. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(2), 702–706. https://doi.org/10.1073/pnas.1214804110

Nakano, T., & Miyazaki, Y. (2019). Blink synchronization is an indicator of interest while viewing videos. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *135*, 1–11. https://doi.org/10.1016/j.ijpsycho.2018.10.012

Nakano, T., Yamamoto, Y., Kitajo, K., Takahashi, T., & Kitazawa, S. (2009). Synchronization of spontaneous eyeblinks while viewing video stories. *Proceedings. Biological Sciences / The Royal Society*, *276*(1673), 3635–3644. https://doi.org/10.1098/rspb.2009.0828

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, 117254. https://doi.org/10.1016/j.neuroimage.2020.117254

Nicolas, G., Bai, X., & Fiske, S. T. (2022). A spontaneous stereotype content model: Taxonomy, properties, and prediction. *Journal of Personality and Social Psychology*, *123*(6), 1243–1263. https://doi.org/10.1037/pspa0000312

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, *13*(3), 135–143. https://doi.org/10.1016/j.tics.2008.12.006

Nummenmaa, L., Glerean, E., Viinikainen, M., Jaaskelainen, 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., Hietanen, J. K., Calvo, M. G., & Hyönä, J. (2011). Food catches the eye but not for everyone: a BMI-contingent attentional bias in rapid detection of nutriments. *PloS One*, *6*(5), e19215. https://doi.org/10.1371/journal.pone.0019215

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2010). Semantic categorization precedes affective evaluation of visual scenes. *Journal of Experimental Psychology. General*, *139*(2), 222–246. https://doi.org/10.1037/a0018858

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

Nummenmaa, L., Lukkarinen, L., Sun, L., Putkinen, V., Seppala, K., Karjalainen, T., Karlsson, H. K., Hudson, M., Venetjoki, N., Salomaa, M., Rautio, P., Hirvonen, J., Lauerma, H., & Tiihonen, J. (2021). Brain Basis of Psychopathy in Criminal Offenders and General Population. *Cerebral Cortex* , *31*(9), 4104–4114. https://doi.org/10.1093/cercor/bhab072

Nummenmaa, L., Saarimaki, H., Glerean, E., Gotsopoulos, A., Jaaskelainen, I. P., Hari, R., & Sams, M. (2014). Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. *NeuroImage*, *102 Pt 2*, 498–509. https://doi.org/10.1016/j.neuroimage.2014.07.063

Nummenmaa, L., Smirnov, D., Lahnakoski, J. M., Glerean, E., Jaaskelainen, I. P., Sams, M., & Hari, R. (2014). Mental action simulation synchronizes action-observation circuits across individuals. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(3), 748–757. https://doi.org/10.1523/JNEUROSCI.0352-13.2014

Nuthmann, A., Schütz, I., & Einhäuser, W. (2020). Salience-based object prioritization during active viewing of naturalistic scenes in young and older adults. *Scientific Reports*, *10*(1), 22057. https://doi.org/10.1038/s41598-020-78203-7

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(24), 9868–9872. https://doi.org/10.1073/pnas.87.24.9868

Oliva, M., & Anikin, A. (2018). Pupil dilation reflects the time course of emotion recognition in human vocalizations. *Scientific Reports*, *8*(1), 4871. https://doi.org/10.1038/s41598-018-23265-x

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2012). Viewpoint (in)dependence of action representations: an MVPA study. *Journal of Cognitive Neuroscience*, *24*(4), 975–989. https://doi.org/10.1162/jocn_a_00195

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

OpenAI. (2024). *ChatGPT*. https://chatgpt.com/

Osgood, C. E., & Suci, G. J. (1955). Factor analysis of meaning. *Journal of Experimental Psychology*, *50*(5), 325–338. https://doi.org/10.1037/h0043965

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: the role of stimulus structure in recognition and classification. *Memory & Cognition*, *26*(1), 146–160. https://doi.org/10.3758/bf03211378

