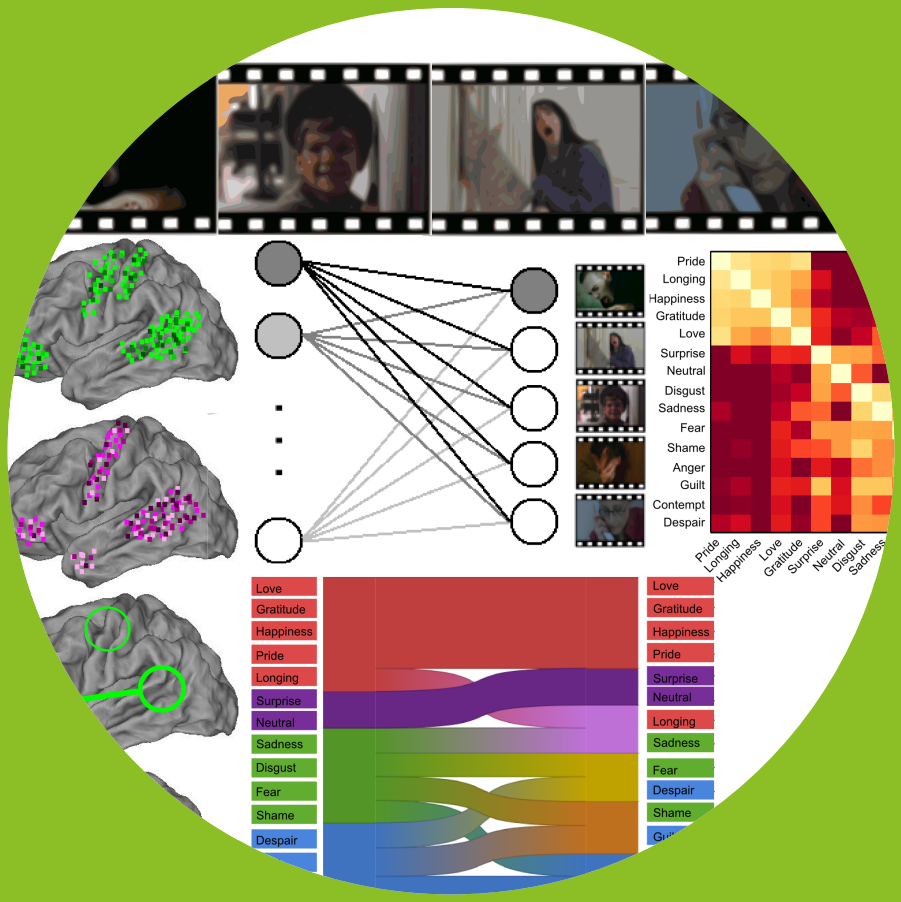
Department of Neuroscience and Biomedical Engineering

Decoding emotions from brain activity and connectivity patterns

Heini Saarimäki



Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 17/2018

Decoding emotions from brain activity and connectivity patterns

Heini Saarimäki

A doctoral dissertation completed for the degree of Doctor of Philosophy to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall A2 of Kandidaattikeskus (Otakaari 1) on 16 February 2018 at 12.

Aalto University
School of Science
Department of Neuroscience and Biomedical Engineering
Brain and Mind Laboratory

Supervising professor
Professor Mikko Sams, Aalto University, Finland

Thesis advisors
Professor Lauri Nummenmaa, University of Turku, Finland
Professor Iiro P. Jääskeläinen, Aalto University, Finland

Preliminary examiners
Professor Ralph Adolphs, California Institute of Technology, USA
Professor Kevin S. LaBar, Duke University, USA

Opponent
Professor Christian Keysers, Netherlands Institute for Neuroscience, Netherlands

Aalto University publication series
DOCTORAL DISSERTATIONS 17/2018

ISBN 978-952-60-7817-5 (printed)
ISBN 978-952-60-7818-2 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-7818-2

Unigrafia Oy
Helsinki 2018

Finland

441 697
Printed matter

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Heini Saarimäki

Name of the doctoral dissertation
Decoding emotions from brain activity and connectivity patterns

Publisher School of Science

Unit Department of Neuroscience and Biomedical Engineering

Series Aalto University publication series DOCTORAL DISSERTATIONS 17/2018

Field of research Systems neuroscience

Manuscript submitted 6 October 2017 **Date of the defence** 16 February 2018

Permission to publish granted (date) 29 November 2017 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

Emotions guide both human and animal behavior providing the means for survival in a constantly changing environment. Different emotions seem to be distinct from each other in several aspects, including physiological changes, bodily sensations, facial expressions, and subjective experience. Whether and how such emotion categories exist at the neural level remains however under debate. The goal of this dissertation was to employ pattern classification methods to investigate the neural underpinnings of different emotion states. Specifically, it was hypothesized that if different emotions have distinct neural bases, we should be able to reliably classify them from brain activity and connectivity patterns. Further, it was hypothesized that the classifier confusions presumably reveal which emotions have similar neural substrates.

Multiple emotional states were induced in four studies with altogether 109 participants using emotional movies, mental imagery, and narratives while participants' brain activity was measured with functional magnetic resonance imaging (fMRI). Several approaches to the fMRI data analyses were employed: multivariate pattern classification to distinguish voxel activity and functional connectivity patterns underlying different emotions, representational similarity analysis to compare experienced and neural similarity of different emotions, functional connectivity analysis to reveal emotional modulations in brain connectivity, univariate methods such as general linear model (GLM) to visualize the neural substrates of different emotions, and correlation analyses to compare the relationship of different emotions at different emotion-related components.

Results from these studies show that specific emotions can be classified from both voxel activity and functional connectivity patterns. Successful pattern classification of voxel activity across the whole brain shows that different emotions have distinct brain activity patterns that generalize across participants and across emotion induction techniques. Further, emotions that subjectively feel more similar also have more similar neural underpinnings. Functional connectivity is modulated by emotional content and shows distinct patterns for different emotions especially within the default mode network (DMN). DMN regions especially in the cortical midline, together with somatomotor, sensory, and subcortical areas, support most emotions. Finally, distinctness of emotions is related at the level of different components, including facial expressions, bodily sensations, emotional evaluations, subjective experiences, and neural substrates.

To conclude, emotions have distinct brain activity and connectivity patterns that encompass large extent of the brain. Emotions can thus be viewed as systemic states that, at a given moment, facilitate and constrain other mental functions.

Keywords Emotion; brain; pattern classification; functional connectivity; fMRI; MVPA; RSA

ISBN (printed) 978-952-60-7817-5 **ISBN (pdf)** 978-952-60-7818-2

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2018

Pages 180 **urn** http://urn.fi/URN:ISBN:978-952-60-7818-2

A! Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Heini Saarimäki

Väitöskirjan nimi
Tunteiden luokittelu aivojen aktivaatiosta ja konnektiviteetista

Julkaisija Perustieteiden korkeakoulu

Yksikkö Neurotieteen ja lääketieteellisen tekniikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 17/2018

Tutkimusala Systeeminen neurotiede

Käsikirjoituksen pvm 06.10.2017 **Väitöspäivä** 16.02.2018

Julkaisuluvan myöntämispäivä 29.11.2017 **Kieli** Englanti

☐ **Monografia** ☒ **Artikkeliväitöskirja** ☐ **Esseeväitöskirja**

Tiivistelmä

Tunteet ohjaavat ihmisten ja eläinten käyttäytymistä tarjoamalla keinoja selviytyä jatkuvasti muuttuvassa ympäristössä. Eri tunteet eroavat toisistaan monin tavoin: eri tunnetiloihin liittyy erilaisia fysiologisia muutoksia, kehon tuntemuksia, kasvonilmeitä ja yksilöllisiä tunnekokemuksia. Tunnetutkimuksessa kuitenkin kiistellään siitä, miten eri tunnetilat eroavat toisistaan aivoissa. Tämän väitöskirjan tavoitteena oli hyödyntää koneoppimisen menetelmiä eri tunnetilojen hermostollisen perustan tutkimiseksi. Erityisesti oletettiin, että jos eri tunteilla on erillinen aivoperusta, ne pitäisi pystyä luokittelemaan aivojen aktiivisuuden ja yhteyksien muutosten perusteella. Lisäksi oletettiin, että käytetyn luokittelualgoritmin tekemät virheet paljastavat, millä tunteilla on keskenään samankaltaisempi hermostollinen perusta.

Eri tunnetiloja tuotettiin neljässä tutkimuksessa (yhteensä 109 vapaaehtoista osallistujaa) tunnepitoisten elokuvien, eläytymistehtävän ja tarinoiden avulla samalla, kun osallistujien aivojen aktivaatiota mitattiin toiminnallisella magneettikuvantamisella (fMRI). Datan analyysissa käytettiin useita eri lähestymistapoja: eri tunteisiin liittyviä vokselikohtaisia aktivaatioita ja funktionaalista konnektiviteettia eriteltiin koneoppimisen luokittelualgoritmeja käyttäen, eri tunteiden koettua ja hermostollista samankaltaisuutta vertailtiin samankaltaisuusanalyysia hyödyntäen, tunteiden aikaansaamia muutoksia aivojen konnektiivisuudessa tutkittiin funktionaalista konnektiivisuusanalyysia käyttäen, yksimuuttujamenetelmiä kuten lineaarista regressiota käytettiin eri tunteiden aivoperustan visualisointeihin ja korrelaatioanalyyseilla verrattiin tunnetilojen eroja tunteiden eri komponenteissa.

Tulokset osoittavat, että tunteita voidaan luokitella sekä vokseliaktivaatioiden että funktionaalisen konnektiviteetin muutosten perusteella. Onnistunut koko aivojen aktivaatioon perustuva luokittelu osoittaa, että eri tunteilla on erillinen aivoperusta, joka yleistyy henkilöstä ja tunteiden herättämistekniikasta toiseen. Lisäksi tunteilla, jotka koetaan samankaltaisempina, on myös samankaltaisempi aivoperusta. Tunne muokkaa funktionaalista konnektiviteettia, jonka tunnekohtaiset erot ovat selvimpiä aivojen lepotilaverkostoissa (default mode network, DMN). Aivojen keskilinjan rakenteiden lisäksi erityisesti somatomotoriset, sensoriset ja subkortikaaliset alueet aktivoituvat useimpien tunteiden aikana. Tunteiden erillisyys ilmenee eri komponteissa, kuten kasvonilmeissä, kehon tuntemuksissa, tunnesisällön arvioinnissa, yksilöllisessä tunnekokemuksessa sekä aivoperustassa.

Yhteenvetona voidaan todeta, että eri tunteisiin liittyy kullekin tunteelle tyypillinen aivojen aktivaatio ja konnektiviteetti, jotka kattavat suuren osan aivoista. Tunteita voidaan siis pitää aivojen tilana, joka kullakin ajanhetkellä vaikuttaa muihin mielen toimintoihin.

Avainsanat Tunteet; aivot; koneoppiminen; funktionaalinen konnektiviteetti; fMRI; MVPA; RSA

ISBN (painettu) 978-952-60-7817-5 **ISBN (pdf)** 978-952-60-7818-2

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2018

Sivumäärä 180 **urn** http://urn.fi/URN:ISBN:978-952-60-7818-2

Acknowledgements

Time has come to finally pay my respects to the people with whom I have shared these years of doctoral studies which were conducted at the Department of Neuroscience and Biomedical Engineering (NBE) - former Department of Biomedical Engineering and Computational Science (BECS) - part of Aalto University School of Science. My supervision team has been fantastic; working with and learning from such people has been a privilege. I would like to thank professor Mikko Sams for guidance and support throughout the years. Thank you for giving me the opportunity to conduct the PhD in this great research team, and thank you for the many discussions revolving around science and life. I really look up to you both as a leader and a person. Professor Lauri Nummenmaa, thank you for the enthusiasm, inspiring ideas and drive to excel, and trust in the good work that we have been conducting over the years. Professor Iiro Jääskeläinen, thank you especially for guidance and practical discussions regarding life, science, and life as a scientist.

I have had the great pleasure to work with many inspiring scientists over the years. I am especially grateful for all co-authors for valuable input and discussions. Professor emerita Riitta Hari, thank you for inspiration and guidance and challenging us youngsters to think it through. Professor Patrik Vuilleumier, I am grateful for the insightful discussions regarding both emotions and research methods. It has been a wonderful opportunity to learn from such a great scholar in the field.

During the final stage of the doctoral studies, I was lucky to have world-renowned emotion researchers commenting my work. I wish to thank the pre-examiners of my thesis, professor Ralph Adolphs and professor Kevin LaBar, for encouraging feedback and kind words. Professor Christian Keysers kindly agreed on being my opponent and I wish we share an interesting discussion during the defence.

The current and former members of the Brain and Mind Lab have always been there through joy and sorrow - my warmest thank you for that and the cakes, too. This has been a wonderful work community and I have learned a lot during the years, even of topics I did not necessarily expect or wish to learn from. I want to especially thank Dmitry Smirnov, Enrico Glerean, Athanasios Gotsopoulos, Juha Lahnakoski, Satu Saalasti and Mareike Bacha-Trams for all mental and practical support during the years. For all BMLers, I am grateful that our paths crossed and I hope we will cherish collaboration and friendship in future. I also would like to thank all other colleagues that I have worked with and met at NBE / BECS, especially the BECS sports club and NBE choir, as well as the administrative staff members and IT team

for making everything happen so smoothly in the background. I am also grateful for the AMI center personnel, especially Marita Kattelus, Tuomas Tolvanen, and Toni Auranen, for their help with fMRI scanning.

This work would not have been possible without the financial support from the aivoAALTO research project, Academy of Finland and European Research Council support for the projects I have worked with, and personal grants from Finnish Cultural Foundation and Kordelin Foundation.

Friends inside and outside academia have made sure that life has been in a balance also during the PhD even though work and life have kept me busy and sometimes absent-minded. I am grateful especially to Hanne Hellstedt, Tiina Lehtonen, Riikka Sanchez, Amy Lindroos, Kasper Lindroos, and Saara Huhtanen for friendship and support. Also many past and present friends and colleagues at University of Helsinki and University of Edinburgh, at the crossfit gym, in the orienteering club, and in various dog sports have made sure that unwinding from work is possible - thank you. Finally, peer support is of enormous importance, and I especially wish to express my gratitude to Otto Lappi for mentoring, inspiring discussions and reading and commenting my work, Saana Sipari for being the shoulder, and Johanna Forsman for keeping me running through all these years.

Finally, I would like to thank my family and extended family for all support - both mental and practical - which have become immensely important in these past years when combining work and family life. Special thanks to my parents Helena and Jarmo Heikkilä for always believing in me no matter what I decided to strive for. My big sister Pia, my big brother Tomi and his family Miia, Otso, Oona, and Ella, and my uncle Markus and his family Helena and Tilda have always been not just beloved family members but also inspiring mentors for life, academia, and arts - thank you. Also my extended family, especially my parents-in-law Tuula and Kimmo Saarimäki, have played a big role in our small family's life during the PhD. I wish to thank you all for your caring support.

My most loving thank you goes to my husband Jarno, my son Touko, and the little light within: you have taught me more about emotions than a lifetime as an affective neuroscientist would. Thank you for all the love, oxytocin and curiosity for the universe that you spread around. To our furry companion Nano, I am forever grateful for showing me how animal emotions and cross-species synchronization work in practice.

Espoo, January 2, 2018
Heini Saarimäki

Contents

List of Abbreviations and Symbols

ACC	Anterior cingulate cortex
Amy	Amygdala
ANET	Affective Norms for English Text
ANOVA	Analysis of variance
aPFC	Anterior prefrontal cortex
BH-FDR	Benjamini and Hochberg (1995) false discovery rate
BOLD	Blood-oxygen-level dependent
Cer	Cerebellum
DMN	Default mode network
EEG	Electroencephalography
EPI	Echo planar imaging
FDR	False discovery rate
fMRI	Functional magnetic resonance imaging
FOV	Field of view
GLM	General linear model
Hi	Hippocampus
HRF	Hemodynamic response function
IFG	Inferior frontal gyrus
Ins	Insula
LOC	Lateral occipital cortex
MEG	Magnetoencephalography
MFC	Medial frontal cortex
MNI	Montreal Neurological Institute
MP-RAGE	Magnetization prepared rapid gradient echo

MRI	Magnetic resonance imaging
MVPA	Multi-voxel pattern analysis
NAcc	Nucleus accumbens
OFC	Orbitofrontal cortex
PaC	Paracingulate cortex
PCC	Posterior cingulate cortex
PCun	Precuneus
PET	Positron emission tomography
PFC	Prefrontal cortex
postCG	Postcentral gyrus
preCG	Precentral gyrus
RSA	Representational similarity analysis
SBPS	Seed-based phase synchronization
SMA	Supplementary motor area
Th	Thalamus
TMS	Transcranial magnetic stimulation

List of Publications

This doctoral dissertation consists of a summary, three articles published in peer-reviewed journals, and two manuscripts under review. Publications are referred to by their roman numerals.

I Saarimäki H, Gotsopoulos A, Jääskeläinen IP, Lampinen J, Vuilleumier P, Sams M, Hari R, Nummenmaa L (2016) Discrete neural signatures of basic emotions. Cerebral Cortex 26: 2563-2573.

II Saarimäki H, Ejtehadian LF, Glerean E, Jääskeläinen IP, Vuilleumier P, Sams M, Nummenmaa L (under revision) Distributed affective space represents multiple emotion categories across the brain. Social Cognitive and Affective Neuroscience.

III Nummenmaa L, Saarimäki H, Jääskeläinen IP, Glerean E, Gotsopoulos A, Hari R, Sams M (2014) Emotional speech synchronizes brain across listeners and engages large-scale dynamic brain networks. NeuroImage 102:498-509.

IV Saarimäki H, Glerean E, Smirnov D, Mynttinen H, Jääskeläinen IP, Sams M, Nummenmaa L (under review) Classification of emotions from brain connectivity patterns. Journal of Neuroscience.

V Nummenmaa L, Saarimäki H (in press) Emotions as discrete patterns of systemic activity. Neuroscience Letters.

Author's Contribution

Publication I: Discrete neural signatures of basic emotions

The candidate gathered the data, analyzed the data, and wrote the manuscript. Assistance for the data acquisition was received from Athanasios Gotsopoulos and Marita Kattelus. All co-authors gave valuable input during writing of the manuscript.

Publication II: Distributed affective space represents multiple emotion categories across the brain.

The candidate designed the experiment and the stimuli, gathered the data, analyzed the data, and wrote the manuscript. Assistance for the data acquisition was received from Lara Farzaneh Ejtehadian and Marita Kattelus. All co-authors gave valuable input during writing of the manuscript.

Publication III: Emotional speech synchronizes brain across listeners and engages large-scale dynamic brain networks

The candidate designed the stimuli, gathered the data, analyzed part of the data, and contributed to the writing of the manuscript. Assistance for the data acquisition was received from Athanasios Gotsopoulos and Marita Kattelus. All co-authors gave valuable input during writing of the manuscript.

Publication IV: Classification of emotions from brain connectivity patterns

The candidate designed the stimuli, gathered the data, analyzed the data with Dr Enrico Glerean, and wrote the manuscript. Assistance for the data acquisition was received from Henri Mynttinen and Marita Kattelus. All co-authors gave valuable input during writing of the manuscript.

Publication V: Emotions as discrete patterns of systemic activity

The candidate conducted the meta-analysis and contributed to the writing of the manuscript.

1. Introduction

1.1 Emotions: more than a feeling

From scientists to parents of any two-year old, from philosophers since ancient Greece to young lovers, the nature of emotions is one of the mysteries continuously intriguing the human mind. Emotions guide our behavior to protect our body and mind by modulating the activation of cardiovascular, skeletomuscular, neuroendocrine, and autonomic nervous systems as well as higher-order cognitive functions to respond to the dangers and possibilities around us (Levenson, 2003; LeDoux, 2012). Thus, they constantly modulate our brain activity and prepare us to act and survive in a changing environment by orienting actions and modulating approach versus avoidance motivation (Lang, 1995; Elliot et al., 2013; Anderson and Adolphs, 2014). In our everyday life, we can subjectively differentiate between discrete, phenomenological emotional states such as feeling disgusted, happy, or proud. However, it is yet unresolved how such emotional feelings are brought about by the central nervous system. Given the prevalence of emotional problems in many psychiatric disorders and the widespread consequences these problems bring to the lives of individuals, their families, and the society, it is of the highest interest to understand the organization of the neural circuits underlying different emotions.

Emotion as a topic of scientific enquiry has proven surprisingly problematic for both psychologists and neuroscientists. This is partly due to the difficulty in defining what emotion is in scientific terms, but also in our inability to distance ourselves from the everyday use of the term. These have led to some suggestions that the scientific community should abandon the term altogether (LeDoux, 2012). Also, it is unclear to what extent our concept of emotions picks out a homogeneous kind of state (Griffiths, 1997). To illustrate the large variation under the umbrella term emotion, commonly used emotion categories such as happiness and disgust share little in common - they vary in function, neurobiology, experience and so forth - yet there is something similar in both states that makes us categorize them under the term emotion rather than group them together with other cognitive functions (Nummenmaa and Saarimäki, in press).

Emotion can be conceptualized as a concerted, adaptive, phasic or episodic (meaning it has clear on- and off-sets and limited duration) change in multiple physiological systems including somatic and neural components in response to the value of a stimulus (Adolphs, 2002b; see also Damasio, 1995, 1999; Plutchik, 1980; Scherer, 2000). Emotional response typically involves concerted changes in a very large number of somatic parameters, including endocrine, visceral, autonomic, neural, and musculoskeletal changes such as facial expressions, all of which unfold in a complex fashion over time (Adolphs, 2002b).

According to Mulligan and Scherer (2012), x is an emotion only 1) if x is an affective episode, 2) if x has the property of intentionality (i.e., of being directed), 3) x contains bodily

changes (arousal, expression etc.) that are felt, 4) x contains a perceptual or intellectual episode, y, which has the property of intentionality, 5) the intentionality of x is inherited from the intentionality of y, 6) x is triggered by at least one appraisal, and 7) x is guided by at least one appraisal.