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications*, *9*(1), 332. https://doi.org/10.1038/s41467-017-02722-7

Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, *112*(4), 642–681. https://doi.org/10.1037/pspp0000111

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, *59*(1–2), 185–198. https://doi.org/10.1016/s1071-5819(03)00017-x

Pauling, L., & Coryell, C. D. (1936). The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, *22*(4), 210–216. https://doi.org/10.1073/pnas.22.4.210

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology*, *93*(1), 603–608. https://doi.org/10.1152/jn.00513.2004

Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T., & McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: an FMRI study of eye, mouth and hand movements. *Cerebral Cortex (New York, N.Y.: 1991)*, *15*(12), 1866–1876. https://doi.org/10.1093/cercor/bhi064

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, *56*(4), 2356–2363. https://doi.org/10.1016/j.neuroimage.2011.03.067

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences*, *25*(2), 100–110. https://doi.org/10.1016/j.tics.2020.11.006

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, *209*(4), 481–493. https://doi.org/10.1007/s00221-011-2579-1

Poldrack, R. A., Mumford, J. A., & Nichols, T. E. (2011). *Handbook of functional MRI data analysis*. Cambridge University Press.

Politis, D. N., & Romano, J. P. (1992). A circular block-resampling procedure for stationary data. *Exploring the Limits of Bootstrap*. https://www.google.com/books?hl=fi&lr=&id=ZJzIpNZNVLgC&oi=fnd&pg=PA263&dq=+politis+circular+block-resampling&ots=Yho2o052Di&sig=8HupZEB45JfdEunAwDRwTdz1uHw

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Prolific. (2025). *Prolific*. https://www.prolific.com/

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

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *16*(16), 5205–5215. https://doi.org/10.1523/jneurosci.16-16-05205.1996

Putkinen, V., Nazari-Farsani, S., Karjalainen, T., Santavirta, S., Hudson, M., Seppälä, K., Sun, L., Karlsson, H. K., Hirvonen, J., & Nummenmaa, L. (2023). Pattern recognition reveals sex-dependent neural substrates of sexual perception. *Human Brain Mapping*, *44*(6), 2543–2556. https://doi.org/10.1002/hbm.26229

Quiñones-Camacho, L. E., Fishburn, F. A., Belardi, K., Williams, D. L., Huppert, T. J., & Perlman, S. B. (2021). Dysfunction in interpersonal neural synchronization as a mechanism for social impairment in autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *14*(8), 1585–1596. https://doi.org/10.1002/aur.2513

Ranti, C., Jones, W., Klin, A., & Shultz, S. (2020). Blink rate patterns provide a reliable measure of individual engagement with scene content. *Scientific Reports*, *10*(1), 8267. https://doi.org/10.1038/s41598-020-64999-x

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: a taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718. https://doi.org/10.1037/a0037250

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, *7*, 13289. https://doi.org/10.1038/ncomms13289

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network*, *10*(4), 341–350. https://www.ncbi.nlm.nih.gov/pubmed/10695763

Ro, T., Friggel, A., & Lavie, N. (2007). Attentional biases for faces and body parts. *Visual Cognition*, *15*(3), 322–348. https://doi.org/10.1080/13506280600590434

Rolls, E. T., Joliot, M., & Tzourio-Mazoyer, N. (2015). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, *122*, 1–5. https://doi.org/10.1016/j.neuroimage.2015.07.075

Rösler, L., End, A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. *PloS One*, *12*(7), e0182037. https://doi.org/10.1371/journal.pone.0182037

Roth, N., Rolfs, M., Hellwich, O., & Obermayer, K. (2023). Objects guide human gaze behavior in dynamic real-world scenes. *PLoS Computational Biology*, *19*(10), e1011512. https://doi.org/10.1371/journal.pcbi.1011512

Russell, J. A., Lewicka, M., & Niit, T. (1989). A cross-cultural study of a circumplex model of affect. *Journal of Personality and Social Psychology*, *57*(5), 848–856. https://doi.org/10.1037/0022-3514.57.5.848