A recently outlined framework defines emotions as central, functional states, implemented in the activity of neural systems, that are caused by external sensory stimuli or internal memories and that regulate a multitude of complex behavioral, cognitive, and somatic changes (Anderson and Adolphs, 2014). Adolphs (2017) divides emotion into its specific aspects: emotional states, feelings, emotion concepts, emotion attributions, and emotional expressions. In this framework, the emotion words we use to describe our functional emotional state are concepts that describe our inner feelings. Feelings constitute the consciously experienced part of the emotional state, derived from the internal or external situation surrounding us using emotion attributions, and expressed to others using emotional expressions.

Affective neuroscience investigates the neural systems underlying different aspects of emotion. The advances in neuroimaging techniques during the past three decades have increased research in this field tremendously. Also, since approaches where emotions are defined as functional states of the system have gained support (LeDoux, 2012; Kober et al., 2008; Barrett, 2006; Anderson and Adolphs, 2014; Nummenmaa & Saarimäki, in press), the importance of emotions to all our functioning - behavior, information processing, social interaction, practically all aspects of mental life - are clearer and the benefits of understanding emotions thus have more implications to other branches of neuroscience. Yet, despite the efforts, the neural basis of emotions is far from clear and there is an ongoing debate regarding the neural circuits underlying different emotions. Therefore, the current work aims at elucidating this question.

1.2 Neural basis of emotions

1.2.1 How are emotions organized in the brain?

> *Currently, we do not know what is the level of biological organization or function at which any functional state, emotional or not, is instantiated in the brain. They could be a neuromodulatory system, a neuroanatomical structure, a distributed neural network, a type of firing pattern, or all of the above. (Anderson & Adolphs, 2014)*

The organization of emotions in the brain remains an open question and hotly debated topic in the field of affective neuroscience (for a review, see Hamann, 2012). While different approaches - including for instance basic emotion theories and constructivist theories - acknowledge that the emotion space is carved up into discrete emotions, they differ in terms of how different emotions originate biologically (Lewis and Liu, 2011).

Basic emotion theories (e.g., Ekman, 1992; Ekman and Cordaro, 2011; Panksepp, 1982; Panksepp and Watt, 2011; Levenson, 2011; Izard, 1993, 2011) posit that there exists a set of specific emotions that are evolutionary scripts differing from each other in terms of subjective experience, physiology, and neural basis. These scripts have been shaped during the evolution to serve distinct survival functions via distinguishable neural circuits and physiological systems (Ekman, 1992, 1999; Panksepp, 1982; Damasio, 1999). Therefore, the basic emotions should show unique neural signatures.

Another way to view the behavioural, physiological and subjective bases of emotions is to define them in terms of a limited number of general-purpose systems (e.g. Barrett, 2006). Accordingly, the contemporary constructivist theories build on dimensional theories that suggest that emotions can be reduced to a low-dimensional space (Russell, 1980), typically one governing valence (pleasure versus displeasure) and one arousal (intensity of the emotional response), although the exact number of suggested dimensions varies (Fontaine et al., 2007). The relative activity of these systems could then generate different patterns of emotional behaviour and experiences. The constructivist theories suggest that all emotions are constructed from the activation of different brain regions that may not be specific to emotion, but may combine in various ways to produce the emotional states (Barrett, 2006; Barrett et al., 2007; Russell, 2003). Therefore, they suggest that no unique neural systems exist for specific emotions, a view that is largely shared by contemporary basic emotion theories (for a review, see e.g. Hamann, 2012). Further, constructivist theories stress that two occasions of the same emotion are never the same, thus leading to different neural activations (Clark-Polner et al., 2016) and, accordingly, we could not identify brain structures involved in processing of a particular emotion. However, the variety between occasions is true in most other systems: for instance, exact occasions of visual stimuli are likely always different, yet we can identify regions that deal with rough similarities of stimuli, including faces and objects.

Strong advocates of these two theory traditions sometimes see the theories mutually exclusive (Barrett, 2006, 2017; Clark-Polner et al., 2016), but according to others they can co-exist (Panksepp, 2007; Hamann, 2012). Specific emotions - both basic and other - can be presented in the dimensional space and rely on similar underlying components (Hamann, 2012). Yet, as is probable for all mental states that differ from each other, the differences between specific emotions should be somehow present also in the neural system, and one of the aims of the current work is to investigate how.

Where in the brain can we, then, expect to observe different responses between emotions? When an emotional state takes over, a cascade of automatic changes occurs in mere split seconds for instance in emotional expressions in face and voice, preset and learned actions, autonomic nervous system (ANS) activity that regulates our body, regulatory patterns that continuously modify our behaviour, retrieval of relevant memories and expectations, and how we interpret what is happening within us and in the world (Ekman and Cordaro, 2011). All this suggests that a multitude of brain regions dealing with these different components is also activated. This is supported by neuroimaging findings where a set of core emotion processing regions is consistently engaged during multiple emotions (Figure 1; Phan et al., 2002; Murphy et al., 2003; Wager et al., 2003; Kober et al., 2008; Vytal & Hamann, 2010). These include cortical midline regions (Peelen et al., 2010; Chikazoe et al., 2014; Trost et al., 2012), somatomotor regions (Adolphs et al., 2000; de Gelder et al., 2004; Pichon et al., 2008; Nummenmaa et al., 2012), as well as subcortical regions including amygdala, brainstem, and thalamus (Adolphs, 2010; Damasio and Carvalho, 2013; Kragel and LaBar, 2014). None of the regions is unique to emotional processing alone but is also engaged during non-emotional processes (Salzman and Fusi, 2010). This is in line with the evolutionary psychological view to emotions, where emotions are seen as superordinate mechanisms that developed during evolution to coordinate the activity of other systems to solve adaptive problems (Al-Shawaf et al., 2016).

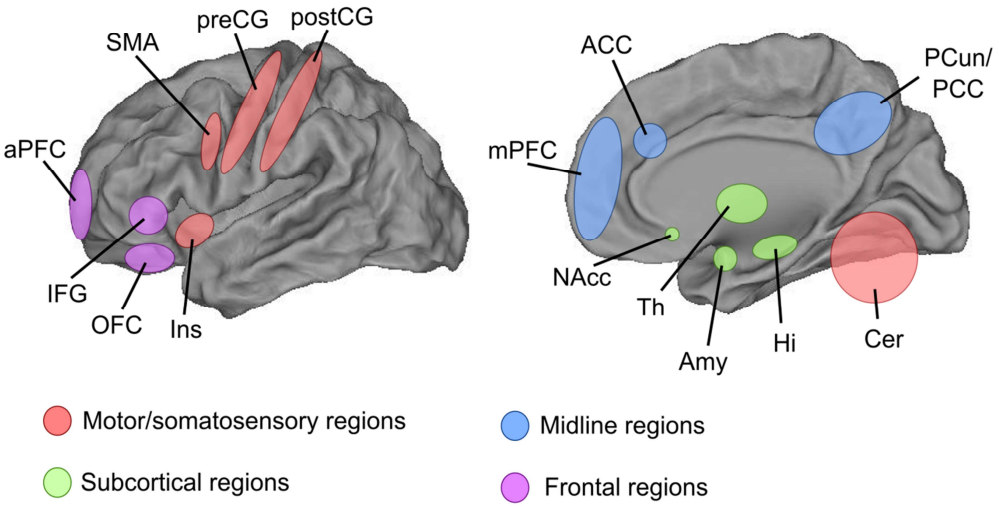


Figure 1. Schematic depiction of emotion-related brain areas. The areas are shown on their approximate locations on the PALS12 cortical atlas template (Van Essen, 2005). Most of the depicted regions are bilateral but shown on one hemisphere for simplicity. Some areas reside within the surface and subcortical regions and cerebellum are not shown on the template, therefore, areas belonging to these structures are shown on their approximate locations. Abbreviations: ACC - anterior cingulate cortex, Amy - amygdala, aPFC - anterior prefrontal cortex, Cer - cerebellum, Hi - hippocampus, IFG - interior frontal gyrus, Ins - insula, mPFC - medial prefrontal cortex, NAcc - nucleus accumbens, PCC - posterior cingulate cortex, PCun - precuneus, postCG - postcentral gyrus, preCG - precentral gyrus, SMA - supplementary motor area, Th - thalamus.

While it is clear that multiple regions are involved in emotional processing, there is growing interest in how these regions work together as circuits to produce the emotional response. For instance, Pessoa (2017) suggests that emotions should be understood in terms of large-scale network interactions spanning the entire neural system. However, the functions and organization of such emotion networks remain elusive.

1.2.2 Classification schemes for emotions

Recently, research has shed light on the organization of semantic and object categories in the human brain (Huth et al., 2012, 2016). Emotions form categories at multiple levels - including those of facial expressions, autonomic nervous system activation, subjective experience, and potentially neural basis - yet the similarities and differences between different emotion categories remain largely unexplored. Many different taxonomies of emotion categories have been proposed (see Table 1 for a summary). As the current work focuses on different emotion categories, the relevant taxonomies include especially the ones concerning different emotion families (marked in bold).

The basic emotions traditionally include at least fear, anger, disgust, happiness, sadness, and surprise (but see Ekman and Cordaro, 2011, for discussion of the number of basic emotions). They are usually characterized by distinct facial expressions, physiological activation, and subjective feelings (Ekman and Cordaro, 2011). Despite the efforts on characterizing the neural basis of basic emotions (for meta-analyses, see Phan et al., 2002; Murphy et al., 2003; Vytal and Hamann, 2010), no consensus regarding their neural substrates has been reached

Table 1. Possible ways to categorize emotional processes. Adopted from Adolphs (2002a) with additions marked with an asterisk (*). The categorizations and emotional states used in the current work are marked in bold.

Behavioral state	Motivational state	Moods, background emotions	**Basic emotions**	**Social emotions**
Approach Withdrawal	Reward Punishment Thirst Hunger Pain Craving Reproduction *	Depression Anxiety Mania Cheerfulness Contentment Worry	**Happiness** **Fear** **Anger** **Disgust** **Sadness** **Surprise**	**Pride** **Guilt** **Shame** **Maternal love** **Sexual love** Embarrassment Infatuation Admiration Jealousy **Contempt *** **Gratitude *** **Despair *** **Longing ***

(Barrett and Wager, 2006; Lindquist et al., 2012). This can be partly due to a real lack of differences. Alternatively, another possibility is that this lack of differences may be an artefact stemming from limitations, in particular 1) multiple emotions have rarely been included in the same study, 2) conventional univariate analysis methods cannot distinguish spatially overlapping neural activation patterns, and 3) functional connectivity profiles of different emotional states have been completely overlooked.

A wide array of other non-basic emotions, including 'secondary' or 'social' emotions (see reviews and proposed taxonomies in Damasio, 1999; Adolphs, 2002a) also serve adaptive survival functions and are characterized by distinctive facial expressions (Baron-Cohen et al., 2001; Shaw et al., 2005), bodily sensations (Nummenmaa et al., 2014a), and neural activity patterns (Kassam et al., 2013; Kragel and LaBar, 2015). Certain classes of emotions - the so-called social, moral, or self-conscious emotions - function explicitly to regulate social behaviour. These emotions include at least shame, embarrassment, pride, and guilt. Such social emotions require a more extensive self-representation and contextual information than does the feeling of the basic or primary emotions, as it involves representing oneself situated in a web of social relations and requires representing the internal mental states of other individuals, such as representing how others feel about oneself (Adolphs, 2002a). For instance, while basic emotions such as fear are more sensory-driven and therefore possible to elicit using simple stimuli, inducing social emotions such as embarrassment is not as simple. The psychological and neural mechanisms of these and other non-basic emotions, as well as their commonalities or differences relative to basic emotions, remain largely unresolved (Ekman, 1999; Ekman and Cordaro, 2011; Adolphs, 2002b). In particular, these emotions may involve more elaborate cognitive representations acquired through experience, education, and social norms (Panksepp and Watt, 2011), and hence, recruit brain systems partly distinct from those implicated in more “primitive” (and possibly partly innate) basic emotions. It is thus possible that also non-basic emotions may have distinct neural bases which would, however, be discernible from that of basic emotions.

1.2.3 Subjective experience of emotions

Another open question in the field of affective neuroscience is how feelings, that is, the subjective experience of an emotional state forms. If our brain is in the state of fear, how and when do we consciously interpret this state as fear? Humans are usually aware of their current emotional state, which may help to fine-tune the behavior adaptively to better match to the challenges of the environment (Damasio et al., 1996).

A few formulations exist for the causal route from emotional stimulus to subjective experience of emotion. Many psychological and neurobiological views of emotion define emotional state as the coordinated effects of multiple components including cognitive, motivational, somatic, and behavioral responses caused by an emotional stimulus, and this emotional state is consciously interpreted as a subjective feeling (Russell, 2003; Scherer, 2009; Barrett et al., 2007; Salzman and Fusi, 2010). Therefore, subjective experience results from the sum of orchestrated activation of these various component systems. An alternative view sees the causal path and, consequently, the origin of subjective experience differently: emotional stimulus affects the central emotion state of the system, which then gives rise to effects in other systems which can be observed in the form of behavior, subjective reports, psychophysiology, cognitive changes, and somatic responses (Anderson and Adolphs, 2014). Note that according to the latter view, subjective experience (as observed in subjective reports) is a direct result of the central emotion state and does not require the integration of responses from other component systems.

One way to investigate the underpinnings of subjective feelings is to focus on the similarities of brain activity patterns underlying emotions that feel either similar or different to each other. Damasio et al. (2000) suggested that emotion-dependent neural patterns across regions could explain why each emotion feels subjectively different. If the subjective experience is a sum of some basic emotion circuits, emotions that share more similar neural underpinnings should be experienced in a more similar way. Some of the aforementioned emotion categories - such as happiness, fear, disgust, love - feel more similar to each other, and some more distant: feelings of happiness and love share something in common, whereas feelings of happiness and fear seem remote. The similarities of emotional states have been previously described in terms of the similarities in experiences underlying emotions (Toivonen et al., 2012). However, it is yet unresolved whether there exist separate neural circuits responsible for different kinds of emotional behavior and experiences. For instance, it is possible that ANS activation separates only more general categories such as more exciting and calm emotions, whereas neural circuits processing facial expressions could already separate between at least the basic emotions.

1.3 Measuring brain activity and connectivity with functional magnetic resonance imaging

1.3.1 Functional magnetic resonance imaging (fMRI)

The neural underpinnings of human emotions as an object for scientific study pose a number of methodological challenges. Emotions have rapid onset but they may be relatively long lasting - even fairly mild emotional stimulation continues to affect brain activity minutes after the stimulation ends (Eryilmaz et al., 2011). Emotions modulate both subcortical and cortical brain regions implicating that the method used to measure brain activity should cover the whole brain. Finally, emotions can be challenging to induce in restricting laboratory conditions. Especially, certain emotional reactions - such as fleeing or aggression - are implausible in neuroimaging settings where participant is supposed to lie still during the scanning, and it is not easy to bring to the scanner such stimuli that would bear personal relevance that compares to the range of experiences in every-day life (Nummenmaa and Saarimäki, in press).

With these requirements in mind, functional magnetic resonance imaging (fMRI) is the most widely employed, well-suited brain imaging technique for studying human emotions in

healthy subjects. fMRI measures changes in oxygen levels of the blood which serves as an indirect measure of brain activity: it is based on the assumption that brain regions with more neural activity receive more oxygenated blood (Logothetis, 2008). Haemoglobin and deoxyhaemoglobin have opposite magnetic properties, therefore, changes in brain activity can be measured if the MRI signal is made sensitive to the haemoglobin / deoxyhaemoglobin ration. This is usually done using blood oxygenation level-dependent (BOLD) signal, which increases in active brain areas (Ogawa et al., 1990; Bandettini et al., 1992; Kwong et al., 1992). During neural activation, more blood flows to areas which require energy in form of glucose. This leads to increased blood flow and expansion of blood vessels and, therefore, to an increase in oxygenated blood and consequently an increased BOLD signal in active brain regions.

fMRI has a good spatial resolution in the range of millimeters which is especially suitable if we are interested in the spatial activity patterns. As emotions are relatively long-lasting, the ~2s temporal accuracy is enough to capture at least the slower aspects of emotional reactions, however, temporally fine-grained interplay between different brain regions cannot be measured. With fMRI, it is possible to reach activity changes in both cortical and subcortical regions. Since emotions are global events, fMRI's ability to detect activity from the whole brain is useful. Furthermore, fMRI is non-invasive and thus healthy subjects can be studied without ethical concerns in contrast to intracranial recordings that can be conducted only in specific patient populations scheduled to undergo neurosurgery.

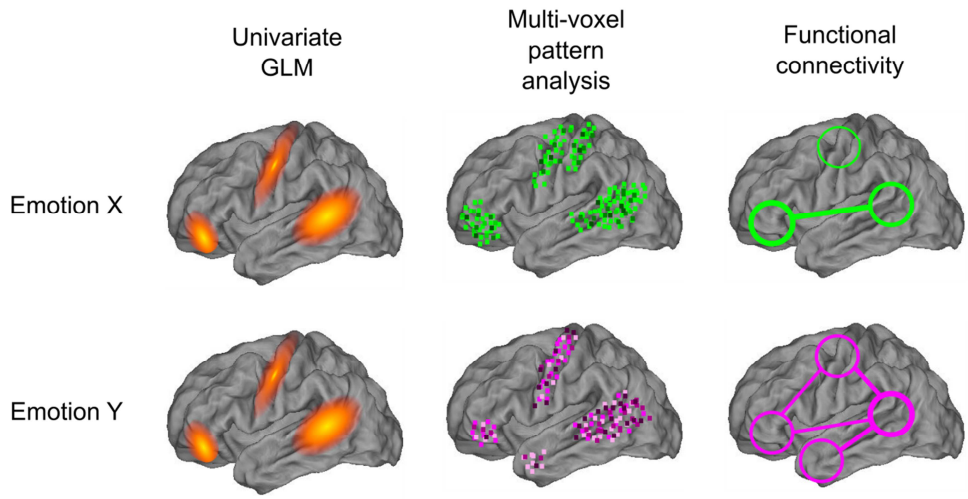


Figure 2. Comparison of fMRI analysis methods employed in the dissertation. In traditional univariate GLM, hot spots denote the strength of activation related to emotional states, which in this case are similar for both exemplar emotions. However, multi-voxel pattern analysis can detect the underlying differences of pattern structure for both emotions and also patterns in areas that are below the statistical significance threshold in GLM, i.e., it is assumed that it is the distributed pattern of activations and deactivations that underlies a given emotional state. Finally, functional connectivity analysis, in turn, detects the co-activation between different brain regions.

Traditionally, fMRI data has been analyzed using univariate methods. For instance, the widely used general linear model (GLM) includes fitting a stimulus model time series to the time series of each voxel and testing the model fit using parametric or nonparametric statistics. This approach considers each voxel independently and does not take into account the multivariate nature of the fMRI data. It is possible that the combined activation values from multiple voxels are important for the function we are interested in, and go unnoticed with traditional GLM. As such, GLM measures the net activation within an area. However, same net

activation can result from different configurations of individual voxel activations (Figure 2). Moreover, GLM ignores how different brain regions work together. Therefore, the current work takes advantage of multivariate methods including multi-voxel pattern analysis (MVPA), representational similarity analysis (RSA), and functional connectivity analysis to reveal the differences in fine-grained patterns and connectivity of brain areas between specific emotions.

1.3.2 Multivariate pattern analysis (MVPA)

Pattern recognition, often called as decoding, is a branch of machine learning that focuses on the recognition of patterns and regularities in data (Bishop, 2006). These types of algorithms generally aim to perform the most probable matching of an input to an output, taking into account the statistical variation in inputs. A successful solution to the classification problem is defined as the classifier's ability to predict the underlying true input category with above chance performance. The idea that fMRI data analysis can be defined as a pattern-classification problem where we try to recognize a pattern of brain activity as being associated with one mental state versus another has led to an enormous increase in pattern classification applications in fMRI (Norman et al., 2006). Unlike classical univariate methods, multivariate approaches such as multi-voxel pattern analysis (MVPA) extract information based on the complete pattern of brain activity, rather than intensity differences between individual voxels (Haxby et al., 2001; Norman et al., 2006). Thus, they can overcome the limitations of univariate methods in identifying differences in activation patterns that overlap spatially between conditions such as brain circuits involved in generating multiple emotional states. Furthermore, while univariate methods focus on average activity of a brain region across multiple repetitions of a stimulus condition, thus creating a statistical summary of the corresponding experimental condition, MVPA considers more fine-grained patterns within that brain region and can reveal more information regarding the condition-specific voxel activity changes (Cox et al., 2003; Kriegeskorte et al., 2006). MVPA typically involves implementing a machine learning algorithm that tries to learn associations between a set of *a priori* categories and the multivariate data patterns associated with them. The algorithm is then tested on an independent dataset that was not used in training the algorithm to see whether the differences between categories are consistent and to avoid peeking of the data.

Pattern recognition analyses are theoretically well-suited for tackling the organization of human emotions in the brain. First, emotions cause changes across the brain, and pattern recognition techniques are focused on analysing system-level patterns (e.g. Kober et al., 2008). Second, successful classification of two or more emotional states would require that they elicit consistently different activation patterns, thus allowing testing the distinctness of different emotional states. Third, successful cross-validation of the classifier across independent samples of subjects would require consistent emotion-specific activations across individuals, possibly providing evidence for the biological versus acquired basis of emotions. Fourth, investigating the similarities and distances of the emotion-evoked patterns across a large array of emotions allows revealing the (categorical or dimensional) structure of the emotion space. Preliminary evidence suggests that this approach can be applied to classify sensory emotional signals conveyed by visual (Peelen et al., 2010) and auditory (Ethofer et al., 2009) cues and in the absence of external stimuli (Kassam et al., 2013). However, previous studies have applied classification to a restricted set of voxels or performed classification

between only two emotional states at the time. Consequently, the large-scale brain networks supporting multiple emotional states remain poorly understood.