Saarimäki, H. (2021). Naturalistic Stimuli in Affective Neuroimaging: A Review. *Frontiers in Human Neuroscience*, *15*, 675068. https://doi.org/10.3389/fnhum.2021.675068

Saarimäki, H., Nummenmaa, L., Volynets, S., Santavirta, S., Aksiuto, A., Sams, M., Jääskeläinen, I. P., & Lahnakoski, J. M. (2023). Cerebral Topographies of Perceived and Felt Emotions. *BioRxiv*, 2023.02.08.521183. https://doi.org/10.1101/2023.02.08.521183

Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision*, *10*(5), 11. https://doi.org/10.1167/10.5.11

Salmi, J., Roine, U., Glerean, E., Lahnakoski, J., Nieminen-von Wendt, T., Tani, P., Leppamaki, S., Nummenmaa, L., Jaaskelainen, I. P., Carlson, S., Rintahaka, P., & Sams, M. (2013). The brains of high functioning autistic individuals do not synchronize with those of others. *Neuroimage Clin*, *3*, 489–497. https://doi.org/10.1016/j.nicl.2013.10.011

Santavirta, S. (2023a). *Functional organization of social perception*. GitHub. https://github.com/santavis/functional-organization-of-social-perception

Santavirta, S. (2023b). *Functional organization of social perception in the human brain*. NeuroVault. https://neurovault.org/collections/IZWVFEYI/

Santavirta, S. (2024a). *Modelling human social vision with cinematic stimuli*. GitHub. https://github.com/santavis/social-vision-in-cinema

Santavirta, S. (2024b). *Taxonomy of human social perception*. GitHub; GitHub. https://github.com/santavis/taxonomy-of-human-social-perception

Santavirta, S., Karjalainen, T., Nazari-Farsani, S., Hudson, M., Putkinen, V., Seppälä, K., Sun, L., Glerean, E., Hirvonen, J., Karlsson, H. K., & Nummenmaa, L. (2023). Functional organization of social perception in the human brain. *NeuroImage*, 120025. https://doi.org/10.1016/j.neuroimage.2023.120025

Santavirta, S., Malén, T., Erdemli, A., & Nummenmaa, L. (2024). A taxonomy for human social perception: Data-driven modeling with cinematic stimuli. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspa0000415

Santavirta, S., Paranko, B., Seppala, K., Hyona, J., & Nummenmaa, L. (2024). Modelling human social vision with cinematic stimuli. In *bioRxiv* (p. 2024.10.18.618846). https://doi.org/10.1101/2024.10.18.618846

Santavirta, S., Wu, Y., & Nummenmaa, L. (2024). GPT-4V shows human-like social perceptual capabilities at phenomenological and neural levels. In *bioRxiv*. https://doi.org/10.1101/2024.08.20.608741

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842. https://doi.org/10.1016/s1053-8119(03)00230-1

Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture*, *2*(1). https://doi.org/10.9707/2307-0919.1116

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, *103*(4), 663–688. https://doi.org/10.1037/a0029393

Scott, S. K., Lavan, N., Chen, S., & McGettigan, C. (2014). The social life of laughter. *Trends in Cognitive Sciences*, *18*(12), 618–620. https://doi.org/10.1016/j.tics.2014.09.002

Shin, Y. S., Chang, W.-D., Park, J., Im, C.-H., Lee, S. I., Kim, I. Y., & Jang, D. P. (2015). Correlation between inter-blink interval and episodic encoding during movie watching. *PloS One*, *10*(11), e0141242. https://doi.org/10.1371/journal.pone.0141242

Simms, L. J. (2007). The big seven model of personality and its relevance to personality pathology. *Journal of Personality*, *75*(1), 65–94. https://doi.org/10.1111/j.1467-6494.2006.00433.x

Smirnov, D., Saarimaki, H. G. E., Hari, R., Sams, M., & Nummenmaa, L. (2019). Emotions amplify speaker-listener neural alignment. *Human Brain Mapping*, *40*(16), 4777–4788. https://doi.org/10.1002/hbm.24736