1.3.3 Functional connectivity

Different brain regions work together as a network to support information processing. While traditional univariate fMRI analyses investigate the specialization of brain regions for some aspects of a mental function, which can be called functional segregation, functional connectivity investigates functional integration: the study of connected processes (Friston, 2011). Functional connectivity is defined as the statistical association or dependency among two or more anatomically distinct time-series (Friston et al., 1996). Functional connectivity analysis thus measures how different brain regions are linked together during different tasks. Usually, connectivity is defined as correlation between the time series of two voxels or regions. Note that changes in correlation of the time series do not themselves indicate the direction or neurochemistry of causal influences, structural connectivity, or synaptic connections between brain regions, that is, they do not tell how regions are coupled. However, they indicate functional interactions between regional systems that contribute to a shared task.

Despite suggestions that emotions require activation of large-scale brain networks (e.g. Hamann, 2012; Pessoa, 2017), the connectivity modulations by emotional content and especially the differences in connectivity between emotions have been largely overlooked in affective neuroscience. Experience of separate emotional states likely stems from differences in the functional interplay and connectivity between widespread brain regions, which we can further investigate using functional connectivity analyses. So far only a handful of studies have compared how specific emotions modulate functional brain connectivity either by focusing on a limited set of *a priori* defined brain regions (Eryilmaz et al., 2011; Tettamanti et al., 2012; Raz et al., 2016) or by looking at emotion-specific intrinsic connectivity (Touroutoglou et al., 2015), which tends to remain largely unaffected by differences between mental tasks in general (Cole et al., 2014).

1.3.4 Representational similarity analysis (RSA)

Representational similarity analysis (RSA; Kriegeskorte et al., 2008) provides another alternative to investigate multivariate patterns in BOLD-fMRI data and other domains. It allows combining data from different modalities such as from neuroimaging, behavioral ratings, or theoretical models. Data from different domains is first transformed into separate dissimilarity matrices that, in the case of fMRI, describe the dissimilarity of neural activity patterns between each pair of stimulus conditions. Using representational similarity analysis, we can directly compare how similarities in neural activity correspond to similarities of subjective experience or investigate how different models fit to the neural or behavioral data. Practically any kind of data - single-cell recordings, behavioral ratings, MEG or fMRI data, theory-based models - can be described as a dissimilarity matrix and compared to each other using correlation of the dissimilarity matrices which reveals whether similarities in stimuli in one domain correspond to the similarity structure in the other domain.

In affective neuroscience, RSA appears as a promising technique for comparing different theoretical models and combining data across domains that previously have been difficult to pair, such as neuroimaging data and animal studies. However, applications have so far been sparse. A recent study showed that similarities in emotional valence correspond to the simi-

larities in neural activation in orbitofrontal cortex (OFC) but not in similarities in neural activation in parts of occipital or temporal lobes (Chikazoe et al., 2014). However, the similarity structure of specific emotion categories remains unexplored.

1.3.5 Naturalistic stimuli

With the development of more advanced computational methods such as the multivariate and connectivity methods described above, we can now use more complex stimulation during the fMRI scanning (see e.g. Hasson et al., 2004; Jääskeläinen et al., 2008). Traditional GLM analysis requires an event-related experimental design with simple emotion elicitation using for instance sounds, short movie clips, or pictures. With more advanced methods, we can move towards more naturalistic stimuli such as longer movies, narratives, and emotional imagery. For affective neuroscience, these provide a way to elicit more natural emotions during fMRI scanning. For instance, recent studies have elicited strong emotions during fMRI using emotional movies (Nummenmaa et al., 2012; Tettamanti et al., 2012). Compared to static images such as emotional facial expressions or pictures depicting emotional events, naturalistic stimuli have been shown to elicit stronger and more vivid emotions that engage widespread brain emotion circuits (Costa et al., 2010; Nummenmaa et al., 2012, 2014b).

2. Objectives

The goal of this dissertation was to investigate the neural underpinnings of different emotional states using multivariate methods. We took advantage of state-of-the-art analytic techniques and naturalistic stimulation setups to elicit strong and reliable emotional states during fMRI scanning. We focused on four research questions tackled in five studies (Table 2):

In research question 1, we asked whether different emotions have distinguishable neural bases. We hypothesized that emotions that subjectively feel different also have distinct, robust, and identifiable neural bases. If this hypothesis holds, we should be able to successfully classify emotions based on their neural activity patterns.

In research question 2, we investigated the core regions supporting emotions. We hypothesized that different emotional states would be characterized by widely-spread changes in brain activity, and that this activity would overlap between different emotional states.

In research question 3, we asked how the large-scale brain connectivity is modulated by different emotional states. We hypothesized that the core regions identified in research question 2 would be connected and contain emotion-specific connectivity patterns.

Finally, in research question 4, we returned to our starting point by asking how the neural similarity of emotions relates to their subjectively felt similarity. We hypothesized that the emotions that feel more similar also share more similar neural underpinnings.

Table 2. Research questions. Summary of the research questions and how different studies were designed to answer them.

Research question (RQ)	Study I	Study II	Study III	Study IV	Study V
RQ1: Do different emotions have distinct neural bases?	x	x			
RQ2: What are the core regions supporting emotions?	x	x	x		
RQ3: How does the large-scale functional connectivity vary during emotions?			x	x	
RQ4: How does the neural similarity of emotions relate to their subjectively felt similarity?	x	x			x

3. Methods

3.1 Participants

Participants were altogether 109 healthy adults aged from 19 to 38 years (see Table 3). Participants were all right-handed and Finnish speaking, with no reported neurological disorders. All participants gave their informed consent according to the Declaration of Helsinki. All studies were approved by the Aalto University ethical committee.

Table 3. Summary of Studies I-V. The number and age of participants and a description of the study type in Studies I-V.

	Study I			Study II	Study III	Study IV	Study V
	Movie experiment (Exp. 1)	**Imagery experiment (Exp. 2)**	**Crossmodal experiment (Exp. 3)**				
Participants	21 healthy adults (12 males) ages 19–33 years, mean age 24.9. years	14 healthy females ages 19–30 years, mean age 23.6 years	13 healthy females ages 21–29 years, mean age 25.4 years	25 healthy females ages 19–38 years, mean age 23.6 years	20 healthy adults (8 males) ages 19–30 years, mean age 25 years	16 healthy females ages 20–30 years, mean age 24.3 years	
Study type	fMRI study	fMRI study	fMRI study	fMRI study	fMRI study	fMRI study	Review with meta-analysis

3.2 Stimuli

We induced emotions in human participants using naturalistic stimuli including movies, mental imagery, and guided imagery based on narratives (see Figure 3).

In Study I, we employed movie clips and mental imagery to elicit emotions during fMRI. In the Movie experiment part of the study (Exp. 1), we induced disgust, fear, happiness, sadness, and a neutral state using 10-s movie clips (10 per category) chosen from a video database (Tettamanti et al., 2012). We presented the clips without sound to avoid attentional and linguistic confounds, as most movies contained English speech and the participants were native Finnish speakers. The participants were instructed to view the movies similarly as they would watch TV. Each clip was preceded by a 5-s fixation cross and followed by a 15-s washout period. In the Imagery experiment part of the study (Exp. 2), we induced emotions using mental imagery. Before the fMRI scanning, the participants were given a list of 36 emotion words each representing a variant of one of six emotion categories (anger, fear, happiness, sadness, disgust, and surprise) and were asked to devise and practice their own method to elicit each emotion in the list. Sample methods of emotion elicitation (such as imagining a past event, thinking about a corresponding movie scene, or recreating the bodily state associated with

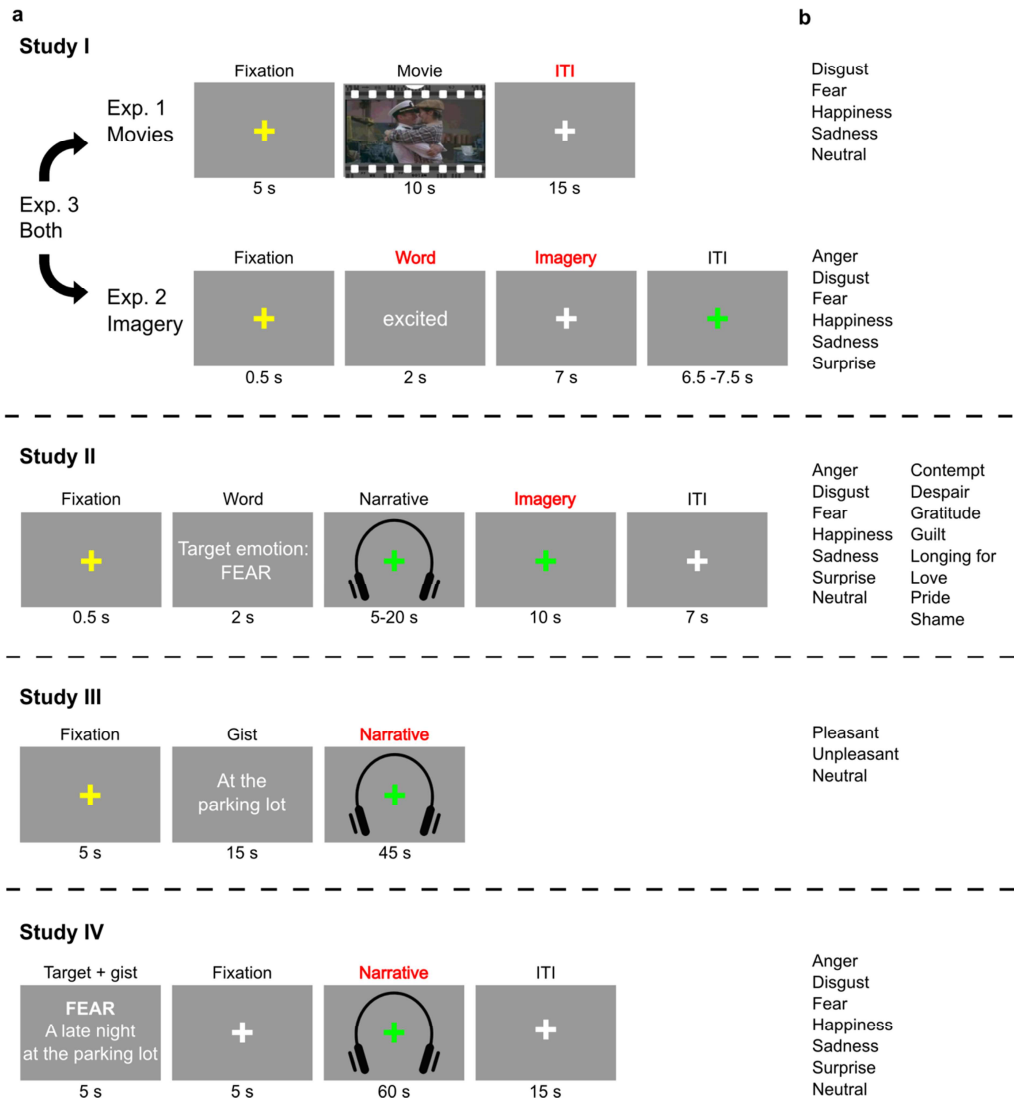


Figure 3. Experimental designs and emotion categories in Studies I-IV. (a) Experimental design for Studies I-IV. Time period marked in red was used in the analyses. See text for details. (b) Emotion categories used in Studies I-IV.

the emotion) were provided but participants were free to choose whatever method they considered best for each emotion. Participants were asked to practice the imagery task at home for at least 1 hour prior to the fMRI experiment and again immediately before the fMRI scanning. During the scanning, each trial began with a fixation cross shown for 0.5 s, followed by the presentation of the word for 2 s, and an imagery period of 7 s. Participants were instructed to imagine the emotional state described by the emotion word they saw and to continue imagery until the subsequent intertrial interval. In Exp. 3, we use trial designs from both Movie and Imagery experiments of the study in separate runs to elicit emotions in same participants using both elicitation techniques.

For Study II, we created sixty 5-20 second long auditory narratives to support guided affective imagery. Each narrative elicited primarily one out of possible 14 emotions or a neutral

state. Participants were instructed to imagine that the events of the narrative happen to them. Targeted emotions included six basic or primary emotions (anger, fear, disgust, happiness, sadness, and surprise) and eight social or secondary emotions (shame, pride, longing, guilt, love, contempt, gratitude, and despair). The narratives were spoken by a female speaker using neutral prosody without cues for the affective content of the narrative. Each trial started with a fixation cross shown for 0.5 seconds, followed by a 2-s presentation of a word describing the target emotion, a spoken narrative, and a 10-s imagery period. The trial ended with a 10-s wash-out period to counter for possible carryover effects.

In Study III, we developed thirty 45-s-long spoken narratives describing unpleasant, neutral, and pleasant events (10 stories per each category). The recorded narratives were read by a neutral female voice that provided no prosodic cues for the affective significance of the story contents. During fMRI, participants were instructed to listen to the narratives similarly as if they would listen to radio or a podcast, and to try to get involved in the stories by imagining the described events vividly. Each narrative was preceded, for 5 s, by a fixation cross and, for 15 s, by a short text that explained the general setting of the forthcoming narrative without revealing its actual content. The 45-s-long narrative was followed by a short wash-out period.

In Study IV, emotions were induced using thirty-five 60-s-long narratives triggering six emotional states (anger, fear, disgust, happiness, sadness, surprise) and a neutral state. The narratives described personal life events spoken by a female speaker with varying emotional prosody. A trial started with a fixation cross presented for 5 seconds. It was followed by a 5s presentation of the target emotion (e.g. 'happy') and a short description of the narrative gist (e.g. 'lovers under a tree'), after which a fixation cross appeared on the screen and the 60-second-long narrative was played through earphones. The trial ended with a 15s wash-out period. Subjects were instructed to listen to the narratives similarly as if they would listen to their friend describing a personal life event.

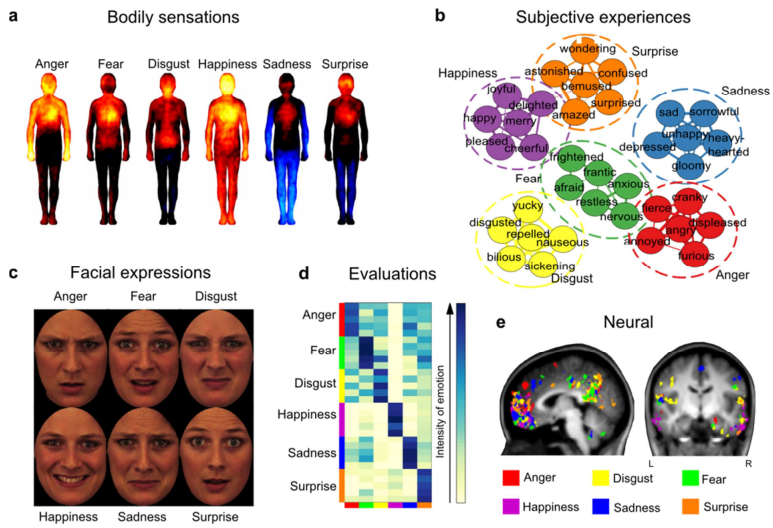


Figure 4. Summary of the datasets used in the meta-analysis in Study V. (a) Self-reported bodily sensations corresponding to different emotions (Nummenmaa et al., 2014a). (b) Rated similarity of subjective experiences related to 36 different emotions (Saarimäki et al., 2016). (c) Ratings for the emotion best corresponding to each facial expression (Calvo and Lundqvist, 2008). (d) Evaluations of intensity of emotions evoked by short emotional narratives (Nummenmaa et al., 2014a). (e) Voxels contributing most significantly to pattern classification of emotions from BOLD-fMRI data (Saarimäki et al., 2016).

In Study V, we collected existing datasets from three published studies (Nummenmaa et al., 2014a; Saarimäki et al., 2016; Calvo and Lundqvist, 2008). The data included similarity values of facial expressions (human observers' confusions between facial expression categories), bodily sensations (Linear Discriminant Analysis classifier run on the bodily maps of emotions), evaluations (Euclidean similarity of intensity profiles of discrete emotion ratings for short narratives), subjective experiences (direct pairwise ratings of emotion concepts), and neural data (confusions of a pattern classifier on BOLD-fMRI data) underlying six basic emotions: anger, fear, disgust, happiness, sadness, and surprise (Figure 4).

3.3 Measuring subjective emotional experiences

In all experiments, ratings of emotional qualities of the stimuli were acquired post-experiment rather than during fMRI, as a reporting task is known to influence neural response to emotional stimulation (Hutcherson et al., 2005; Lieberman et al., 2007) and as repeating a specific emotional stimulus has only a negligible effect on self-reported emotional feelings (Hutcherson et al., 2005).

3.3.1 Similarity ratings

In Exp. 2 of Study I and in Study II, we collected similarity ratings of the emotions used in the study (36 variants from 6 categories in Exp. 2 of Study I, 14 emotions and a neutral state in Study II) using direct pairwise ratings. After the fMRI scanning, participants were shown one pair of emotion words at the time and asked to rate the similarity between the subjective emotional experiences related to these emotions (ranging from no similarity [0] to full similarity [5]). The ratings were then scaled to range between 0 and 1.

3.3.2 Emotion intensity ratings

In Studies I and II, we measured the intensity of the specific emotions evoked during the fMRI scanning. In Exp. 1 of Study I, the participants viewed the movie clips again and chose the emotion (disgust, fear, happiness, sadness, neutral, anger, surprise) that best described their feelings during each movie. They also rated the intensity (1–9) of the experienced emotion. In Exp. 2 of Study I, the participants rated the intensity (1-9) of the elicited emotion. In Study II, the participants listened to the narratives again and, for each narrative, rated how strongly they felt each of the possible 14 (and neutral state) emotions using a scale ranging from 0 (not at all) to 9 (very much), which was for further analyses scaled to range from 0 to 1.

3.3.3 Valence and arousal ratings

In Study III and IV, we collected ratings of valence and arousal content of the stimuli. In Study II, participants listened to the narratives after the fMRI experiment and rated the valence and arousal of each narrative using a scale ranging from 0 (negative valence/low arousal) to 9 (positive valence/high arousal). In Study III, we collected continuous ratings of valence and arousal of the narratives on separate runs after the fMRI scan. While listening to each narrative, participants used a mouse to move a small cursor at the edge of the screen up

and down in order to indicate their current experience; data were collected at 5 Hz. The actual valence–arousal scale was arbitrary, but for the analyses the responses were rescaled to range from 1 (negative valence/low arousal) to 9 (positive valence/high arousal).

3.4 Measuring brain activity: Functional magnetic resonance imaging

3.4.1 (f)MRI data acquisition and preprocessing

In all studies, MRI data were collected with a 3T Siemens Magnetom Skyra scanner at the Advanced Magnetic Imaging Centre (Aalto NeuroImaging, Aalto University) using a 20-channel Siemens volume coil. Whole-brain functional scans were collected using a whole-brain T2*-weighted EPI sequence with the following parameters: 33 axial slices, TR = 1.7 s, TE = 24 ms, flip angle = 70°, voxel size = 3.1 × 3.1 × 4.0 mm^3, matrix size = 64 × 64 × 33, FOV 198.4 × 198.4 mm^2, using ascending interleaved acquisition with no gaps between slices. A custom-modified bipolar water-excitation radio-frequency pulse was used to avoid signal from fat. High-resolution anatomical images with isotropic 1 × 1 × 1 mm^3 voxel size were collected using a T1-weighted MP-RAGE sequence. Standard preprocessing of fMRI data in all studies included slice timing correction, motion correction, non-brain matter removal, and high-pass temporal filtering.

For voxel-based pattern classification (Studies I and II), we created participant-wise gray matter masks from the T1-weighted images and transformed them to the 64 × 64 × 33 native space for within-participant classification, and to the 2-mm Montreal Neurological Institute (MNI) 152 standard space template for across-participants classification. For univariate GLM analyses (Studies II and III), the preprocessed data were registered to the 2-mm MNI 152 standard space template and spatial smoothing was applied. For functional connectivity analyses (Studies III and IV), respiratory and heart rate signal were removed from fMRI data using the DRIFTER toolbox (Särkkä et al., 2012). Functional data were registered to the 2-mm MNI 152 standard space template, detrending was performed, and spatial smoothing was applied.

3.4.2 Multivariate pattern analysis

Voxel-based within-participant classification

In Studies I and II, classification of emotion categories within participants (Figure 6) was performed using each participant's data in native space. We used the whole-brain data since recent studies have shown that emotional processing relies on large-scale cortical and subcortical circuits, rather than on isolated regions (Kober et al., 2008; Vytal and Hamann, 2010; Nummenmaa et al., 2012). Voxels outside gray matter were masked out and the functional data were temporally normalized to a mean of zero and unit variance in each voxel by subtracting the mean response across all categories. Feature selection was performed using ANOVA to select voxels with a significant ($p<0.05$) main effect for emotion, i.e., to select the voxels whose mean activation differed between at least some of the included emotion conditions. Finally, the hemodynamic lag was corrected by convolving the category regressors with the canonical double-gamma hemodynamic response (HRF) function and thresholding the convolved regressors using a sigmoid function to return the regressors to the binary form.