Smirnov, Dmitry, Lachat, F., Peltola, T., Lahnakoski, J. M., Koistinen, O.-P., Glerean, E., Vehtari, A., Hari, R., Sams, M., & Nummenmaa, L. (2017). Brain-to-brain hyperclassification reveals action-specific motor mapping of observed actions in humans. *PloS One*, *12*(12), e0189508. https://doi.org/10.1371/journal.pone.0189508

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), 16–16. https://doi.org/10.1167/13.8.16

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences*, *23*(8), 699–714. https://doi.org/10.1016/j.tics.2019.05.004

SR Research. (2025). *EyeLink Data Viewer*. https://www.sr-research.com/data-viewer/

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *52*(1), 77–86. https://doi.org/10.1016/j.ijpsycho.2003.12.005

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14425–14430. https://doi.org/10.1073/pnas.1008662107

Stoeckel, C., Gough, P. M., Watkins, K. E., & Devlin, J. T. (2009). Supramarginal gyrus involvement in visual word recognition. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *45*(9), 1091–1096. https://doi.org/10.1016/j.cortex.2008.12.004

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour*, *4*(4), 361–371. https://doi.org/10.1038/s41562-019-0800-6

Stoll, J., Thrun, M., Nuthmann, A., & Einhäuser, W. (2015). Overt attention in natural scenes: objects dominate features. *Vision Research*, *107*, 36–48. https://doi.org/10.1016/j.visres.2014.11.006

Suda, M., Takei, Y., Aoyama, Y., Narita, K., Sakurai, N., Fukuda, M., & Mikuni, M. (2011). Autistic traits and brain activation during face-to-face conversations in typically developed adults. *PloS One*, *6*(5), e20021. https://doi.org/10.1371/journal.pone.0020021

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

Sutherland, C. A. M., Rhodes, G., Burton, N. S., & Young, A. W. (2020). Do facial first impressions reflect a shared social reality? *British Journal of Psychology* , *111*(2), 215–232. https://doi.org/10.1111/bjop.12390

Sutherland, C. A. M., & Young, A. W. (2022). Understanding trait impressions from faces. *British Journal of Psychology* , *113*(4), 1056–1078. https://doi.org/10.1111/bjop.12583

Tarhan, L., & Konkle, T. (2020). Sociality and interaction envelope organize visual action representations. *Nature Communications*, *11*(1), 3002. https://doi.org/10.1038/s41467-020-16846-w

The MathWorks Inc. (2024a). *fitrensemble MATLAB function*. https://se.mathworks.com/help/stats/fitrensemble.html

The MathWorks Inc. (2024b). *predictorImportance MATLAB function*. https://se.mathworks.com/help/stats/classreg.learning.classif.compactclassificationensemble.predictorimportance.html

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition*, *13*(6), 657–665. https://doi.org/10.1080/13506280500410949

Thornton, M. A., & Tamir, D. I. (2020). People represent mental states in terms of rationality, social impact, and valence: Validating the 3d Mind Model. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *125*, 44–59. https://doi.org/10.1016/j.cortex.2019.12.012

Thornton, M. A., & Tamir, D. I. (2022). Six dimensions describe action understanding: The ACT-FASTaxonomy. *Journal of Personality and Social Psychology*, *122*(4), 577–605. https://doi.org/10.1037/pspa0000286

Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology*, *44*(3), 187–194. https://doi.org/10.1002/ejsp.2015

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, *63*, 483–509. https://doi.org/10.1146/annurev-psych-120710-100412

Trevor, C., Arnal, L. H., & Frühholz, S. (2020). Terrifying film music mimics alarming acoustic feature of human screams. *The Journal of the Acoustical Society of America*, *147*(6), EL540. https://doi.org/10.1121/10.0001459