Classification was performed on the standardized, HRF-convolved fMRI volumes from the imagery period following the movie, word, or narrative stimulus to extract only brain activity

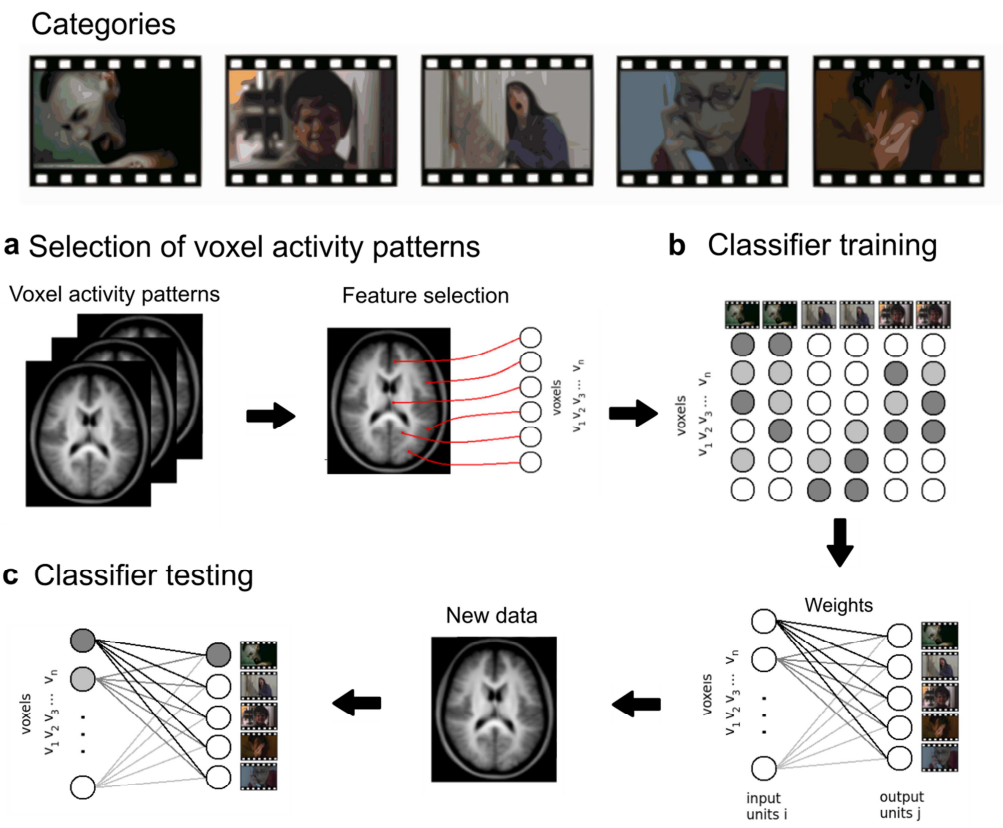


Figure 5. Pipeline for the within-participant classification. (a) Selection of voxel activity patterns for the classifier. In the current work, we used both whole-brain activity patterns (voxels within the grey matter) and a selection of regions-of-interest. Feature selection was performed with voxel-wise ANOVA to remove voxels whose activity did not differ between emotion conditions. (b) Classifier training. The voxel activations were fed as an input for the classifier, here, a linear multi-class classifier implemented as a neural network without hidden layers which, during training, learns the weights from each voxel to a specific category. We used both within-participant classification (with leave-one-run-out cross-validation) and across-participants classification (with leave-one-participant-out cross-validation). (c) In the classifier testing, the classifier algorithm is presented with new data and asked to guess the category it belongs to. The classifier performance can then be evaluated as the number of correct guesses per category and on average.

related to the emotion, and to minimize the activity differences related to the acoustic, semantic, and visual features of the stimuli. A linear neural network classifier without hidden layers was trained to recognize the correct emotion category out of the possible ones (multiclass classification, see Polyn et al. 2005). The classifier starts with random weights from input *i* (voxels) to output *j* (categories). During training, the weights are adjusted for each given input with scaled conjugate gradient algorithm for optimization and mean squared error as an error function. During testing, each input is mapped to values from 0 to 1 for each output category using logistic functions. This corresponds to the confidence that the input belongs to a specific category. In all experiments, the classifier was trained using a leave-one-run-out procedure where training was performed with $N - 1$ runs and testing was then applied to the remaining one run. Cross-validation was performed across all runs and the participant-wise classification accuracy was calculated as an average percentage of correct guesses across all the cross-validation runs. Naïve chance level was derived as a ratio of 1 over the

number of categories. To test whether classification accuracy exceeded chance level, we used permutation tests to simulate the probability distribution of the classification. Each permutation step included shuffling of category labels of the training set (across training set runs) and re-running the whole classification pipeline, repeated 1,000 times for each subject. FDR correction at $p<.05$ was used for multiple comparisons.

To visualize the brain regions contributing most to the classifier's selection of each emotion category, voxel-wise importance values were calculated and plotted separately for each category. Importance values were calculated by defining importance $imp = a \times w$, where a is the activation of a voxel for a specific category and w is the trained weight from this voxel assigned to a specific category (Polyn et al., 2005). This method reveals which voxels are most important in driving the classifier's output for a specific category, and it highlights voxels that have concordant activation values and weights. Participant-wise importance maps were first calculated using the mean importance values over cross-validation runs and subsequently registered to MNI space. Then, mean importance maps were calculated across all participants for each emotion. These maps were plotted on a standard brain volume after selecting the highest 10 000 importance values (corresponding to ca. 1%). Clusters smaller than 27 (3 × 3 × 3) voxels (216 mm^3) were excluded from visualizations. It should be noted that all voxels that passed the feature selection were taken into account in the classification and the importance maps simply highlight the most important clusters of voxels.

Voxel-based across-participants classification

To test whether the neural signatures of different emotions generalize across participants, we ran whole-brain across-participants MVPA in Studies I and II. This analysis was performed with the same steps as the within-participant classification but using fMRI data that was registered to MNI space with 2-mm isotropic voxels. For each experiment, a linear classifier was trained using a leave-one-participant-out procedure where the training was performed with N – 1 participants and the testing of the classifier with the remaining one participant. Cross-validation was then performed across all participants, and the classification accuracy was calculated as an average percentage of correct guesses across all the cross-validation runs.

Voxel-based crossmodal classification

Study I also included an experiment where emotions were induced in same participants using both movies and mental imagery (Exp. 3). For this combined crossmodal experiment, we trained a classifier using the imagery periods following the movies and words. The classifier was trained with either the movie data and tested with the imagery data, or vice versa for cross-validation, and it was trained to select the correct category out of four possible ones (disgust, fear, happiness, sadness).

Voxel-based region-of-interest classification

In addition to the whole-brain analyses described above, we also applied a region-of-interest (ROI) analysis to test whether the BOLD signal in any of our *a priori* -defined ROIs would allow a reliable classification of the emotional states when considered alone. ROI analyses were performed in Studies I and II. Cortical regions showing consistent emotion-related activation in the literature were selected as candidate ROIs for coding emotional content (Kober et al., 2008; Vytal and Hamann, 2010): orbitofrontal cortex (OFC), anterior prefrontal cortex (aPFC), inferior frontal gyrus (IFG), insula (Ins), anterior cingulate cortex (ACC), posterior cingulate cortex (PCC), medial frontal cortex (MFC), precuneus (PCun), paracingulate

gyrus (PAC), precentral gyrus (preCG), supplementary motor area (SMA), and postcentral gyrus (postCG). The subcortical regions were amygdala (Amy), nucleus accumbens (NAcc), hippocampus (Hi), and thalamus (Th). Bilateral masks for these ROIs were first defined in MNI standard space using the Harvard-Oxford cortical and subcortical atlases (Desikan et al., 2006) and then transformed into native space. Feature selection and classifier training was then performed for each ROI separately similarly to the whole brain analyses.

3.4.3 Functional connectivity

Modelling emotion-dependent functional connectivity

In Study III we assessed whether time-variable valence and arousal are associated with changes of functional connectivity in large-scale brain networks. Data were first down-sampled to isotropic 6 x 6 x 6 -mm³ voxels and voxels outside the gray matter were masked out. To reveal the pairs of regions for which the dynamic connectivity depended most strongly on valence and arousal, we computed instantaneous seed-based phase synchronization (SBPS; Glerean et al., 2012) as a time-varying group measure of connectivity between every pair of voxels. We then extracted valence and arousal regressors from behavioral ratings, gamma-convolved them used them to predict each connection's time series in the general linear model (GLM) to assess the positive and negative effects of valence and arousal on functional connectivity. The mean voxel-wise connectivity changes were stored in connectivity maps, where link intensities reflect the degree to which SBPS is dependent on valence and arousal. Statistically significant functional connections were plotted on cortical flatmaps using the Gephi software (Bastian et al., 2009). Statistical significance of the association between emotion ratings and SBPS time series was based on a nonparametric voxel-wise permutation test for *r* statistic (Kauppi et al., 2010). We approximated the full permutation distribution independently for each connection with 10,000 permutations per connection using circular block resampling (Politis and Romano, 1992). Due to the large number of links, we used positive FDR (Storey and Tibshirani, 2003) of $q < 10\%$ to control false discovery rate for the connectivity time-series; this choice is equivalent to the cluster network correction, which takes into account the large number of links in the network without being overly conservative (Zalesky et al., 2010).

Connectivity-based pattern classification

In Study IV, we tested whether functional connectivity patterns underlying different emotions can be separated using pattern classification (Figure 10). To create the functional networks for classification, we selected 264 nodes based on a functional parcellation (Power et al., 2011) and extracted the BOLD time course for each node. For each of the 35 emotional narratives belonging to six emotion categories or to the neutral state (thus totaling 5 narratives per category), we calculated the Pearson correlation coefficient between the BOLD time course of each of the nodes during the 60-s-long story, which resulted in a connectivity matrix of 264 x 264 nodes for each narrative. Next, we removed the baseline connectivity pattern from emotion-wise connectivity matrices by taking the average of the five neutral connectivity matrices and subtracting it from each of the remaining 30 connectivity matrices separately using linear regression, and kept the residuals in the connectivity matrices. In addition to the full network of 264 x 264 nodes, we extracted also subnetworks based on the 10 functional systems of interest as proposed by Power et al. (Power et al., 2011). The included subnetworks were motor and somatosensory (35 nodes), cingulo-opercular (14 nodes), audi-

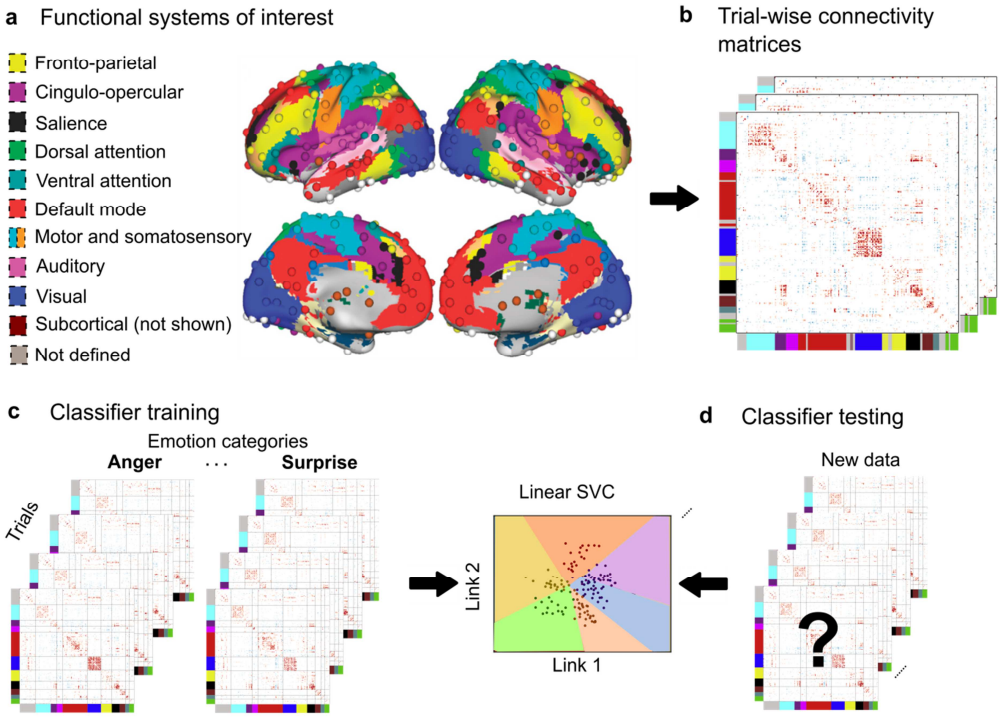


Figure 6. Pipeline for classification of functional connectivity patterns. (a) Selection of nodes for the connectivity analysis. Functional connectivity is calculated between selected nodes, which in the current work consist of 264 nodes (denoted by dots) belonging to ten functional brain systems (denoted by colors) from Power et al. (2011). Adopted from Cole et al. (2013) with permission. (b) Extraction of trial-wise connectivity matrices. Here, connectivity matrices for each trial are calculated using Pearson correlation between each pair of 264 node time series for each subject and for each 60-s narrative. (c) Classifier training. The connectivity matrices are fed as input for the classifier, here, a linear support vector classifier. Here, we used an across-participants classifier with leave-one-participant-out cross-validation. (d) Classifier testing. The testing is performed with new data, in this case, the left-out participant. The classifier performance was evaluated by calculating the classification accuracy (defined as percentage of correct classifier guesses per category) and the confusion matrix (predicted vs. true labels of each category).

tory (13 nodes), default mode (58 nodes), visual (31 nodes), fronto-parietal (25 nodes), salience (18 nodes), subcortical (13 nodes), ventral attention (9 nodes), and dorsal attention (11 nodes) networks.

The classification of emotion categories was then performed across participants. We trained a between-subjects support vector machine classification algorithm with linear kernel to recognize the correct emotion category out of 6 possible ones (anger, disgust, fear, happiness, sadness, surprise). Naïve chance level was defined as the ratio of 1 over the number of categories (14.2%). The samples for the classifier consisted of the 30 connectivity matrices (5 matrices for each emotion category) from each subject, resulting in altogether 480 samples (80 per category). A leave-one-subject-out cross-validation was performed and the classification accuracy was calculated as an average percentage of correct guesses across all the cross-validation runs. For full network classification, we included the full connectivity matrix of each sample. For subnetwork (region-of-interest) classification, we included the connectivity matrix of each sample either within one subnetwork or between two subnetworks. A separate classifier was trained for each within/between subnetwork division. Based on subnetwork classifier results, we wanted to investigate the default mode system subnetworks in more de-

tail. Therefore, we split the default mode system into separate subnetworks and trained a separate classifier for each within/between subnetwork. P-values were computed with permutations by generating 5,000 surrogate accuracy values for full network and for each subnetwork. The null cumulative distribution function was obtained using kernel smoothing. Multiple comparisons were corrected for by using FDR correction (Benjamini and Hochberg, 1995).

3.4.4 Representational similarity analysis (RSA)

In Study I, we used RSA to investigate whether the neural similarities between different variants of basic (anger, fear, disgust, happiness, sadness, surprise) emotional states correspond to their experiential (subjectively felt) differences. To construct a neural similarity matrix, we trained a within-participant classifier to separate between brain responses to all 36 emotion categories in Exp. 2 of Study I and computed the mean confusion matrix across the basic emotion categories for each participant. All other classifier parameters remained as in the between-category classification described above. As an indicator for neural similarity, we then averaged these confusion matrices across participants and averaged the upper and lower triangles to make the matrix symmetrical and to estimate the mean confusion regardless of which category was the target and which was the guess. Experiential similarity matrices were extracted from behavioral similarity ratings, thus each link denotes the experienced similarity between a pair of emotions. To examine the correlation between the two similarity matrices, we applied the Mantel test to examine the correlation between the two similarity matrices using an in-house algorithm (available at http://users.aalto.fi/~eglerean/permutations.html, last accessed on June 15, 2017). The probability distribution was obtained with permutation repeated for 10^6 times.

In Study II, we used RSA to examine the similarity structure of brain responses and subjective feelings associated with both basic and social emotions. To construct the neural similarity matrix, we took the classifier confusion matrix from the whole-brain within-participant classification of 14 emotions and a neutral state. From the group-averaged confusion matrix, we calculated a distance matrix by taking the category confusion vectors for each pair of emotions and by calculating the Euclidean distance between these vectors (see Reyes-Vargas et al., 2013). Experiential similarity matrices were again calculated based on pairwise similarity ratings of experienced emotions and averaged over the participants. Subsequently, the mean neural and subjective similarity matrices were correlated using Spearman's rank correlation coefficient. The p level for the Spearman test was obtained with a permutation test by shuffling the neural matrix and re-calculating the correlation for 10^6 times.

Finally, we visualized how different emotions cluster together based on their neural similarities and whether same cluster structure is present at the level of subjective experience. For this, we employed hierarchical cluster analysis. To visualize the similarities in subjective and neural organization of emotions, we extracted the clusters in both neural and behavioral data, and subsequently plotted the cluster solutions using alluvial diagrams (Rosvall and Bergstrom, 2010).

3.4.5 Univariate GLM analyses

Univariate general linear model (GLM) analyses were employed mainly to complement multivariate analyses. As visualization of results is one of the caveats of pattern classification, we

took advantage of the more conventional GLM analysis for mapping the emotion-dependent responses in the brain. Voxel-based GLM tests for the null hypothesis that the time course of the experimental manipulation (in the current study, usually a particular emotion) is not related to the time course of the activation at each voxel. In functional connectivity analyses, we can contrast the connectivity matrices for different conditions, here, emotions, with that for the other conditions by using pairwise *t* tests for each link at the time.

In Study II, we visualized the emotion-related activation by calculating cumulative activation maps. We first ran separate GLMs to compare each of the 14 emotions versus the neutral baseline: first-level GLM analysis was performed to obtain contrast maps for each participant, and second-level (i.e., subjects as the random factor) analysis was run to reveal emotion-specific activations at the population level. Cluster correction for multiple comparisons was employed at $p<0.05$ (Eklund et al., 2016). Next, we qualitatively summarized the results across emotions by calculating a cumulative map where each voxel shows the number of emotions showing statistically significant activation in the random effects model (cluster corrected at $p<0.05$). Similar analysis was also run for deactivations. The resulting cumulative activation / deactivation maps show the number of emotions for which the voxel activates at the given threshold, but does not reveal which emotions are activating the voxel. To answer this, we used hierarchical clustering of emotions (see section 3.4.4) to obtain three principal clusters of emotions and mapped them on the cortical and subcortical maps using R, G, B color space. For each emotion, we took the unthresholded second-level *t* maps obtained from the GLM analysis that approximate the average activation for each emotion, summed them for emotions belonging to the same cluster, and assigned the summed values to the corresponding R, G, B channels. The color channels were subsequently visualized in MNI space. Consequently, the RGB color at each voxel reflects the cluster distribution of that voxel, and can be used for localizing brain regions contributing to different emotions.

In Study III, we used univariate GLM to model the effects of emotional valence and arousal on brain activity. For each participant, we used the GLM to assess regional effects of the valence and arousal parameters on BOLD activation. The model included the orthogonalized valence and arousal time series obtained from average behavioral ratings. Individual contrast images were generated for the activation and deactivation effects of valence and arousal. The second-level analysis used these contrast images to generate statistical *t* maps (FDR corrected at $p<0.05$ for multiple comparisons).

4. Results

4.1 RQ 1: Do different emotions have distinct neural bases? (Studies I and II)

Different emotions - both basic (primary) and non-basic (secondary, including social emotions) - form categories at the level of subjective experience, facial expressions, and bodily sensations. In Studies I and II we investigated whether different emotions are also characterized by distinct neural bases. We hypothesized that specific emotions could be classified using brain activity patterns, which would support the view that they have unique neural underpinnings.

To elicit emotional states from various emotion categories, we induced a set of canonical basic emotions (anger, fear, disgust, happiness, sadness, and surprise; Studies I and II) and non-basic emotions (pride, gratitude, love, contempt, guilt, shame, longing, despair; Study II) in participants while their brain activity was measured with BOLD-fMRI. To validate that we indeed induced the target emotions, participants gave ratings of the emotions they had experienced during the fMRI scanning. The distinctness of brain activity patterns underlying different emotions was investigated with MVPA by training a classifier algorithm to recognize the emotion the participant was experiencing. We specifically focused on three aspects of distinctness: 1) whether emotions have distinct brain activity patterns within the participant (within-participant classification with leave-one-run-out cross-validation; Studies I and II), 2) whether the distinct brain activity patterns generalize across individuals (across-participants classification with leave-one-participant-out cross-validation; Studies I and II), and 3) whether the distinct brain activity patterns generalize across emotion induction conditions (crossmodal classification; Study I). All classifiers were trained with feature-selected whole-brain voxel activity patterns.

Behavioral ratings confirmed that the stimuli elicited robust, specific emotions in participants. In Exp. 1 of Study I, subjects selected our *a priori* defined target emotion category to best correspond with their experienced emotion elicited by movie clips with 93.1% accuracy (Figure 7a). In Exp. 2 of Study I, similarity ratings confirmed that the emotion variants imagined by participants belonged to separate emotion clusters representing the basic emotions (Figure 7b). In Study II, participants reported the intensity of different emotions during each narrative stimulus, and the highest intensities were observed in our target emotion categories (Figure 7c). Also, narratives belonging to the same emotion category elicited similar emotional experiences (Figure 7d).

Within-participant classification in all experiments showed that both basic (anger, disgust, fear, happiness, sadness, and surprise) and non-basic (pride, gratitude, love, contempt, guilt, and despair) could be classified from voxel-activity patterns (Figures 8a-b and 9). A comparison between basic and non-basic emotions revealed that, on average, basic emotions were assigned to the correct category with higher accuracy. Across-participants classification was

also successful for basic emotions, suggesting that the neural underpinnings of these emotions generalize across participants (Figure 8d-e). Moreover, crossmodal classification using two different emotion induction techniques including external (movies) and internal (mental imagery) conditions was successful for basic emotions, implying that their neural signatures generalize across stimulus modalities (Figure 8c).

Taken together, the findings show that both basic and non-basic emotions have distinct brain activity patterns within the same participant. Moreover, for at least the basic emotions, the emotion-specific brain activity patterns generalize across participants and across emotion elicitation techniques.