Tucciarelli, R., Turella, L., Oosterhof, N. N., Weisz, N., & Lingnau, A. (2015). MEG multivariate analysis reveals early abstract action representations in the lateral occipitotemporal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(49), 16034–16045. https://doi.org/10.1523/JNEUROSCI.1422-15.2015

Tucciarelli, R., Wurm, M., Baccolo, E., & Lingnau, A. (2019). The representational space of observed actions. *ELife*, *8*. https://doi.org/10.7554/eLife.47686

Tzschaschel, E., Brooks, K. R., & Stephen, I. D. (2022). The valence-dominance model applies to body perception. *Royal Society Open Science*, *9*(9), 220594. https://doi.org/10.1098/rsos.220594

Vaidya, A. R., Pujara, M. S., Petrides, M., Murray, E. A., & Fellows, L. K. (2019). Lesion studies in contemporary neuroscience. *Trends in Cognitive Sciences*, *23*(8), 653–671. https://doi.org/10.1016/j.tics.2019.05.009

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research: JMLR*. https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbcl

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(32), E3353-61. https://doi.org/10.1073/pnas.1409860111

Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia*, *112*, 31–39. https://doi.org/10.1016/j.neuropsychologia.2018.02.023

Walker, M., & Vetter, T. (2016). Changing the personality of a face: Perceived Big Two and Big Five personality factors modeled in real photographs. *Journal of Personality and Social Psychology*, *110*(4), 609–624. https://doi.org/10.1037/pspp0000064

Wan, J. (2016). *The Conjuring 2*. New Line Cinema, RatPac-Dune Entertainment, The Safran Company, Atomic Monster.

Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. J. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision*, *12*(1), 16. https://doi.org/10.1167/12.1.16

Wegrzyn, M., Riehle, M., Labudda, K., Woermann, F., Baumgartner, F., Pollmann, S., Bien, C. G., & Kissler, J. (2015). Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *69*, 131–140. https://doi.org/10.1016/j.cortex.2015.05.003

Wilkowski, B. M., Fetterman, A., Lappi, S. K., Williamson, L. Z., Leki, E. F., Rivera, E., & Meier, B. P. (2020). Lexical derivation of the PINT taxonomy of goals: Prominence, inclusiveness, negativity prevention, and tradition. *Journal of Personality and Social Psychology*, *119*(5), 1153–1187. https://doi.org/10.1037/pspp0000268

Williams, C. C., & Castelhano, M. S. (2019). The Changing Landscape: High-Level Influences on Eye Movement Guidance in Scenes. *Vision (Basel, Switzerland)*, *3*(3). https://doi.org/10.3390/vision3030033

Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). *Detectron2*. Detectron2. https://github.com/facebookresearch/detectron2

Wurm, M. F., & Caramazza, A. (2019). Lateral occipitotemporal cortex encodes perceptual components of social actions rather than abstract representations of sociality. *NeuroImage*, *202*, 116153. https://doi.org/10.1016/j.neuroimage.2019.116153

Wurm, M. F., Caramazza, A., & Lingnau, A. (2017). Action Categories in Lateral Occipitotemporal Cortex Are Organized Along Sociality and Transitivity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(3), 562–575. https://doi.org/10.1523/JNEUROSCI.1717-16.2016

Wurm, M. F., & Lingnau, A. (2015). Decoding actions at different levels of abstraction. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(20), 7727–7735. https://doi.org/10.1523/JNEUROSCI.0188-15.2015

Wurm, M. F., & Schubotz, R. I. (2018). The role of the temporoparietal junction (TPJ) in action observation: Agent detection rather than visuospatial transformation. *NeuroImage*, *165*, 48–55. https://doi.org/10.1016/j.neuroimage.2017.09.064

Wyly, S., Jinon, N., Francis, T., Evans, H., Kao, T. L., Lambert, S., Montgomery, S., Newlove, M., Mariscal, H., Nguyen, H., Cole, H., Aispuro, I., Robledo, D., Tenaglia, O., Weinberger, N., Nguyen, B., Waits, H., Jorian, D., Koch-Kreher, L., … Wilson, R. C. (2024). The psychophysiology of Mastermind: Characterizing response times and blinking in a high-stakes television game show. *Psychophysiology*, *61*(3), e14485. https://doi.org/10.1111/psyp.14485

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *The American Psychologist*, *35*(2), 151–175. https://doi.org/10.1037/0003-066X.35.2.151

Zemeckis, R. (1994). *Forrest Gump*. Paramount Pictures.