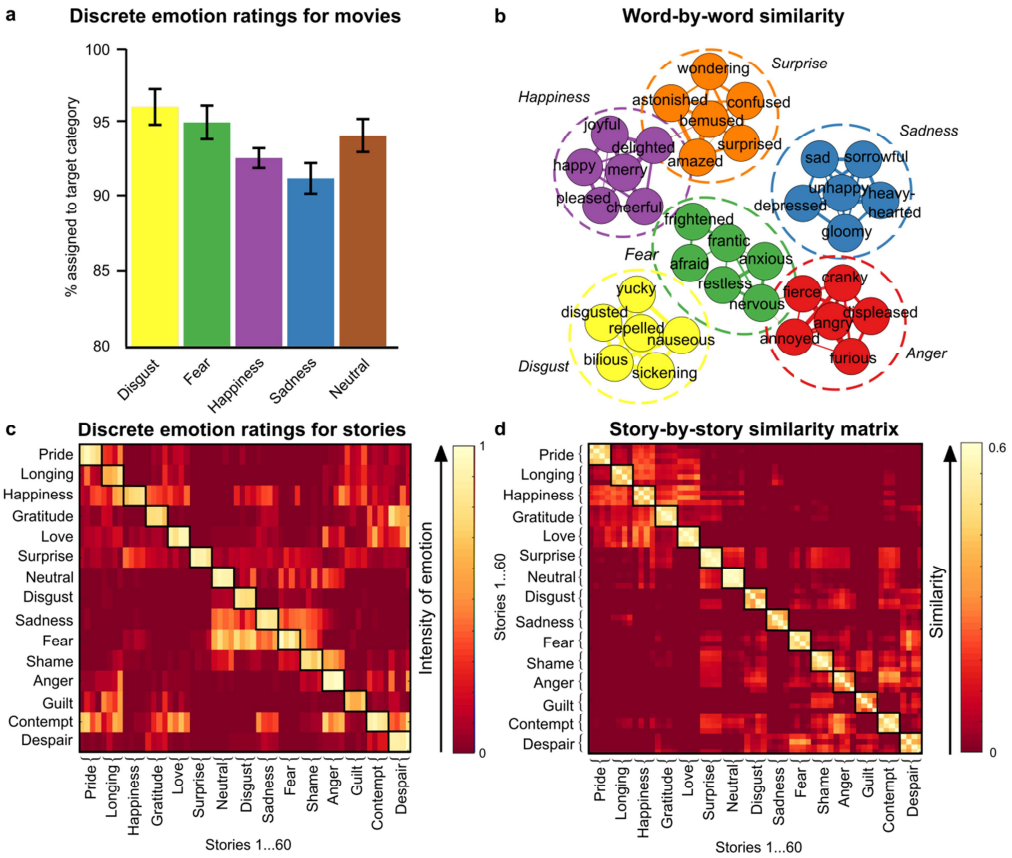


Figure 7. Behavioral results from Studies I and II show that stimuli elicited distinct and robust emotions. (a) Behavioral results in the Movie experiment (Exp. 1) of Study I: mean ± SEM percentages of movie clips per emotion category during which participants reported feeling the corresponding emotion. The clips were assigned to the predefined target category with 93.1% overall accuracy. (b) Behavioral results in the Imagery experiment (Exp. 2) of Study I: mean network of basic emotion concepts based on the participant's behavioral similarity ratings. Link width denotes similarity between words. (c) - (d) Behavioral results in Study II. Participants rated on a scale from 0 -1 how much of each emotion was elicited by each of the 60 narratives that targeted 14 emotional states and a neutral state (c). Based on the ratings, we calculated the similarity of emotion content between narratives by using Euclidean distances (d).

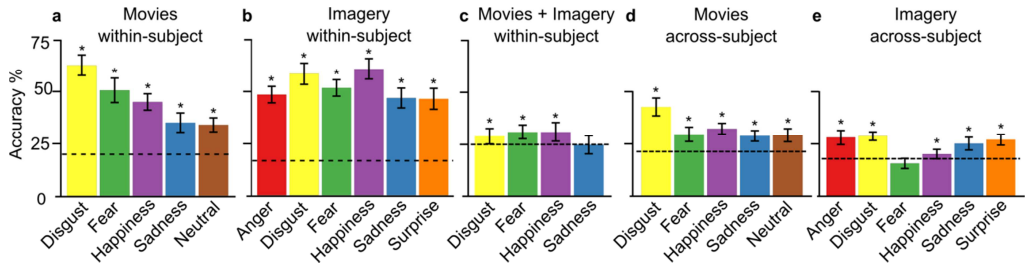


Figure 8. Classification results from Study I. Mean ± SEM classification accuracy for each emotion category. Dashed line represent the chance level (20% in the Movie experiment [a and d], 16.7% in the Imagery experiment [b and e], 25% in the cross-modal experiment [c]. Asterisks denote accuracies above a permuted chance level. (a) In the Movie experiment, the mean within-participant classifier accuracy was 47% for distinguishing one emotion against all others (averaged across all categories) and the classifier was able to classify each of the 5 emotion categories statistically significantly above chance level (20%, $p < 0.05$). (b) In the Imagery experiment, the mean within-participant classifier accuracy was 55% and the classifier was able to classify each emotion category statistically significantly above chance level (16.7%, $p < 0.05$). (c) The mean within-participant classifier accuracy in the crossmodal experiment was 29%. The classifier was able to classify all emotion categories except sadness statistically significantly above chance level (25%, $P < 0.05$). (d) The mean across-participant classifier accuracy in the Movie experiment was 34% and significantly above chance level (20%, $p<0.05$) for all emotion categories. (e) The mean across-participant classifier accuracy in Imagery experiment was 23% and significantly above chance level (16.7%, $p<0.05$) for all categories except for fear.

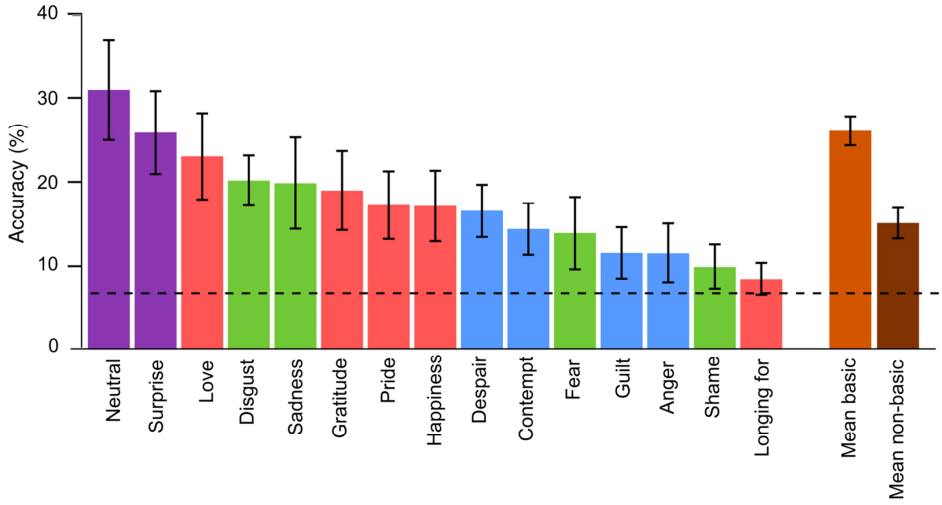


Figure 9. Classification results from Study II. Mean ± SEM classification accuracy for each emotion category using a whole-brain within-participant classifier. Dashed line represents chance level (6.7%). Colors reflect the clusters formed on the basis of experienced similarity of emotions (see RQ 4).

4.2 RQ 2: What are the core regions supporting emotions? (Studies I, II & III)

To unravel the brain mechanisms supporting different emotions, we examined the brain regions activating during basic and non-basic emotions. We specifically tested 1) how dimensions of valence and arousal and 2) and distinct emotions are coded by the brain. We hypothesized that the brain regions known from previous studies to be related to emotional processing would activate during all emotions which would support the view that there is no one-to-one mapping between any single brain region and a specific emotion.

BOLD-fMRI data from Studies I, II and III was examined to reveal the neural basis of different emotions. We specifically aimed 1) to examine whether emotions are better characterized by whole-brain activity patterns rather than activity patterns of single brain regions (comparison of whole-brain vs. *a priori* selected ROI pattern classification accuracies; Studies I and II), 2) to illustrate the core brain regions underlying specific emotions (voxels important for classification of basic emotions, Study I; cumulative activation maps across different basic and non-basic emotions, Study II; maps of superordinate emotion clusters, Study II), and 3) to investigate which of these core brain regions are modulated by lower-order dimensions of valence and arousal (univariate GLM with dynamic valence and arousal ratings; Study III).

The comparison of whole-brain and ROI pattern classification of basic emotions showed that whole-brain classification accuracy exceeded that of any single region of interest (Figure 10). ROIs with highest classification accuracies were found in prefrontal cortex, precuneus, pre- and postcentral gyri, and IFG, suggesting that emotion-specific brain activity patterns exist especially within these areas. However, the net brain activity pattern across all regions together gave a more accurate signature for each emotion. Illustrations of core brain regions constituting these emotion-specific net activity patterns show that voxels important for the classification of basic emotion were found especially in cortical midline, subcortical areas, and somatomotor regions for all emotions with no one-to-one mapping between a single brain region and a specific emotion (Figure 11). Furthermore, the activity of these same regions together with activity in frontal areas, brainstem and sensory (visual) areas was modulated by a majority of tested basic and non-basic emotions (Figure 12). To summarize emotion-specific activity in the brain, we defined superordinate categories of emotions by applying hierarchical clustering to the confusion matrices obtained from MVPA. The resulting emotion categories differ slightly in how they are distributed in the brain: especially, positive emotions activate the anterior prefrontal cortex, negative basic emotions tend to activate especially the somatomotor regions and negative social emotions show specific activation in left insula (Figure 13). When investigating the modulations of valence and arousal alone, we found that negative valence modulated the subcortical structures, cerebellum, and PCC, and positive valence modulated OFC and operculum (Figure 14a). High arousal modulated brain activity especially in midline structures, subcortical areas, and posterior STS, whereas the effects of low arousal were small (Figure 14b).

In summary, both basic and non-basic emotions have distinguishable neural bases characterized by specific, distributed activation patterns in widespread cortical and subcortical circuits, including especially the cortical midline, somatomotor and visceral regions, and subcortical areas, but extending also to frontal and sensory regions. Locally differentiated engagement of these globally shared circuits defines the unique neural signature activity pattern. Brain activity in these areas is modulated by valence and arousal, but for specific emotions, activation are also seen in areas exceeding these, suggesting that dimensions of valence

and arousal are not enough to completely characterize the neural differences between emotions.

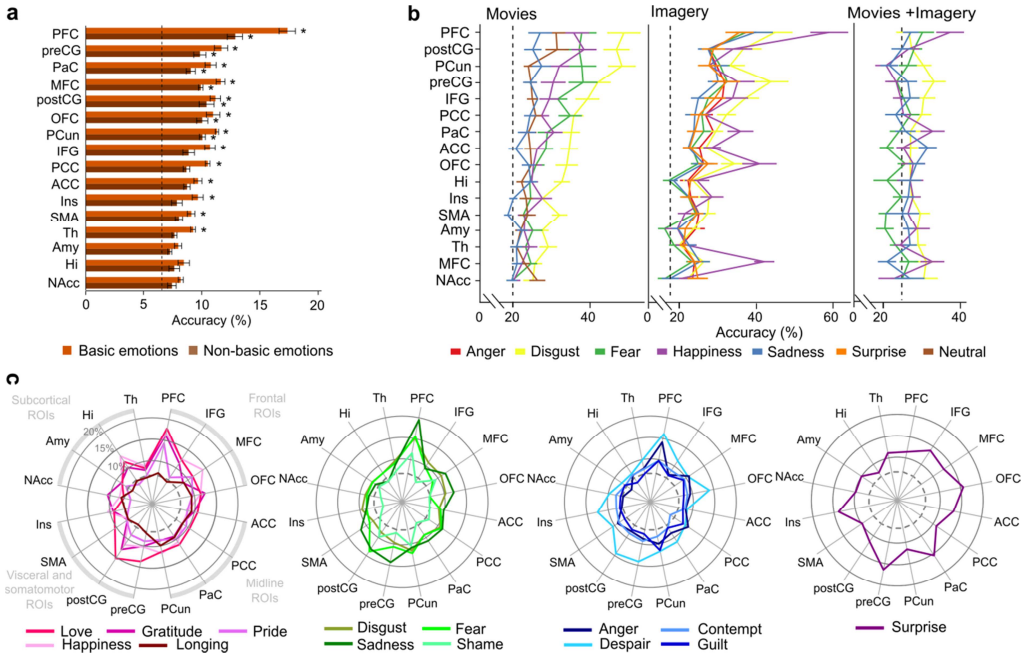


Figure 10. Region-of-interest-wise mean classification accuracies in Studies I and II. (a) Comparison of mean ± SEM classification accuracies between basic and non-basic emotions for each region-of-interest (ROI) in Study II. Asterisks denote above-chance-level accuracies ($p<0.05$). Dashed line represents chance level (6.7%). (b) Mean ± SEM classification accuracies for each ROI and each emotion separately in Study I. Dashed line represents chance level (20% in Movie experiment, 16.7% in Imagery experiment, 25% in crossmodal classification in crossmodal experiment). (c) Mean classification accuracies for each ROI and each emotion separately in Study II. Dashed line represents chance level (6.7%). PFC = prefrontal cortex; MFC = medial frontal cortex; OFC = orbitofrontal cortex; ACC = anterior cingulate cortex; PCC = posterior cingulate cortex; PaC = paracingulate cortex; PCun = precuneus; preCG = precentral gyrus; postCG = postcentral gyrus; SMA = supplementary motor area; Ins = insula; NAcc = nucleus accumbens; Amy = amygdala; Hi = hippocampus; Th = thalamus.

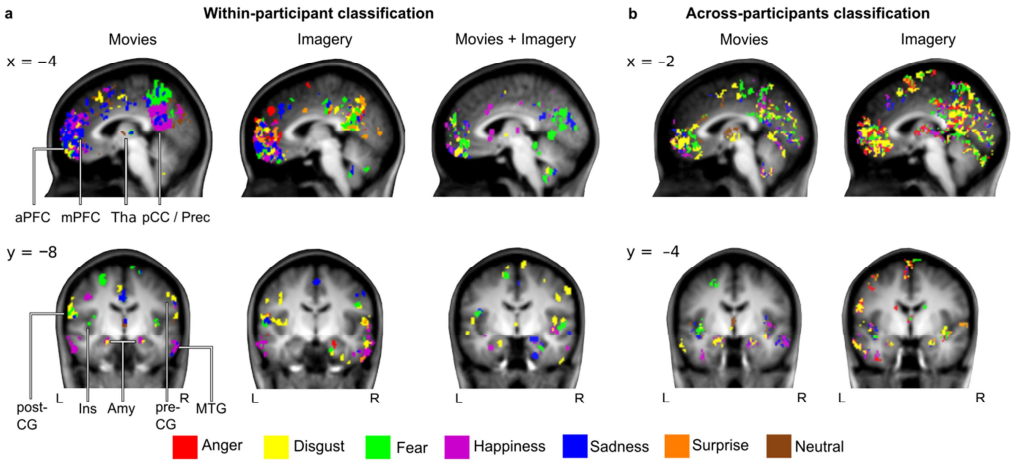


Figure 11. Voxels with the largest importance for the classification of basic emotions in Study I. The thresholding of the importance values (defined as voxel activation * classifier weight) is arbitrary and the maps are shown for visualization only. mPFC = medial prefrontal cortex; PCC = posterior cingulate cortex; Prec = precuneus; aPFC = anterior prefrontal cortex; LOC = lateral occipital cortex; postCG = postcentral gyrus; preCG = precentral gyrus; Ins = insula; Amy = amygdala; MTG = middle temporal gyrus.

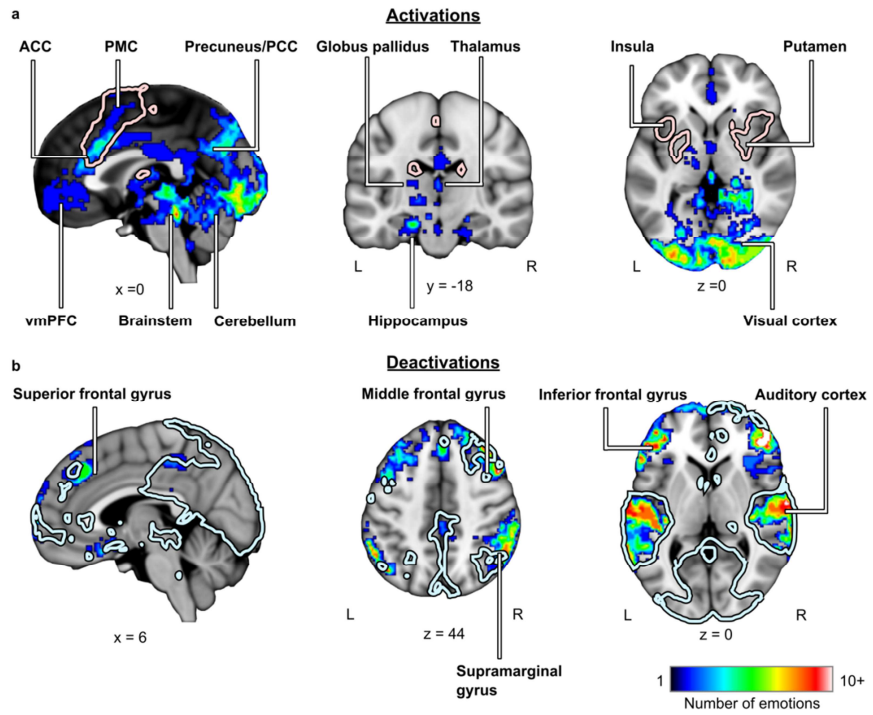


Figure 12. Cumulative activation/deactivation maps from Study II showing areas that activate/deactivate for multiple emotions. (a) Cumulative activation map shows the cumulative sum of binarized *t* maps ($p<0.05$, cluster-corrected) across each emotion vs. neutral condition. Outline shows the GLM results for all emotions contrasted against the neutral condition, $p<0.05$, cluster-corrected). (b) Cumulative deactivation map shows the cumulative sum of binarized *t* maps ($p<0.05$, cluster-corrected) across neutral condition vs. each emotion. Outline shows the GLM results for the neutral condition contrasted against all emotions, $p<0.05$, cluster-corrected).

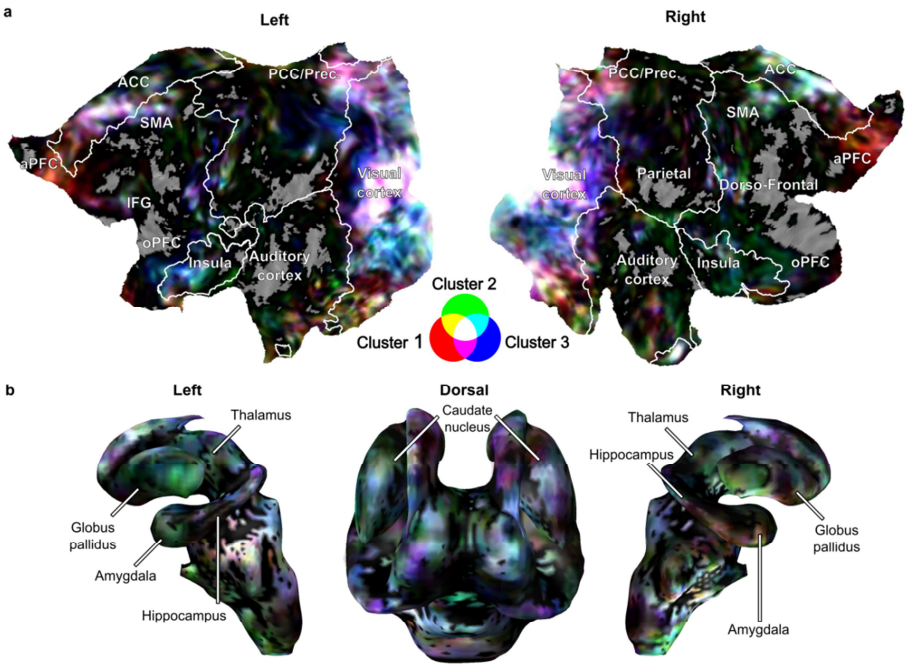


Figure 13. Cluster-specific activation patterns in cortical and subcortical regions from Study II. The maps show the averaged uncorrected *t* maps for emotions belonging to each cluster obtained from the hierarchical clustering analysis in cortical regions (a) and subcortical regions (b). Colors represent the three clusters: cluster 1 (red) = positive emotions (happiness, pride, love, longing for, gratitude); cluster 2 (green) = negative basic emotions (disgust, sadness, fear, shame); cluster 3 (blue) = negative non-basic emotions (anger, contempt, guilt, despair).

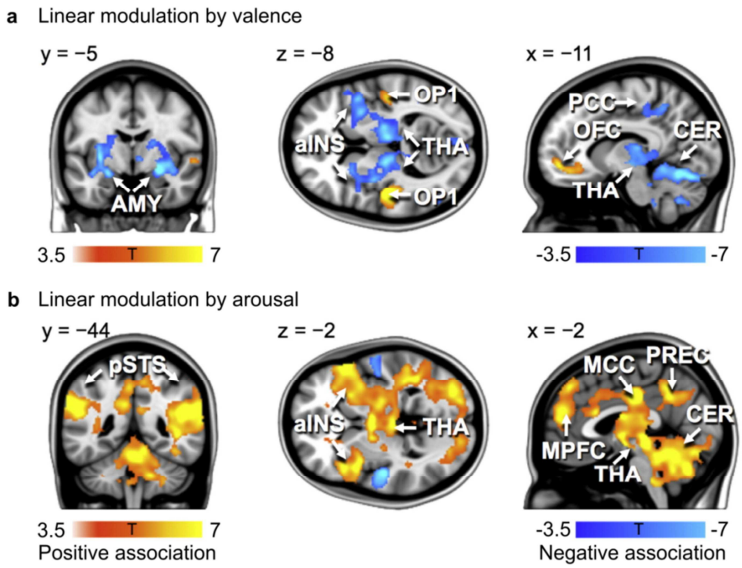


Figure 14. GLM results showing areas whose activation is modulated by valence and arousal from Study III. (a) Regions that are modulated by emotional valence of the stimulus. (c) Regions that are modulated by emotional arousal of the stimulus.

4.3 RQ 3: How does the large-scale functional connectivity vary during emotions? (Studies III & IV)

Previous results showed that specific emotions have distinguishable activity patterns distributed across the brain. As the brain functions as a network, also the connections between different regions might vary between different emotions. We hypothesized that if different emotions are supported by functional changes in large-scale neural networks, we should be able to decode them from functional connectivity data.

To elicit various emotional states, we varied either valence and arousal (positive, negative and neutral valence, high, low and neutral arousal; Study III) or discrete emotion content (basic emotions: anger, fear, disgust, happiness, sadness, and surprise; Study IV) of auditory stimuli that were presented to participants while their brain activity was measured with BOLD-fMRI. We then specifically examined 1) how valence and arousal modulate functional connectivity in the brain (combining dynamic ratings of valence and arousal with functional connectivity calculated with SBPS; Study III), 2) whether the whole-brain connectivity patterns differ between discrete emotional states (across-participants pattern classification of whole-brain functional connectivity patterns; Study IV), and 3) whether specific subnetworks contribute to the classification of functional connectivity patterns underlying different emotions (across-participants pattern classification of subnetwork functional connectivity patterns; Study IV).

In general, high arousal and negative valence increased functional connectivity across the whole brain, whereas the connectivity modulations due to positive valence and low arousal were more limited (Figure 15). Further, basic emotions could be classified from whole-brain functional connectivity patterns (Figure 16). Subnetwork classification revealed that the differences between emotions lie especially within the default mode network (DMN) connections (Figure 17), which was the only subnetwork with above-chance-level classification accuracy and where all emotions could be classified above chance level. We further investigated the subnetworks within the DMN and found that emotions could be classified from each other based on connectivity within the posterior midline DMN, between left temporal and frontal midline DMN, and between right temporal and posterior midline DMN (Figure 18).

In summary, while valence and arousal modulate functional connectivity markedly, the connectivity patterns underlying anger, fear, disgust, happiness, sadness, and surprise are more similar outside the DMN, where they show distinct functional connectivity patterns.

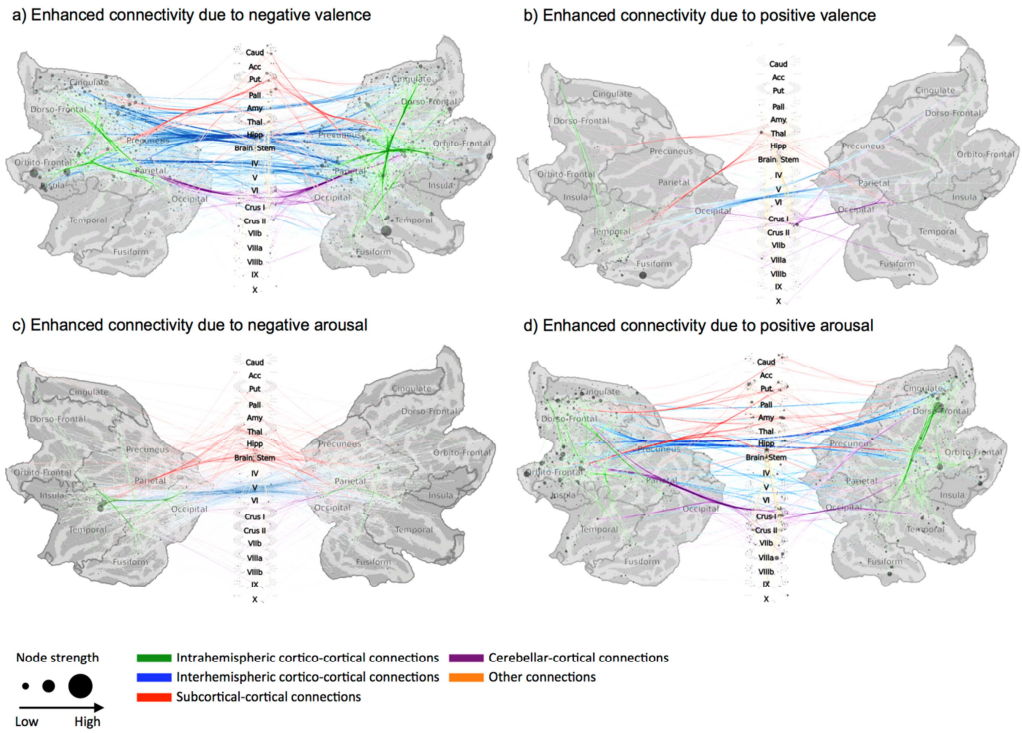


Figure 15. Functional connectivity modulations by valence and arousal from Study III. Connectivity graphs show the brain regions whose functional interconnectivity increased as a function of emotional valence (top row, a-b) and arousal (bottom row, c-d). Negative valence and arousal are shown in the left panel, positive valence and arousal in the right. Colour denotes link type and node circle size indicates node strength (all data thresholded at q < 0.1, FDR-corrected).

Figure 16. Classification of basic emotions based on whole-brain functional connectivity patterns (Study IV). (a) Emotion-wise classification accuracies for the full-network classification. Dashed line represents naïve chance level (16.6%). Asterisks denote significance relative to chance level (*p<0.01, ***p<0.0001). Thick line represents median of classification accuracies. Boxes show the 25th to 75th percentiles of classification accuracies and values outside this range are plotted as dots. Whiskers extend from box to the largest value no further than 1.5 * inter-quartile range from the edge of the box. (b) Classifier confusions from full network classification. Off-white colour code denotes classifier accuracy; cells shown in white have accuracies below naïve chance level.

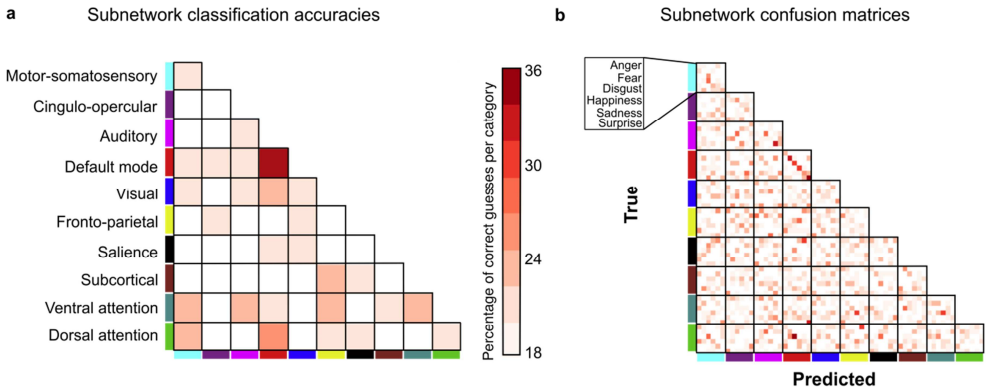


Figure 17. Subnetwork classification accuracies and confusion matrices from Study IV. (a) Classification accuracies for connectivity within and between each subnetwork. Colour code denotes classifier accuracy; cells shown in white have accuracies below naïve chance level. All non-white accuracies are above the naïve chance level (16.6%), but after correcting for multiple comparisons, only the accuracy for within default mode network connections remained significant. (b) Classifier confusions for subnetwork classification.

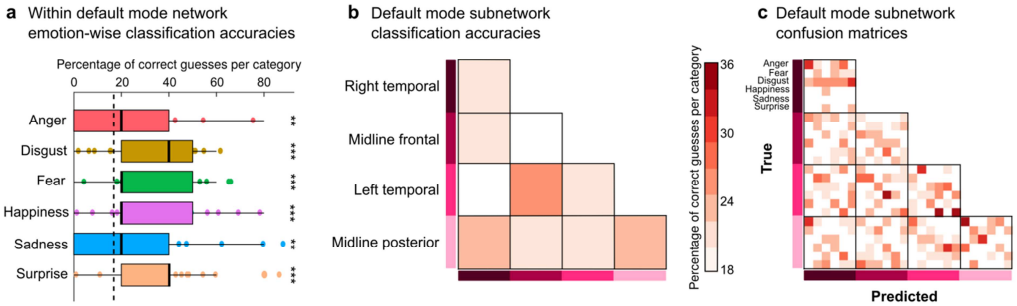


Figure 18. Classification results from default mode subnetwork (DMN) classification (Study IV). (a) Emotion-wise classification accuracies of the classification for connections within the DMN. Dashed line represents naïve chance level (16.6%). Asterisks denote significance relative to chance level (**p<0.001, ***p<0.0001). Thick line represents median of classification accuracies. Boxes show the 25th to 75th percentiles of classification accuracies and values outside this range are plotted as dots. Whiskers extend from box to the largest value no further than 1.5 * inter-quartile range from the edge of the box. (b) - (c) Classification accuracies (b) and subnetwork confusion matrices (c) for DMN subnetwork classification. Colour code denotes classifier accuracy; cells shown in white have accuracies below naïve chance level.

4.4 RQ 4: Do emotions that have similar neural bases also feel subjectively similar? (Studies I, II and V)

Many emotions feel distinct from each other in the mind and the body, and our results so far suggest that different emotions also have distinguishable neural underpinnings characterized by differences in distributed voxel activity patterns throughout the brain and in functional connectivity patterns especially within the default mode network. In this final part of the project we investigated the link between subjective feelings and their neural bases, a domain in affective neuroscience that remains unknown. We hypothesized that if the subjective experience is a sum of emotion circuits, emotions that share more similar neural underpinnings should be experienced in a more similar way.

To compare the similarity structures of emotions at the level of both brain activity and subjective feelings, we combined the BOLD-fMRI data and behavioral ratings of experienced basic and non-basic emotions (Studies I and II). Further, we took advantage of existing datasets that measured similarities in brain activity, subjective feelings, facial expressions, cognitive evaluations, and bodily sensations between the canonical basic emotions in a small meta-analysis (Study V). We specifically investigated 1) whether the similarity structure of brain activity and subjective feelings is correlated (RSA; Studies I and II), 2) whether the superordinate clusters of emotions are similar across brain activity and subjective feelings (hierarchical clustering analysis; Studies I and II), and 3) whether the similarity of subjective feelings resembles the similarity in brain activity, bodily sensations, cognitive evaluations, and facial expressions (Study V).

The correlation between experiential and neural similarity matrices was significant ($rho=0.43$, $p<0.001$ in Study I, $rho=0.68$, $p<0.0001$ in Study II), supporting our hypothesis that emotions that feel similar also have more similar neural bases. While variants of basic emotions clustered around the basic emotions in both neural and experiential data (Figure 19), different basic and non-basic emotions were clustered together to form a set of superordinate clusters - positive emotions, negative basic emotions, negative social emotions, and neutral emotions - that differed slightly of those in the experiential data (Figure 20). Clear clustering of basic emotions was present at multiple levels, and the similarity of subjective feelings correlated especially with similarities in neural basis, cognitive evaluations, and facial expressions, while correlation with bodily sensations was weaker (Figure 21).

We conclude that the more similar two emotions feel, the more similar are their neural underpinnings. Also, emotions tend to cluster to superordinate categories, which are distinguishable similar in both brain activity and subjective feelings for basic emotions, but differ slightly when examining a larger set of emotions. Especially, the discreteness of basic emotions is present at multiple levels, including those of underlying brain activity, bodily sensations, cognitive evaluations, facial expressions, and subjective feelings.

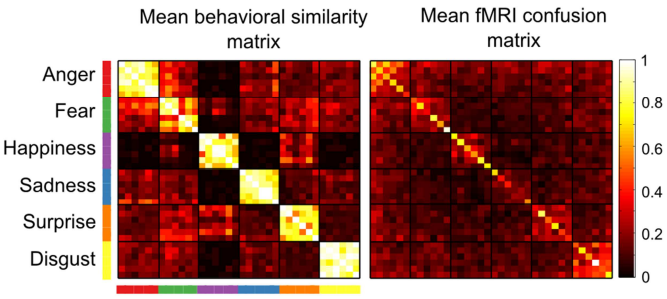


Figure 19. Comparison of neural and experiential similarities in Exp. 1 of Study I. Left: Experiential (behavioral) similarity matrix based on the rated similarity of emotional experiences evoked by each pair of emotions. Right: Neural similarity matrix based on the fMRI confusion matrix from word-by-word within-participant classification. Correct categories on the x-axis, classifier guesses on the y-axis.

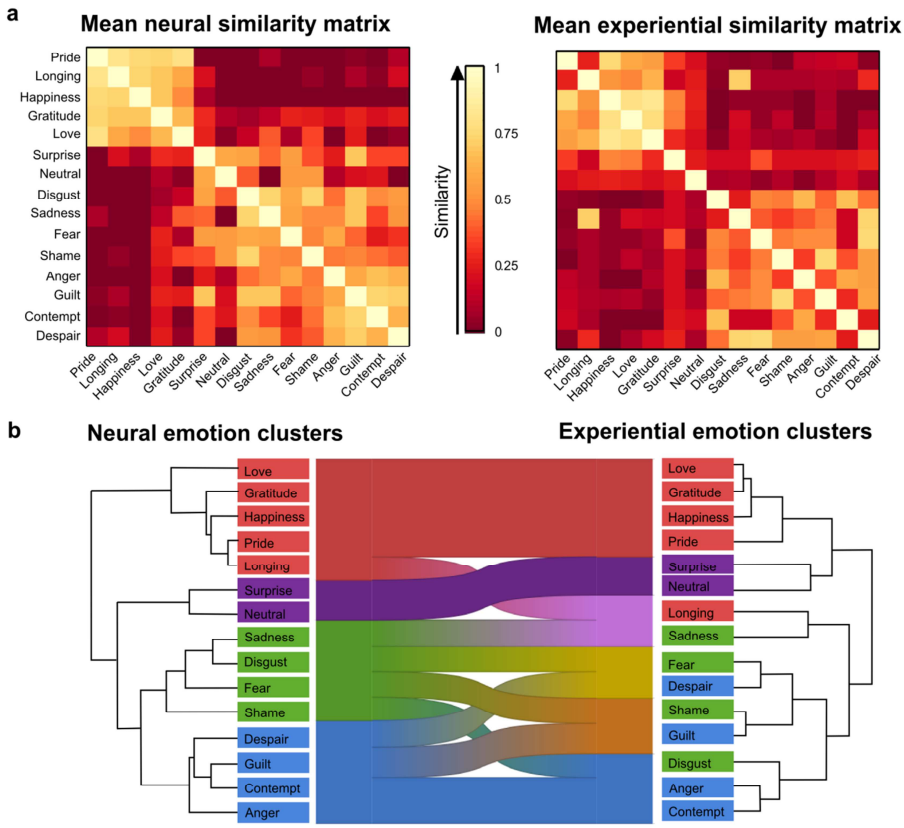


Figure 20. Comparison of neural and experiential similarities in Study II. (a) Left: Mean neural similarity matrix extracted from the classifier confusion matrix. The similarity matrix was created by calculating the Euclidean distance between each pair of emotions based on their category confusion vectors. Right: Mean experiential similarity matrix shows the rated similarity of emotional experiences evoked by each pair of emotion. (b) Alluvial diagram showing the similarity of hierarchical cluster structure of the neural and experiential similarities. Coloring of the emotion categories is based on the cluster of the neural similarity matrix.

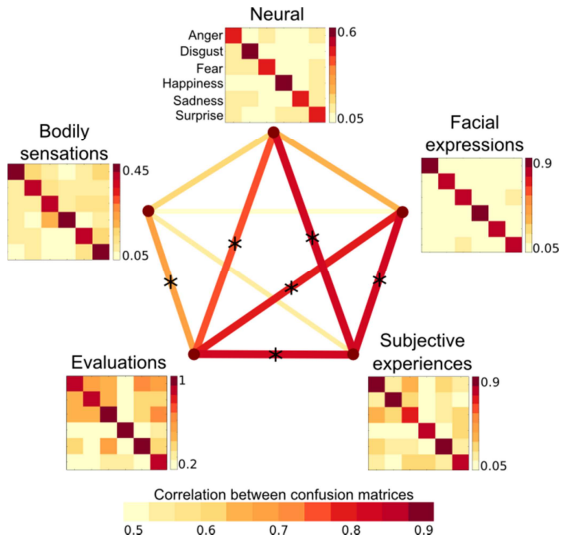


Figure 21. Cross-modal similarities between bodily, cognitive-evaluative, subjective, expressive, and neural representation of emotions. Matrices show the correlation matrices for data from different modalities using data from Calvo and Lundqvist (2008), Nummenmaa et al. (2014a), and Saarimäki et al. (2016). In all matrices, larger values denote higher similarity. Lines connecting different confusion matrices show the correlations between them calculated with Spearman correlation coefficient denoted. Line colour and width denotes the strength of the correlation, all shown similarities are significant, $p<0.0001$, in a parametric test. Asterisks (*) denote significant similarities in a complementary permution-based test where values were obtained by permuting the row and corresponding column elements ($p<0.05$, BH-FDR-corrected).

5. General discussion

5.1 Emotions as discrete patterns of systemic activity

Taken together, the current findings are that

1) both basic and non-basic emotions are characterized by distinct brain activity patterns distributed across the brain,
2) emotion-specific patterns are found especially in default mode network areas, somatomotor and visceral areas, subcortical areas, frontal cortex, and sensory areas,
3) default mode network connectivity is modulated differently by different emotions, and
4) similarities in subjective experience of emotions are linked to similarities in their neural underpinnings.

Based on these findings, emotions are best understood as widespread, system-level patterned activity, rather than selective regions or systems engageing during specific emotions. A data-driven meta-analysis of functional imaging studies (Kober et al., 2008) proposed a functional subdivision of emotional brain circuits into six groups, each responsible for processing different types of information (see also Meaux and Vuilleumier, 2015). These functional circuits were suggested to code for different components of emotions, such as attentional, motor, or mnemonic processes engaged during emotional episodes. However, previous experimental work has failed to establish different neural signatures for different emotions within these circuits, while the current work shows that these functional circuits contain emotion-specific brain activity patterns.

A speculative framework for how the systemic patterns might code for distinct emotions could be formulated by describing the components and the likelihood of a specific activity pattern within a component being associated with a particular emotion. When relevant emotional input - either external such as visual stimulus or internal such as a memory of a past event - arrives in the brain, it triggers changes in various parts of the nervous system that serve for different functions. For instance, processing of the relevance of the input signal causes changes in salience-processing areas (e.g., threatening stimuli cause more activation), sensory processing (e.g., attention targets visual search for relevant objects in the environment), modulation of sympathetic and parasympathetic nervous systems and endocrine systems (e.g., increasing heart rate, releasing adrenaline for quicker responses), and activation of related memory traces (e.g., what happened last time I was in this situation) and motor action tendencies (e.g., flight or fight, clenching the fists). In textbook terms of cognitive science where different components - in the case of emotions, including at least modules related

to cognitive, motivational, somatic, and behavioral changes - and their possible operations and responses are described as boxes, the emotional state will frame or narrow down the possible outcomes a specific module can have. For instance, during the state of fear, likelihood of freezing, fighting or fleeing would increase, whereas likelihood of stretching your legs to a relaxed position on a couch would decrease. This way, the emotional state modulates or restricts the outputs of different components. For pattern classification in neuroscience, this would lead to specific brain activity patterns within a brain region responsible for some component, which at times could be rather similar between some emotions: for instance, psychophysiological activation pattern related to both fear and disgust would probably be quite similar since they both are characterized by a state of readiness and adrenaline-related increase of heart rate. However, the patterns underlying these emotions would differ, for instance, in their memory or motor action components, leading to a different net activity pattern of the whole nervous system.

5.2 Distinct neural bases of different emotions

Different emotions - in the current studies, anger, fear, disgust, happiness, sadness, surprise, gratitude, love, guilt, contempt, despair, and pride - can be classified from brain activity patterns suggesting that they have distinct neural bases. Our results accord with recent studies that have achieved successful classification of discrete emotional states, usually focusing on the basic emotions only or a subset of these (Peelen et al., 2010; Said et al., 2010; Ethofer et al., 2009; Kotz et al., 2013; for reviews, see Kragel and LaBar, 2014, 2016). While the canonical basic emotions have attracted most attention in psychological and neurophysiological studies, they constitute only a small portion of the emotions humans universally experience (Edelstein and Shaver, 2007). Furthermore, accumulating behavioral evidence suggests that also other emotions are characterized by distinctive features in facial expressions (Baron-Cohen et al., 2001; Shaw et al., 2005), bodily changes (Nummenmaa et al., 2014a), and physiological activation patterns (Kreibig, 2010; Kragel and LaBar, 2013). Our data corroborate these findings by showing that also emotions not considered as 'basic' also have distinct brain activation patterns.

The within-participant classification accuracy in the current studies was always higher than the intersubject classification accuracy. This is a general tendency in pattern classification studies and is at least partly explained by that the brains of two individuals are never exactly the same. Local brain activity patterns underlying different emotions are most probably to some extent variable across individuals and reflect experience-dependent plasticity and genetically determined individual differences. Further, BOLD signal is noisy and fluctuates already within the same individual. Importantly, even if emotions would activate comparable brain regions across subjects, it is unlikely that these regions are anatomically aligned across individual brains. However, comparing the intersubject classification accuracies in different sets of emotions shows that while the canonical basic emotions - anger, fear, disgust, happiness, sadness, and surprise - could be classified also across participants, the more social emotions - in our studies love, gratitude, pride, longing, shame, guilt, despair, and contempt - could not. This result might be due to the lack of power, that is, restricted number of repetitions per stimuli in the current studies, or a real phenomenon that supports the view that basic emotions are universal. Another possibility is that as more social emotions require

more contextual knowledge of the situations, they also evoked more differing brain activity in different individuals.

If we consider emotion systems as wide-spread neural ensembles distinct to each emotion state, successful pattern classification of brain states across emotions would provide support for separate emotion systems. Each emotion likely modulates multiple functional systems of the brain differently, and their spatially distributed configuration might define the specific emotion. For instance, two emotions might share their somatosensory representations, but underlying interoceptive representations could be different. Thus, the general configuration of the nervous system leads to a specific emotion. However, it must be noted that this type of analysis does not readily reveal the actual neural organization or code of each emotion system. The pattern classification only tells us that, on average and at the level measurable with BOLD-fMRI, that activation patterns across emotions are statistically separable, whereas localizing the actual source of differences is more difficult. For this reason, we complemented the pattern classification analysis with visualization of different emotion categories using GLM and clustering. If basic emotions were somehow 'special' or different from non-basic emotions at the neural level, we should observe (i) discrete neural activation patterns for basic emotions but not for non-basic emotions, or (ii) different (or perhaps additional) neural systems underlying basic and non-basic emotions. Our classification results and cumulative maps show that both basic and non-basic emotions could be classified accurately mostly to a similar degree, although basic emotions on average reached higher accuracies, and they elicited activation in largely overlapping brain areas.