List of Figures, Tables and Appendices

Figures

Appendices

Appendices

Appendix 1. Brain regions mentioned in this thesis with their abbreviations and locations.

Region	Abbreviation(s)	Lobe / Location	AAL2 atlas / Functional
Inferior frontal gyrus	F Inf, IFG	Frontal	AAL2
Middle frontal gyrus	F Mid, MidFG	Frontal	AAL2
Precentral gyrus	Precentral	Frontal	AAL2
Superior frontal gyrus	F Sup, SFG	Frontal	AAL2
Superior frontal gyrus, medial part	F Sup Med, SFG Med	Frontal	AAL2
Cingulate gyrus	Cingulate (Ant / Mid / Post)	Frontal	AAL2
Inferior frontal gyrus, opercular part	IFG Oper	Frontal	AAL2
Inferior frontal gyrus, orbital part	IFC Orb	Frontal	AAL2
Inferior frontal gyrus, triangular part	IFG Tri	Frontal	AAL2
Olfactory bulb	Olfactory	Frontal	AAL2
Orbitofrontal cortex	OFC (Lat / Post)	Frontal	AAL2
Supplementary Motor Area	Supp Motor Area	Frontal	AAL2
Calcarine sulcus	Calcarine	Occipital	AAL2
Cuneus	Cuneus	Occipital	AAL2
Inferior occipital gyrus	Occ Inf, IOccG	Occipital	AAL2
Lingual gyrus	Lingual	Occipital	AAL2
Medial occipital gyrus	Occ Med, MOccG	Occipital	AAL2
Superior occipital gyrus	Occ Sup, SOccG	Occipital	AAL2
Angular gyrus	Angular	Parietal	AAL2
Inferior parietal gyrus	P Inf, IPG	Parietal	AAL2
Paracentral gyrus	Paracentral	Parietal	AAL2
Postcentral gyrus	Postcentral	Parietal	AAL2
Precuneus	Precuneous	Parietal	AAL2
Rolandic operculum	Rolandic Oper	Parietal	AAL2
Superior parietal gyrus	P Sup, SPG	Parietal	AAL2
Supramarginal gyrus	SupM	Parietal	AAL2
Amygdala	Amygdala	Subcortical	AAL2
Caudate nucleus	Caudate	Subcortical	AAL2
Globus pallidus	Pallidum	Subcortical	AAL2
Insula	Insula	Subcortical	AAL2
Putamen	Putamen	Subcortical	AAL2
Thalamus	Thalamus	Subcortical	AAL2
Fusiform gyrus	Fusiform, FG	Temporal	AAL2
Heschl gyrus	Heschl	Temporal	AAL2
Hippocampus	Hippocampus	Temporal	AAL2
Middle temporal gyrus	MTG, T Mid	Temporal	AAL2
Parahippocampal gyrus	Parahippocampal	Temporal	AAL2
Superior temporal gyrus	STG, T Sup	Temporal	AAL2
Superior temporal pole	T Pole Sup	Temporal	AAL2
Medial frontal cortex	MFC	Frontal	Functional
Occipital face area	OFA	Occipital	Functional
Primary visual cortex	V1	Occipital	Functional
Lateral occipitotemporal cotrex	LOTC	Occipital / Temporal	Functional
Fusiform face area	FFA	Temporal	Functional
Superior temporal sulcus	STS	Temporal	Functional
Temporoparietal junction	TPJ	Temporal / Parietal	Functional

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2025

ISBN 978-952-02-0130-2 (PRINT)
ISBN 978-952-02-0131-9 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)