5.3 Core regions modulated by emotional states

Our findings show that large-scale cortical and subcortical networks support different emotions in a distinctive, category-specific manner. First, voxels important for the classification of anger, fear, disgust, happiness, sadness, and surprise were spread out across the brain and no single region-of-interest alone reached the accuracy of the whole-brain classification for any emotion, suggesting that the anatomically distributed activation patterns contain the most accurate neural signature of an individual's emotional state. This is in line with previous neuroimaging studies which typically show that joint activity from multiple regions best discriminates between different emotions (Baucom et al., 2012; Kotz et al., 2013; Kassam et al., 2013). Second, cumulative activation maps for anger, fear, disgust, happiness, sadness, surprise, shame, guilt, pride, love, contempt, longing, gratitude, and despair show again global activity changes. In general, we see emotion-related activation across the whole brain, and depending on the emotion, the exact configuration of brain activity varies in these areas. A summary of the areas usually found to be active during emotional tasks is presented in Figure 22; note that all these areas have been shown to activate also in tasks that do not have any emotional component. This highligths the view that emotions modulate wide-spread areas and no region might be specialized in emotion-specific processing only.

The medial prefrontal and medial posterior regions (mPFC, ACC, PCC, and precuneus) contributed most significantly to classification between different basic emotions and were activated during most basic and non-basic emotions. Thus, local activation patterns within these areas are most discriminative across emotions and reflect the most distinct neural signatures for different emotions. Also, the activity within mPFC and precuneus was modulated by arousal and, to a smaller degree, also by negative valence. These midline re-

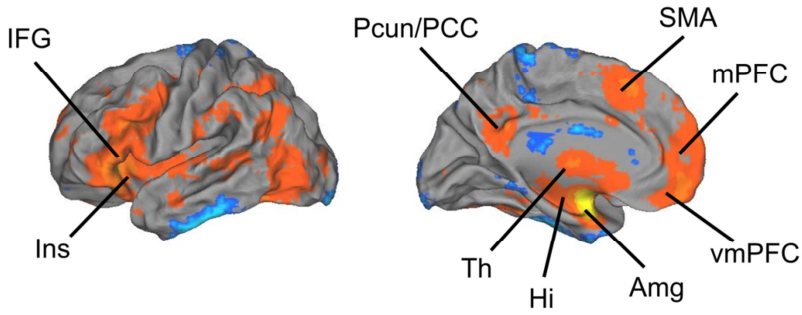


Figure 22. Summary of emotion-related brain areas. Regions are shown on their approximate locations on the PALS12 cortical atlas template (Van Essen, 2005). They were derived from Neurosynth (http://www.neurosynth.org) and show the results of an automated meta-analysis of 1340 studies for the term emotional. All shown activations/deactivations are significant with P<0.01 (FDR corrected). The depicted regions are bilateral but shown on left hemisphere for simplicity. Some activations reside within the surface and subcortical regions are not shown on the template but are shown in their approximate locations. Abbreviations: Amg - amygdala, Hi - hippocampus, IFG - interior frontal gyrus, Ins - insula, mPFC - medial prefrontal cortex, PCC - posterior cingulate cortex, Pcun - precuneus, SMA - supplementary motor area, Th - thalamus.

gions are consistently activated during emotional processing in different sensory modalities (Phan et al., 2002; Kober et al., 2008; Lindquist et al., 2012), particularly coding for emotional valence (Chikazoe et al., 2014; Colibazzi et al., 2010), and contain emotion-specific patterns independent of the task or exact emotion categories used (Peelen et al., 2010; Chikazoe et al., 2014; Skerry and Saxe, 2014; see also Kragel and LaBar, 2014, for a review).

The mPFC and PCC receive inputs from insula which processes visceral information, from amygdala which codes the affective relevance of the stimulus, from medial temporal lobe areas involved in memory, and from thalamus and hypothalamus which govern arousal (Öngür and Price, 2000; Kober et al., 2008; Etkin et al., 2011). Together, mPFC, precuneus, and PCC form the medial part of the default mode network (DMN; see Figure 23), typically linked with self-referential processing and introspection (Amodio and Frith, 2006; Northoff and Bermpohl, 2004; Northoff et al., 2006; Buckner and Carroll, 2007). This anatomical architecture makes these midline regions a plausible candidate for integrating information about one's internal, mental and bodily states (Buckner and Carroll, 2007; Klasen et al., 2011; Mar, 2011) with representations from memory and personal relevance (Summerfield et al., 2009;

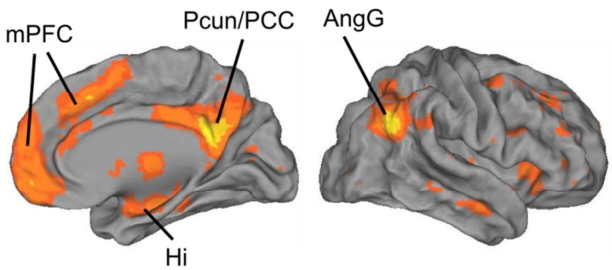


Figure 23. Default mode network. Default mode brain regions are shown on their approximate locations on the PALS12 cortical atlas template (Van Essen, 2005). Regions were derived from Neurosynth (http://www.neurosynth.org) and show the results of an automated meta-analysis of 516 studies for the term default mode. All shown activations/deactivations are significant with P<0.01 (FDR corrected). The depicted regions are bilateral but shown on right hemisphere for simplicity. Some activations reside within the surface and subcortical regions are not shown on the template but are shown in their approximate locations. Abbreviations: Hi - hippocampus, mPFC - medial prefrontal cortex, PCC - posterior cingulate cortex, Pcun - precuneus.

D'Argembeau et al., 2010). The patterns of activity resulting from the binding of these various representations might constitute a core feature of an emotional state regardless of the particular emotion category, and possibly underlie the distinctive signatures of these states as identified by our MVPA analyses. Further, the classification of basic emotions from functional connectivity patterns within DMN was successful.

Somato-motor and visceral regions including postcentral gyrus, posterior insula, and precentral gyrus were also among the most important brain regions for the classification of basic emotions, and premotor cortex, cerebellum (including vermis and the anterior lobe), globus pallidus, caudate nucleus, and posterior insula were activated during most basic and social emotions, but according to the cluster visualizations especially during the processing of emotions that have a strong impact on action tendencies and avoidance-oriented behaviors (fear, disgust, sadness, shame, surprise; Frijda et al., 1989). These findings accord with previous work showing that different emotions elicit discernible patterns of bodily sensations (Nummenmaa et al., 2014a), that primary somatosensory, motor and premotor cortices are reliably engaged during emotion perception (De Gelder et al., 2004; Nummenmaa et al., 2012; Pichon et al., 2008) and that cerebellum has a role in emotion regulation (Schutter and van Honk, 2009). Moreover, damage to somatosensory cortices (Adolphs et al., 2000) or their inactivation by transcranial magnetic stimulation (Pourtois et al., 2004) can cause significant deficits in the recognition of emotions. Similarly, posterior insula mediates the interoceptive awareness of one's own bodily functions (Critchley et al., 2004) and its damage may impair various components of emotion processing, including gustatory processing (Calder et al., 2001) and interoception (Naqvi et al., 2007). Precentral gyrus containing the primary motor cortex is also consistently activated during emotional experience and emotion perception (De Gelder et al., 2004; Hajcak et al., 2007), and it likely plays an important role in motor preparation processes related to emotion and action tendencies (Frijda, 1986; Mazzola et al., 2013; Kohler et al., 2002; Wicker et al., 2003).

Subcortical regions including amygdala and thalamus showed distinct activity patterns for different emotions. Both of these regions are related to salience processing and emotional arousal modulation (Adolphs, 2010; Anders et al., 2004; Damasio and Carvalho, 2013; Kragel and LaBar, 2014) and show discernible activation patterns across basic emotions (Wang et al., 2014), findings that we now extend also to non-basic emotions. However, our ROI analysis in Study I revealed poorer classification accuracy in subcortical versus cortical components of the emotion network. Furthermore, in none of the subcortical ROIs could we separate between all emotion categories. It is possible that this finding reflects the positive association between classification accuracy and ROI size, as the subcortical ROIs were, on average, smaller than their cortical counterparts. However, follow-up analysis established that mere ROI size unlikely accounts the poorer classification accuracy in the subcortical ROIs, particularly as some—such as thalamus—were indeed relatively large. One possibility is that the subcortical circuit contributes to shaping emotional states jointly with the cortical regions. The subcortical regions likely govern elementary functions related to arousal, saliency, and relevance processing, which could be shared across different emotions (Adolphs, 2010; Damasio and Carvalho, 2013; Kragel and LaBar, 2014). Thus, activity in these areas might not be specific enough to separate between all emotions but contributes to the overall brain activity related to the emotion - this is also probably the case with other emotion-related brain areas. Activity in these subcortical regions may then contribute to the generation of discrete emotional states via feed-forward connections to the frontal cortex but the

latter may also shape emotion responses through feedback interactions with subcortical regions. We also found consistent emotion-dependent activity in the brainstem, including periaqueductal grey (PAG), pons, and medulla, for almost all emotions included in our studies (see Damasio et al., 2000; Damasio, 2010). This activation might reflect the control of autonomic nervous system's reactions to different emotions (Critchley et al., 2005; Linnman et al., 2012) and/or covert activation of particular motor programs (Blakemore et al., 2016).

Subregions of the frontal cortex including aPFC and IFG were important for the classification of all emotions. Especially, anterior prefrontal cortex was activated especially during positive emotions (happiness, love, pride, gratitude, longing) according with previous research linking anterior prefrontal cortex with positive affect (Vytal and Hamann, 2010; Bartels and Zeki, 2004; Zahn et al., 2009). Finally, we also found emotion-specific activity in **sensory areas** (auditory and visual) where previous studies have also reported emotion-related effects (Nummenmaa et al., 2012; Holmes and Mathews, 2005; Kassam et al., 2013).

5.4 Large-scale functional connectivity differences between emotions

We found that whole-brain functional connectivity patterns differed significantly during different emotions (anger, fear, disgust, happiness, and surprise), as evidenced by significantly above chance-level classification accuracy of whole-brain functional connectivity patterns in Study IV. A closer examination of the source of differences showed that emotion-specific connectivity patterns were most prominently observed in within the default mode system. Above chance-level intersubject classification confirmed that these connectivity patterns were similar across subjects. The results thus show that not only regional activity patterns (Saarimäki et al., 2016; Kragel & LaBar 2016) but also large-scale connectivity changes across specific brain systems underlie different emotional states.

Prior studies have addressed emotion-triggered changes in functional brain connectivity during discrete emotional states with a limited set of *a priori* selected ROIs (Eryilmaz et al., 2011; Tettamanti et al., 2012; Touroutoglou et al., 2015; Raz et al., 2016). Our results are the first demonstration that emotion-specific connectivity patterns exist both at global and local scales in the brain. These emotion-specific connectivity changes are long-lasting (in the current studies, persisting at least one minute).

Temporally fine-grained functional connectivity analysis also revealed large-scale brain networks underlying the processing of emotional valence and arousal dimensions: both negative and positive valence and arousal were associated with enhanced functional connectivity in large-scale networks spanning from limbic circuits to the neocortex. Particularly the connections between limbic, sensory and association cortices were modulated by dynamic changes in valence and arousal. Emotion systems may thus manage information processing priorities not only in specific sub-systems or circuits, but also at global level. Differences in the effects of valence and arousal on brain connectivity were most prominent in their spatial layout. Second, the spatial layout of the valence and arousal dependent connectivity changes was different. Whereas positive valence resulted in increased frontotemporal, thalamic and striatal connectivity, negative valence resulted in widespread increase in connections from occipito-parietal, limbic (insula, cingulum) and fronto-opercular (primary and premotor cortices, lateral prefrontal cortex) regions. On the contrary, the connectivity changes from the brain's speech processing circuit (specifically Broca's region, auditory cortex and IPC) and limbic emotion circuits (thalamus, striatum, amygdala) as well as frontal cortex were promi-

nently associated with increasing arousal, with only a limited set of occipito-parietal connections being associated with decreasing arousal.

In sum, the current analyses showed that while dynamic changes in valence and arousal resulted in large-scale differences in functional connectivity spanning the whole brain, the differences between specific basic emotions - at least when calculated across the timescale of one minute - were present mostly within the default mode network. This difference can of course be explained by the difference in timescale of these analyses: while valence and arousal modulations were examined on a momentary, 1-second timescale, the emotion-specific connectivity patterns were calculated across a stimulus that lasted for 1-minute. The current analyses did not address the fast changes in dynamic connectivity between different discrete emotions; therefore, we cannot rule out the the fast dynamic connectivity varies during different discrete emotions. However, the current data would allow such analysis to be conducted in future if ratings of dynamic changes in emotion-specific intensity were collected and added to the analyses. Another explanation of the different results between the analyses focusing on valence/arousal and specific emotions might be the different methods chosen for calculating the functional connectivity. In valence/arousal analysis we used seed-based phase synchrony, and in emotion-specific analyses correlation of the node time series was applied. It would be important to extend the seed-based phase synchrony analyses to the emotion-specific data to provide further insight on the differences.

5.5 Similar in mind, similar in brain: how does the neural similarity relate to the subjectively felt similarity of emotions?

Humans are often aware of their current emotional state, which may help to fine-tune the behavior adaptively to better match to the challenges posed by the environment (Damasio et al., 1996). We found that the more similar neural signatures two emotions had, the more similar were the corresponding subjective feeling states. This was the case for both basic and non-basic emotions. Overall, at least for the basic emotions, the similarity structure of emotions is highly similar at least at the level of neural activity, subjective experiences, facial expressions, bodily sensations, and cognitive evaluations, as shown by our meta-analysis in Study V.

Damasio et al. (2000) have suggested that emotion-specific neural patterns across a range of brain regions could explain why each emotion feels subjectively different. Emotions might thus constitute discrete activity patterns in regions processing different emotion-related information, such as somatosensory (bodily sensations), motor (actions), as well as brainstem and thalamocortical loops (physiological arousal). Activation from these areas is then integrated in the cortical midline, such integration then giving rise to the interpretation of the subjective feeling (Northoff and Bermpohl, 2004; Northoff et al., 2006). Our results support this suggestion. We propose that the joint activation of different components is integrated in the mPFC and precuneus/PCC where distributed responses arising in the downstream brain regions are ultimately connected with the context and personal goals, presumably resulting in distinctive neural signatures that reflect the subjective experience of a specific emotion. Thus, a subjective feeling of a specific emotion stems from the net activation of different subprocesses, rather than solely on the basis of any single component of emotion. Also prior studies support the role of medial frontal cortex in the subjective feelings of emotions (Barrett et al., 2007; Etkin et al., 2011; Herbert et al., 2011; Satpute et al., 2012).

The correlations between whole-brain neural and experiential similarity matrices was at 0.5-0.6. Therefore, while the correlations are fairly high, the similarity of at least the whole-brain neural patterns did not fully explained felt similarities. To reveal the role of cortical midline regions in coding of subjective feeling, further analyses focusing on the similarity in activity patterns in this and other regions alone would be important.

5.6 Limitations

The current work uses in particular pattern classification to investigate the neural bases of emotions. While the successful classification reveals that there are some consistent differences between the special patterns underlying different emotions, interpreting the spatial distribution of activity patterns and their roles in coding of specific emotions is difficult. First, an important caveat in fMRI pattern classification studies is the visualization and interpretation of the spatial activity patterns. Whereas traditional univariate analyses directly reveal the voxels that activate above some threshold to a specific experimental condition, pattern classifier takes into account all voxel activations in the area of interest. Therefore, changes in the activity of any input voxel - even those activating very little or deactivating - can result to a different classifier output. To gain a full picture of spatial patterns coding for different conditions, one should show the weights and activity values of each voxel, which makes the visualizations difficult. We have used thresholding of importance values in Study I, however, this only preserves the most important voxels (i.e., those above some arbitrary threshold) while other voxels still contribute to the classification. Therefore, while waiting for more efficient visualization methods, the MVPA society has largely resorted back to GLM to visualize the spatial patterns underlying different conditions, as we have also done in Study II. Second, classification does not tell us anything about the role of different regions in coding of a specific emotional state nor provides any causal evidence for the role of different brain regions in the cascade of emotional processing. All interpretation is therefore speculation based on reverse inference: which functions previous studies have reported related to a specific region.

Pattern recognition is always compromised by the *a priori* class labels, and this is also true for the current study. Despite the careful stimulus selection procedures we employed, it is possible that the stimuli did not evoke exactly the targeted emotions in all participants. Moreover, the category labels were predefined by the research group, that is, we did not take into account subject-specific ratings, but rather we assumed that the stimuli induce same emotions in all participants. Despite these caveats, the classification accuracies were significant, showing that the emotion elicitation was successful. Also, our behavioral ratings showed that at least on average, participants chose the *a priori* defined target category for the intended emotion at high accuracy. However, future studies could take advantage of more subject-specific ratings.

Another related limitation is that the emotion categories we have chosen do not necessarily correspond to the ground truth of existing emotion categories. This is related to a more general problem in the field: we and others use emotion categories that resemble those in everyday life, yet, whether such categories exist in the neurobiological level can be questioned (see e.g. Adolphs, 2017). Take fear, for instance, an emotion that has been extensively studied in affective neuroscience. Our everyday concept of fear might include both social and physical fear, which both have very different roles and therefore also different underlying functions

also at the neural level. Grouping both under the term fear might cause confounds in the data. One possible solution to this ground truth problem would be to use generative models in the analysis. Further, there might be cultural differences in both emotion concepts, their interpretation and underlying functions (see e.g. Wierzbicka, 1999). Our Finnish-speaking sample might limit the generalizability of the results.

Despite the careful stimulus selection and the measures we took to validate emotional induction in participants, it is possible that the stimuli did not fully capture the target emotion only and potentially elicited also a mixture of emotions. Yet, these mixed emotions may arise in different time points during the stimulation and might not be as strong as the main target emotions, thus it is likely that trial-wise activations pertain to the target emotion, as evidenced by the clear primary emotions reported by our participants per each stimulus. Despite this caveat, the observed MVPA pattern might reflect whether each stimulus is dominated by one emotion or at least show the weighted influence of each emotion on a voxel activity. Thus, the successful classification *per se* shows that at least the target emotions were successfully elicited; yet, better classification accuracies may be reached if stimuli could be targeted more carefully to one category at the time.

5.7 Future directions

When studying the human emotions, it is important to consider how natural emotions we can elicit in the restricted laboratory or scanner setting. Do the emotions we elicit correspond to those in everyday life? Using static images has traditionally been common practice in cognitive and affective neuroscience. This is probably because simple stimuli are easy to control for compounds caused by, for instance, visual or auditory differences between them. However, one could argue that emotions are context-dependent: especially the non-basic emotions might require further knowledge about the context to be evoked. Another important question then is whether two occasions of the same emotion can ever be the same, because of contextual differences (see e.g. Barrett, 2006; 2017). In the current work, we aimed to use as naturalistic stimuli as possible - including mental imagery, movies, and narrative-guided imagery - to evoke a multitude of emotional states. However, I suggest we move to even more naturalistic stimuli in future. fMRI studies rarely used ecologically valid stimuli to induce emotions, therefore, future studies should aim to carefully design paradigms that evoke as natural emotions as possible. The few studies that have brought ecologically valid stimulation to brain imaging setup have used autobiographical recall of emotional events, fear of electric shock, or real tarantulas (Damasio et al., 2000; Mobbs et al., 2007, 2010). While eliciting emotions in these natural ways may limit the range of different emotions included in one study potentially to only one, maybe this is the way forward that allows a more in-depth investigation of a particular emotional state (Adolphs, 2017; Nummenmaa and Saarimäki, in press).

The primary current method for studying the brain basis of human emotions is fMRI. However, fMRI has no pharmacological specificity and the time resolution is relatively poor. While the time resolution could be improved by using other brain-imaging methods, such as magnetoencephalography (MEG), transcranial magnetic stimulation (TMS), or electroencephalography (EEG), these methods are limited in their spatial accuracy and scope, both features that are necessary when studying human emotions that spread across the brain. Pharmacological basis of emotions could, however, be studied using patients or positron

emission tomography (PET), which would and do provide important complementary information to the fMRI results.

Finally, the framework presented in Section 5.1 calls for further studies in how emotion-specific patterns code for the output possibilities of different functional components of the brain. A more fine-grained description and analysis of, for instance, mnemonic, visceral, motor, and sensory outputs related to distinct emotional states would be an interesting approach for future. Another important question is how these regions work in synchrony to contribute to the formation of an emotional state, and what are the causal relationships between the different components, that is, how does the activity spread between different brain regions and component processes. These questions require more detailed measurement of the temporal properties of brain activity which could be investigated – regardless of the slow nature of BOLD signal – with fMRI using faster acquisition sequences.

6. Conclusions

Understanding the neural basis of emotions is a timely topic in affective neuroscience, revolving especially around how different, discrete emotional states are coded by the neural system. The current work demonstrates and illustrates the specific and stereotypical changes that take place during different basic and non-basic emotional states. Emotions are best understood by modulations of net activation of multiple functional components which lead to both local and global, emotion-specific activity patterns.

The current findings make a substantial contribution to the search of neural basis of emotions. The set of five studies shows emotion-specific patterns both in large-scale brain activity and connectivity, in areas including the cortical midline, somatomotor and visceral areas, subcortical areas, frontal cortex, and sensory regions. Lower-order dimensions of valence and arousal modulate the dynamic functional brain connectivity differently, and especially default mode network connectivity shows different, sustained patterns for specific emotions. Finally, similarities in subjective experience of emotions are linked to similarities in their neural underpinnings.

References

Adolphs R (2002a) Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and cognitive neuroscience reviews* **1**:21-62.

Adolphs R (2002b) Neural systems for recognizing emotion. *Curr Opin Neurobiol* **12**:169-177.

Adolphs R (2010) What does the amygdala contribute to social cognition? *Ann NY Acad Sci* **1191**:42-61.

Adolphs R (2017) How should neuroscience study emotions? by distinguishing emotion states, concepts, and experiences. *Soc Cogn Affect Neur* **12**:24-31.

Adolphs R, Damasio H, Tranel D, Cooper G, Damasio AR (2000) A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J Neurosci* **20**:2683-2690.

Al-Shawaf L, Conroy-Beam D, Asao K, Buss DM (2016) Human emotions: An evolutionary psychological perspective. *Emot Rev* **8**:173-186.

Amodio DM, Frith CD (2006) Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* **7**:268-277.

Anders S, Lotze M, Erb M, Grodd W, Birbaumer N (2004) Brain activity underlying emotional valence and arousal: A response-related fMRI study. *Hum Brain Mapp* **23**:200-209.

Anderson DJ, Adolphs R (2014) A framework for studying emotions across species. *Cell* **157**:187-200.

Bandettini PA, Wong EC, Hinks RS, Tikofsky RS, Hyde JS (1992) Time course EPI of human brain function during task activation. *Magn Reson Med* **25**:390-397.

Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I (2001) The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psyc* **42**:241-251.

Barrett LF (2006) Are emotions natural kinds? *Perspect Psychol Sci* **1**:28-58.

Barrett LF (2017) The theory of constructed emotion: an active inference account of interoception and categorization. *Soc Cogn Affect Neur* **12**:1-23.

Barrett LF, Wager TD (2006) The structure of emotion: Evidence from neuroimaging studies. *Curr Dir Psychol Sci* **15**:79-83.

Barrett LF, Mesquita B, Ochsner KN, Gross JJ (2007) The experience of emotion. *Annu Rev Psychol.* **58**:373-403.

Bartels A, Zeki S (2004) The neural correlates of maternal and romantic love. *Neuroimage* **21**:1155-1166.

Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. *Proceedings of the Third International ICWSM Conference* **8**:361-362.

Baucom LB, Wedell DH, Wang J, Blitzer DN, Shinkareva SV (2012) Decoding the neural representation of affective states. *Neuroimage* **59**:718-727.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* **57**:289-300.

Bishop CM (2006) Pattern recognition and machine learning. New York: Springer.

Blakemore RL, Rieger SW, Vuilleumier P (2016) Negative emotions facilitate isometric force through activation of prefrontal cortex and periaqueductal gray. *Neuroimage* **124**:627-640.

Buckner RL, Carroll DC (2007) Self-projection and the brain. *Trends Cogn Sci*, **11**:49-57.

Calder AJ, Lawrence AD, Young AW (2001). Neuropsychology of fear and loathing. *Nat Rev Neurosci* **2**:352.

Calvo MG, Lundqvist D (2008) Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behav Res Methods* **40**:109-115.

Chikazoe J, Lee DH, Kriegeskorte N, Anderson AK (2014) Population coding of affect across stimuli, modalities and individuals. *Nat Neurosci* **17**:1114-1122.

Clark-Polner E, Johnson TD, Barrett LF (2016) Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cereb Cortex* **27**:1944-1948.

Cole MW, Reynolds JR, Power JD, Repovs G, Anticevic A, Braver TS (2013) Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci* **16**:1348-1355.

Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE (2014) Intrinsic and task-evoked network architectures of the human brain. *Neuron* **83**:238-251.

Colibazzi T, Posner J, Wang Z, Gorman D, Gerber A, Yu S, Zhu H, Kangarlu A,Duan Y, Russell JA, Peterson BS (2010) Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion* **10**:377.

Costa VD, Lang PJ, Sabatinelli D, Versace F, Bradley MM (2010) Emotional imagery: assessing pleasure and arousal in the brain's reward circuitry. *Hum Brain Mapp* **31**:1446-1457.

Critchley HD, Wiens S, Rotshtein P, Dolan RJ (2004) Neural systems supporting interoceptive awareness. *Nat Neurosci* **7**:189.

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI)"brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* **19**:261-270.

Critchley HD, Rotshtein P, Nagai Y, O'Doherty J, Mathias CJ, Dolan RJ (2005) Activity in the human brain predicting differential heart rate responses to emotional facial expressions. *Neuroimage* **24**:751-762.

Damasio AR (1995) Toward a Neurobiology of Emotion and Feeling: Operational Concepts and Hypotheses. *Neuroscientist* **1**:19-25.

Damasio AR (1999) The feeling of what happens: body and emotion in the making of consciousness. New York: Harcourt Brace.

Damasio AR (2010) Self comes to mind: constructing the conscious mind. New York: Pantheon.

Damasio AR, Everitt BJ, Bishop D (1996) The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos T Roy Soc B* **351**:1413-1420.

Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LL, Parvizi J, Hichwa RD (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat Neurosci* **3**:1049–1056.

Damasio AR, Carvalho GB (2013) The nature of feelings: evolutionary and neurobiological origins. *Nat Rev Neurosci* **14**:143–152.

D'Argembeau A, Stawarczyk D, Majerus S, Collette F, Van der Linden M, Feyers D, Maquet P, Salmon E (2010) The neural basis of personal goal processing when envisioning future events. *J Cognitive Neurosci* **22**:1701-1713.

De Gelder B, Snyder J, Greve D, Gerard G, Hadjikhani N (2004) Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. *P Natl Acad Sci USA* **101**:16701-16706.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ, (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**:968-980.

Edelstein RS, Shaver PR (2007) A cross-cultural examination of lexical studies of self-conscious emotions. In: *The self-conscious emotions: Theory and research* (Tracy JL, Robins RW, Tangney JP, eds), pp194-208. New York: The Guilford Press.

Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *P Natl Acad Sci USA* **113**:7900-7905.

Ekman P (1992) An argument for basic emotions. *Cognition Emotion* **6**:169-200.

Ekman P (1999) Facial expressions. In: *Handbook of Cognition and Emotion* (Dalgleish T, Power M, eds), pp 301–320. New York: John Wiley & Sons Ltd.

Ekman P, Cordaro D (2011) What is meant by calling emotions basic. *Emot Rev* **3**:364-370.

Elliot AJ, Eder AB, Harmon-Jones E (2013) Approach–avoidance motivation and emotion: convergence and divergence. *Emot Rev* **5**:308–311.

Eryilmaz H, Van De Ville D, Schwartz S, Vuilleumier P (2011) Impact of transient emotions on functional connectivity during subsequent resting state: a wavelet correlation approach. *Neuroimage* **54**:2481-2491.

Ethofer T, Van De Ville D, Scherer K, Vuilleumier P (2009) Decoding of emotional information in voice-sensitive cortices. *Curr Biol* **19**:1028-1033.

Etkin A, Egner T, Kalisch R (2011) Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn Sci* **15**:85-93.

Fontaine JR, Scherer KR, Roesch EB, Ellsworth PC (2007) The world of emotions is not two-dimensional. *Psychol Sci* **18**:1050-1057.

Frijda NH, Kuipers P, Ter Schure E (1989) Relations among emotion, appraisal, and emotional action readiness. *J Pers Soc Psychol* **57**:212:228.

Friston KJ (2011) Functional and effective connectivity: a review. *Brain Connect* **1**:13-36.

Friston KJ, Frith CD, Fletcher P, Liddle PF, Frackowiak RS (1996) Functional topography: multidimensional scaling and functional connectivity in the brain. *Cereb Cortex* **6**:156-164.

Glerean E, Salmi J, Lahnakoski JM, Jääskeläinen IP, Sams M (2012) Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional connectivity. *Brain Connect* **2**:91-101.

Griffiths PE (1997) What Emotions Really Are: The Problem of Psychological Categories. Chicago: University of Chicago Press.

Hajcak G, Molnar C, George MS, Bolger K, Koola J, Nahas Z (2007) Emotion facilitates action: a transcranial magnetic stimulation study of motor cortex excitability during picture viewing. *Psychophysiology* **44**:91-97.

Hamann S (2012) Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends Cogn Sci* **16**:458-466.

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. *Science* **303**:1634-1640.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**:2425-2430.

Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**: 404-416.

Herbert C, Herbert BM, Pauli P (2011) Emotional self-reference: brain structures involved in the processing of words describing one's own emotions. *Neuropsychologia* **49**:2947-2956.

Holmes EA, Mathews A (2005) Mental imagery and emotion: a special relationship? *Emotion* **5**:489-497.

Hutcherson CA, Goldin PR, Ochsner KN, Gabrieli JD, Barrett LF, Gross JJ (2005) Attention and emotion: does rating emotion alter neural responses to amusing and sad films? *Neuroimage* **27**:656-668.

Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **76**:1210-1224.

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**:453-458.

Izard CE (1993) Four systems for emotion activation: cognitive and noncognitive processes. *Psychol Rev* **100**:68-90.

Izard CE (2011) Forms and functions of emotions: Matters of emotion–cognition interactions. *Emot Rev* **3**:371-378.

Jääskeläinen IP, Koskentalo K, Balk MH, Autti T, Kauramäki J, Pomren C, Sams M (2008) Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. *Open Neuroim J* **2**:14-19.

Kassam KS, Markey AR, Cherkassky VL, Loewenstein G, Just MA (2013) Identifying emotions on the basis of neural activation. *Plos One* **8**:e66032.

Kauppi JP, Jääskeläinen IP, Sams M, Tohka J (2010) Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency. *Front Neuroinform* **4**:5.

Klasen M, Kenworthy CA, Mathiak KA, Kircher TT, Mathiak K (2011) Supramodal representation of emotions. *J Neurosci* **31**:13635-13643.

Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD (2008) Functional grouping and cortical–subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage* **42**:998-1031.

Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G (2002) Hearing sounds, understanding actions: action representation in mirror neurons. *Science* **297**:846-848.

Kotz SA, Kalberlah C, Bahlmann J, Friederici AD, Haynes JD (2013) Predicting vocal emotion expressions from the human brain. *Hum Brain Mapp* **34**:1971-1981.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *P Natl Acad Sci USA* **103**:3863-3868.

Kragel PA, LaBar KS (2013) Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions. *Emotion*, **13**:681-690.

Kragel PA, LaBar KS (2014) Advancing emotion theory with multivariate pattern classification. *Emot Rev* **6**:160-174.

Kragel PA, LaBar KS (2015) Multivariate neural biomarkers of emotional states are categorically distinct. *Soc Cogn Affect Neur***10**:1437-1448.

Kragel PA, LaBar KS (2016) Decoding the Nature of Emotion in the Brain. *Trends Cogn Sci* **20**: 444-455.

Kreibig SD (2010) Autonomic nervous system activity in emotion: A review. *Biol Psychol* **84**:394-421.

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis–connecting the branches of systems neuroscience. *Front Syst Neurosci* **2**:4.

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R (1992) Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *P Natl Acad Sci USA* **89**:5675-5679.

Lang PJ (1995) The emotion probe – studies of motivation and attention. *Am Psychol* **50**:372–385.

LeDoux J (2012) Rethinking the emotional brain. *Neuron* **73**:653-676.

Levenson RW (2003) Blood, sweat, and fears – the autonomic architecture of emotion. In: *Emotions inside Out: 130 Years after Darwin's the Expression of the Emotions in Man and Animals* (Ekman P, Campos JJ, Davidson RJ, DeWaal FBM, eds), pp 348–366. New York: New York Acad Sciences.

Levenson RW (2011) Basic emotion questions. *Emot Rev* **3**:379-386.

Lewis MD, Liu ZX (2011) Three time scales of neural self-organization underlying basic and nonbasic emotions. *Emot Rev* **3**:416-423.

Lieberman MD, Eisenberger NI, Crockett MJ, Tom SM, Pfeifer JH, Way BM (2007) Putting feelings into words. *Psychol Sci* **18**:421-428.

Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF (2012) The brain basis of emotion: a meta-analytic review. *Behav Brain Sci* **35**:121-143.

Linnman C, Moulton EA, Barmettler G, Becerra L, Borsook D (2012) Neuroimaging of the periaqueductal gray: state of the field. *Neuroimage* **60**:505-522.

Logothetis NK (2008) What we can do and what we cannot do with fMRI. *Nature* **453**:869-879.

Mar RA (2011) The neural bases of social cognition and story comprehension. *Annu Rev Psychol* **62**:103-134.

Mazzola V, Vuilleumier P, Latorre V, Petito A, Gallese V, Popolizio T, Arciero G, Bondolfi G (2013) Effects of emotional contexts on cerebello-thalamo-cortical activity during action observation. *Plos One* **8**:e75912.

Meaux E, Vuilleumier P (2015) Emotion perception and elicitation. *Brain mapping: an encyclopedic reference, vol. 3* (Toga AW, ed), pp. 79-90. Oxford (UK): Elsevier.

Mobbs D, Petrovic P, Marchant JL, Hassabis D, Weiskopf N, Seymour B, Dolan RJ, Frith CD (2007) When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science* **317**:1079-1083.

Mobbs D, Yu R, Rowe JB, Eich H, FeldmanHall O, Dalgleish T (2010) Neural activity associated with monitoring the oscillating threat value of a tarantula. *P Natl Acad Sci USA* **107**:20582-20586.

Mulligan K, Scherer KR (2012) Toward a working definition of emotion. *Emot Rev* **4**:345-357.

Murphy FC, Nimmo-Smith IAN, Lawrence AD (2003) Functional neuroanatomy of emotions: a meta-analysis. *Cogn Affect Behav Neurosci* **3**: 207-233.

Naqvi NH, Rudrauf D, Damasio H, Bechara A (2007) Damage to the insula disrupts addiction to cigarette smoking. *Science* **315**:531-534.

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* **10**:424-430.

Northoff G, Bermpohl F (2004) Cortical midline structures and the self. *Trends Cogn Sci* **8**:102-107.

Northoff G, Heinzel A, De Greck M, Bermpohl F, Dobrowolny H, Panksepp J (2006) Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage* **31**:440-457.

Nummenmaa L, Glerean E, Viinikainen M, Jääskeläinen IP, Hari R, Sams M (2012) Emotions promote social interaction by synchronizing brain activity across individuals. *P Natl Acad Sci USA*, **109**:9599-9604.

Nummenmaa L, Glerean E, Hari R, Hietanen JK (2014a) Bodily maps of emotions. *P Natl Acad Sci USA* **111**:646-651.

Nummenmaa L, Saarimäki H, Glerean E, Gotsopoulos A, Jääskeläinen IP, Hari R, Sams M (2014b) Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. *Neuroimage* **102**:498-509.

Nummenmaa L, Saarimäki H (in press) Emotions as discrete patterns of systemic activity. *Neurosci Lett*.

Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *P Natl Acad Sci USA* **87**:9868-9872.

Öngür D, Price JL (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb Cortex* **10**:206-219.

Panksepp J (1982) Toward a general psychobiological theory of emotions. *Behav Brain Sci* **5**:407-422.

Panksepp J (2007) Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspect Psychol Sci* **2**:281-296.

Panksepp J, Watt D (2011) What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emot Rev* **3**:387-396.

Papez JW (1937) A proposed mechanism of emotion. *Arch Neurol Psychiatry* **38**:725–743.

Peelen MV, Atkinson AP, Vuilleumier P (2010) Supramodal representations of perceived emotions in the human brain. *J Neurosci* **30**:10127-10134.

Pessoa L (2017) A Network Model of the Emotional Brain. *Trends Cogn Sci* **21**:357-371.

Phan KL, Wager T, Taylor SF, Liberzon I (2002) Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage* **16**:331-348.

Pichon S, de Gelder B, Grezes J (2008) Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Soc Neurosci* **3**:199-212.

Plutchik R (1980) A general psychoevolutionary theory of emotion. In: *Emotion. Theory, Research, and Experience. Theories of emotion 1* (Plutchik R, Kellerman H, eds), pp. 3-31. New York: Academic press.

Politis DN, Romano JP (1992) A circular block-resampling procedure for stationary data. In: *Exploring the limits of bootstrap* (LePage R, Billard L, eds), pp. 263-270. New York: John Wiley & Sons Ltd.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. *Science* **310**:1963-1966.

Pourtois G, Sander D, Andres M, Grandjean D, Reveret L, Olivier E, Vuilleumier P (2004) Dissociable roles of the human somatosensory and superior temporal cortices for processing social face signals. *Eur J Neurosci* **20**:3507-3515.

Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE (2011) Functional network organization of the human brain. *Neuron* **72**:665-678.

Raz G, Touroutoglou A, Wilson-Mendenhall C, Gilam G, Lin T, Gonen T, Jacob Y, Atzil S, Admon R, Bleich-Cohen M, Maron-Katz A, Hendler T, Barrett LF (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cogn Affect Behav Neuroscience* **16**:709-723.

Reyes-Vargas M, Sánchez-Gutiérrez M, Rufiner L, Albornoz M, Vignolo L, Martínez-Licona F, Goddard-Close J (2013) Hierarchical clustering and classification of emotions in human speech using confusion ma-

trices. In: *International Conference on Speech and Computer. SPECOM 2013. Lecture notes in Computer Science, vol 8113* (Železný M, Habernal I, Ronzhin A, eds), pp. 162-169. Cham: Springer.

Rosvall M, Bergstrom CT (2010) Mapping change in large networks. *Plos One* **5**:e8694.

Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**:1059-1069.

Russell JA (1980) A circumplex model of affect. *J Pers Soc Psycho*l **39**:1161-1178.

Russell JA (2003) Core affect and the psychological construction of emotion. *Psychol Rev* **110**:145-172.

Said CP, Moore CD, Norman KA, Haxby JV, Todorov A (2010) Graded representations of emotional expressions in the left superior temporal sulcus. *Front Syst Neurosci* **4**:6.

Salzman CD, Fusi S (2010) Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annu Rev Neurosci* **33**:173-202.

Satpute AB, Shu J, Weber J, Roy M, Ochsner KN (2013) The functional neural architecture of self-reports of affective experience. *Biol Psychiatry* **73**:631-638.

Scherer KR (2000) Psychological models of emotion. In: *The neuropsychology of emotion*. (Borod JC, ed), pp. 137-162). Oxford: Oxford University Press.

Scherer KR (2009) The dynamic architecture of emotion: Evidence for the component process model. *Cognition Emotion* **23**:1307-1351.

Schutter DJ, van Honk J (2009) The cerebellum in emotion regulation: a repetitive transcranial magnetic stimulation study. *Cerebellum* **8**:28-34.

Shaw P, Bramham J, Lawrence EJ, Morris R, Baron-Cohen S, David AS (2005) Differential effects of lesions of the amygdala and prefrontal cortex on recognizing facial expressions of complex emotions. *J Cognitive Neurosci* **17**:1410-1419.

Skerry AE, Saxe R (2014) A common neural code for perceived and inferred emotion. *J Neurosci* **34**:15997-16008.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *P Natl Acad Sci USA* **100**:9440-9445.

Summerfield JJ, Hassabis D, Maguire EA (2009) Cortical midline involvement in autobiographical memory. *Neuroimage* **44**:1188-1200.

Särkkä S, Solin A, Nummenmaa A, Vehtari A, Auranen T, Vanni S, Lin FH (2012) Dynamic retrospective filtering of physiological noise in BOLD fMRI: DRIFTER. *Neuroimage* **60**:1517-1527.

Tettamanti M, Rognoni E, Cafiero R, Costa T, Galati D, Perani D (2012) Distinct pathways of neural coupling for different basic emotions. *Neuroimage* **59**:1804-1817.

Toivonen R, Kivelä M, Saramäki J, Viinikainen M, Vanhatalo M, Sams M (2012) Networks of emotion concepts. *Plos One* **7**:e28883.

Touroutoglou A, Lindquist KA, Dickerson BC, Barrett LF (2015) Intrinsic connectivity in the human brain does not reveal networks for 'basic' emotions. *Soc Cogn Affect Neur* **10**:1257-1265.

Trost W, Ethofer T, Zentner M, Vuilleumier P (2011) Mapping aesthetic musical emotions in the brain. *Cereb Cortex* **22**:2769-2783.

Van Essen DC (2005) A population-average, landmark-and surface-based (PALS) atlas of human cerebral cortex. *Neuroimage* **28**:635-662.

Vytal K, Hamann S (2010) Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *J Cognitive Neurosci* **22**:2864-2885.

Wager TD, Phan KL, Liberzon I, Taylor SF (2003) Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *Neuroimage* **19**:513-531.

Wang S, Tudusciuc O, Mamelak AN, Ross IB, Adolphs R, Rutishauser U (2014) Neurons in the human amygdala selective for perceived emotion. *P Natl Acad Sci USA* **111**:E3110-E3119.

Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G (2003) Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron* **40**:655-664.

Wierzbicka A (1999) Emotions across languages and cultures: Diversity and universals. Cambridge: Cambridge University Press.

Zahn R, Moll J, Paiva M, Garrido G, Krueger F, Huey ED, Grafman J (2008) The neural basis of human social values: evidence from functional MRI. *Cereb Cortex* **19**:276-283.

Zalesky A, Fornito A, Bullmore ET (2010) Network-based statistic: identifying differences in brain networks. *Neuroimage* **53**:1197-1207.

Emotions guide both human and animal behavior providing the means for survival in a constantly changing environment. Different emotions seem to be distinct from each other in several aspects, including physiological changes, bodily sensations, facial expressions, and subjective experience. Whether and how such emotion categories exist at the neural level remains however under debate. In the studies of this dissertation multiple emotional states were induced using emotional movies, mental imagery, and narratives while participants' brain activity was measured with functional magnetic resonance imaging. The findings from these studies show that specific emotions can be classified from both voxel activity and functional connectivity patterns, suggesting that emotions have distinct brain activity and connectivity patterns that encompass large extent of the brain and generalize both across individuals and across emotion elicitation techniques.

Aalto-DD 17/2018

ISBN 978-952-60-7817-5 (printed)
ISBN 978-952-60-7818-2 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Neuroscience and Biomedical Engineering
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS