IMT School for Advanced Studies Lucca

Lucca, Italy

Brain topography, connectivity dynamics and coding mechanisms underlying the subjective experience of affective states

PhD Program in Cognitive, Computational and Social Neuroscience

XXXII Cycle

Giada Lettieri
2019

The dissertation of Giada Lettieri is approved.

Program Coordinator: Pietro Pietrini, IMT School for Advanced Studies Lucca

Advisor: Pietro Pietrini, IMT School for Advanced Studies Lucca

Co-Advisor: Luca Cecchetti, IMT School for Advanced Studies Lucca

The dissertation of Giada Lettieri has been reviewed by:

Prof. Francesca Benuzzi, Università di Modena e Reggio Emilia

Prof. Michael Hanke, Forschungszentrum Jülich

Prof. John Van Horn, Keck School of Medicine of University of Southern California

IMT School for Advanced Studies Lucca
2019

Table of contents

List of figures

Acknowledgements

The chapter "Emotionotopy in the human right temporo-parietal cortex" reports material based on a co-authored paper accepted for publication on *Nature Communications* from Giada Lettieri, Giacomo Handjaras, Emiliano Ricciardi, Andrea Leo, Paolo Papale, Monica Betta, Pietro Pietrini and Luca Cecchetti. The affiliation of all authors is IMT School for Advanced Studies Lucca.

The chapter "Emotional experience timecourse explains brain connectivity dynamics during naturalistic stimulation" reports material based on a co-authored paper in progress from Giada Lettieri, Giacomo Handjaras, Emiliano Ricciardi, Andrea Leo, Paolo Papale, Monica Betta, Pietro Pietrini and Luca Cecchetti. The affiliation of all authors is IMT School for Advanced Studies Lucca.

I would like to thank my advisor Prof. Pietro Pietrini for his guidance and insightful comments, which always improved my research.

My sincere gratitude goes to my co-advisor Luca Cecchetti for the continuous support during my PhD, for his patience, motivation, and immense knowledge. Words are limited to express how lucky I was to have his guidance during these three years. Plus, our laughs along the way were really the icing on my PhD cake.

I wish to express my gratefulness to Giacomo Handjaras for his encouragement and advice during these three years and for always showing me how to see life in autumn colors. Also, a special appreciation goes to Emiliano Ricciardi.

Lastly, I thank my family for always be there, no matter how stormy the day is. Thanks for constantly staying right behind me, ready to catch me if I fall and to raise the glass if I succeed.

Vita and publications

1991	Born, Caserta, Italy
2014 MSc	Cognitive Neuroscience (Università Vita Salute San Raffaele, Milano, Italy)
2012 BSc	Science and Psychological Methodologies (Università Vita Salute San Raffaele, Milano, Italy)

Publications (selection)

Lettieri, G., Handjaras, G., Ricciardi, E., Leo, A., Papale, P., Betta, M., ... & Cecchetti, L. (2019). Emotionotopy in the human right temporo-parietal cortex. *Nature Communications*, in press.

Gentili, C., Messerotti Benvenuti, S., Lettieri, G., Costa, C., & Cecchetti, L. (2019). ROI and phobias: The effect of ROI approach on an ALE meta-analysis of specific phobias. *Human brain mapping*, 40(6), 1814-1828.

Cecchetti, L., Lettieri, G., Handjaras, G., Leo, A., Ricciardi, E., Pietrini, P., ... & Train the Brain Consortium. (2019). Brain Hemodynamic Intermediate Phenotype Links Vitamin B12 to Cognitive Profile of Healthy and Mild Cognitive Impaired Subjects. *Neural Plasticity*, 2019.

Tramonti, C., Imperatori, L. S., Fanciullacci, C., Lamola, G., Lettieri, G., Bernardi, G., ... & Chisari, C. (2018). Predictive value of EEG connectivity measures for motor training outcome in multiple sclerosis: an observational longitudinal study. *European journal of physical and rehabilitation medicine*.

Poster presentations (selection)

Lettieri G, Handjaras G, Ricciardi E, Leo A, Papale P, Betta M, Pietrini P, Cecchetti L. (2019). Emotional experience timecourse explains brain connectivity dynamics during naturalistic stimulation. *Organization for Human Brain Mapping Annual Meeting*, Rome, June 09-13, 2019.

Lettieri G, Handjaras G, Ricciardi E, Leo A, Papale P, Betta M, Pietrini P, Cecchetti L. (2019). Emotionotopy: gradients encode emotion dimensions in right temporo-parietal territories. *Social and Affective Neuroscience Society Annual Meeting*, Miami, May 02-04, 2019.

Lettieri G, Cecchetti L, Handjaras G, Leo A, Papale P, Betta M, Ricciardi E, Pietrini P. (2018). Emotions induced by naturalistic stimuli explain right hemisphere activity in an independent sample. *Organization for Human Brain Mapping Annual Meeting*, Singapore, June 17-21, 2018.

Oral presentations

Lettieri G. (2019). Emotional experience timecourse explains brain connectivity dynamics during naturalistic stimulation. Organization for *Human Brain Mapping Annual Meeting*, Rome, June 10, 2019.

Lettieri G. (2018). Perceived emotional experience explains brain activity elicited by an emotionally-charged movie in an independent sample. *XXVI Congresso Nazionale SIPF*, Torino, November 16, 2018.

Lettieri G. (2018). Emotions induced by naturalistic stimuli explain right hemisphere activity in an independent sample. *Organization for Human Brain Mapping Annual Meeting*, Singapore, June 21, 2018.

Honors, awards and accomplishments

Merit Award for the oral presentation: Emotional experience timecourse explains brain connectivity dynamics during naturalistic stimulation. *Organization for Human Brain Mapping Annual Meeting*, Rome, June 09-13, 2019.

Poster Award for the poster: Lettieri G, Handjaras G, Ricciardi E, Leo A, Papale P, Betta M, Pietrini P, Cecchetti L. (2019). Emotionotopy: gradients encode emotion dimensions in right temporo-parietal territories. *Social and Affective Neuroscience Society Annual Meeting*, Miami, May 02-04, 2019.

Best scientific contribution for the oral presentation: Perceived emotional experience explains brain activity elicited by an emotionally-charged movie in an independent sample. *XXVI Congresso Nazionale SIPF*, Torino, November 16, 2018.

Abstract

The present dissertation focuses on the behavioral and neural substrates of socio-affective abilities involved in the subjective emotional experience. Socio-affective abilities are the building blocks of the more general domain of social cognition, which represents the cornerstone of human interactions. In particular, understanding, predicting and responding to others' emotional signals are fundamental aspects necessary for the optimal functioning of human's daily life. Emotions are pervasively present in dyadic interactions, give color to individual experiences, and can rapidly change over time, as they are the consequence of an active interplay between an individual and the environment. Despite their relevance, however, emotions are still an open question for researchers.

Indeed, which mechanisms link descriptions of affective states to brain activity is still unclear, with evidence supporting either local or distributed processing. Moreover, the majority of neuroimaging studies so far did not take into account the dynamism of affective states and their unfolding over time. For this reason, how the temporal characteristics of emotions (e.g., duration, onset, resurgence) are represented in the brain, with the dynamics between specific regions related to different emotional experiences is an open question.

In light of all this, the studies reported in the present dissertation aimed to overcome previous limitations and answer these questions. In the first study we used brain hemodynamic activity evoked by an emotionally charged movie and continuous ratings of the perceived emotion intensity to reveal the topographic organization of affective states. In the second study, we explored the dynamic interplay between different brain regions throughout a naturalistic situation. To do so, we related continuous ratings of the perceived intensity of various emotional states to changes in functional connectivity among distinct brain regions during the watching of the same movie employed before.

Our results showed that moment-by-moment ratings of perceived emotions explain brain activity recorded in independent subjects. Most importantly, we demonstrated the existence of orthogonal and spatially overlapping right temporo-parietal gradients encoding emotion dimensions, a mechanism that we named *emotionotopy*. We also unveiled the central role of the right precentral sulcus during the subjective emotional experience, with changes in the functional connectivity dynamics of this region being modulated by three cardinal emotion dimensions.

1. Preamble

The present dissertation focuses on the behavioral and neural substrates of socio-affective abilities involved in the subjective emotional experience.

Socio-affective abilities are the building blocks of the more general domain of social cognition, which represents the cornerstone of human interactions and can be defined as the sum of those processes that allow individuals of the same species to relate with one another (Fiske and Taylor, 1991; Frith and Frith, 2007). Indeed, successful social exchanges require several skills, including the careful observation of other people's behavior, the prediction of their reactions in relation to ourselves and appropriate responses to their acts (Frith and Frith, 2006). In this sense, understanding, predicting and responding to others' signals are fundamental aspects, necessary for the optimal functioning of human's daily life (Kunda, 1999). Dissecting the social cognition domain, essential abilities include *emotion recognition*, *empathy* and *mentalizing*, which play specific roles in the production of adaptive responses and, by acting altogether, give rise to the complexity of the social world.

In particular, *emotions* play a fundamental part in human life, being pervasively present in dyadic interactions and giving color to individual experiences. Despite their relevance, emotions are still an open question for researchers. Indeed, the term "*emotion*" has been frequently used to interpret and give meaning to a gamut of feelings, behaviors and bodily sensations taking place at the same time, which are significantly difficult to interpret. To do so, humans are used to assign a certain label to these states, supposedly shared and understood by the vast majority of individuals (e.g., anger, regret, sadness). However, there is far more complexity behind this words that researchers are still struggling to clarify.

Empathy has instead been described as the set of processes associated to represent and share affective states and feelings of a person different from ourselves (Bohart et al., 2002; Decety and Jackson, 2004). In particular, *empathy* allows us to imagine as if we were experiencing the other's situation and to perceive the same emotions felt by another individual. It is important to note that this experience is distinct both from *sympathy*, which arise from the awareness of others' affective states and described with feeling of sorrow and concern (Preston and De Waal, 2002), and from *personal distress*, characterized by the desire of ease only one's own uncomfortable feeling (Batson and Shaw, 1991; Decety and Ickes, 2009). An important component and prerequisite for *empathy* is *emotional contagion*, which is the ability of people to feel themselves as physically being into another individual affective state (Hatfield et al., 1994). This process has been considered at the very basis of the entire domain of human social cognition, since it runs below the higher order processes, aiding in mind reading and allowing people to recognize and share other's feelings.

Notably, understanding other's intentions and emotions represents a fundamental pillar to attribute a meaning to their actions and feelings, a process that has been called *mentalizing*. In particular, the act of reading other people's minds is usually described as a *Theory of Mind* (*ToM*). The *ToM* has been first defined by Premack and Woodruff (1978) as the process by which "an individual imputes mental states to himself and others". Furthermore, *ToM* is currently recognized as requiring the integration of several components, including the ability of attributing both intentions (i.e., *cognitive ToM*) and emotions (i.e., *affective ToM*) to individuals (Shamay-Tsoory and Aharon-Peretz, 2007). In this regard, it is evident how *ToM* is strictly associated to *emotion recognition* and *empathy*, in a way that researchers are struggling to identify the existing relationship between these abilities.

In light of all this, it is clear that social interactions are characterized by a series of independent processes which then combine into a complex final percept, difficult to entirely disentangle. Therefore, it is particularly challenging to study each individual component taking part in the social domain. It is also evident the fundamental role that socio-affective abilities play in the human daily life, and their importance for the good functioning of an individual.

The aim of this chapter is to provide a multi-disciplinary background introduction, describing the different socio-affective human abilities and what is currently known about their representation in the brain.

2. Introduction

2.1 Emotions

2.1.1 Definition and perspectives

"Everyone knows what an emotion is, until asked to give a definition" (Fehr and Russell, 1984).

In 1872 Charles Darwin published his *"The Expression of the Emotions in Man and Animals"*, being one of the first scholars to use the term *emotion* as we intend nowadays. In this book, he sustained that emotions are fundamental for human beings as the act of breathing or digesting, are shared with animals, and have been essential to our survival, shaped by biological intervening factors. However, the real debate about the definition of an emotion actually started in the 19th century, when the psychologist William James published a revolutionary essay titled "What is an emotion?", arguing that these states are nothing more than bodily reactions to salient stimuli (James, 1884).

The overwhelming attention that has been directed towards this theory and its definition of affective states shows how emotions are a central topic for researchers, that are presently still struggling to fully unravel. Indeed, notwithstanding a hundred years of debates, today it is still possible to receive as many different answers as many people are asked the question "what is an emotion?".

Across the years, the enduring interest on this topic has involved scholars from a wide range of disciplines, including psychology, philosophy, neuroscience, economics, and sociology. While this attention has resulted in a voluminous body of literature on emotions, it has also delineated a chaotic field of research, surprisingly characterized by less knowledge accumulated compared to other psychological domains as cognitive functions. Indeed, due to this lack of consensus on fundamental questions and definitions, the progresses made so far in the field of affective states have been significantly compromised.

An important part of confusion comes from the use of the term emotion both for physiological responses to certain stimuli (e.g., during a fearful event the heart starts pounding and our pupils dilate) and for conscious experiences that accompany these somatic and cognitive changes, which have also been labeled as *feelings* (e.g., our acknowledgment of perceiving fear). Actually, years of research have been grounded on the dualism between emotion and cognition: the first being associated to implicit, rapid and automatic processing, while the latter to explicit, slow and deliberate processing. These two routes have been mainly viewed in an asymmetric way, with a possible cognitive regulation of emotions but no influence of the latter on cognition.

However, this dualism has been recently challenged, as it has been highlighted how reasoning actually requires emotions, and many of the complex decisions we make in our daily life are impossible without them (Bechara et al., 2000; Goleman, 2006; Lerner et al., 2015). Starting from this point of view, research in the past decades has described how emotions interact and are integrated with cognition (LeDoux, 2000; Pessoa, 2013), and it is now extensively recognized how they are involved and influence a wide range of psychological processes as attention, memory and perception (Fenske and Eastwood, 2003; LaBar and Cabeza, 2006; Phelps et al., 2006; for a review see Brosch et al., 2013). For example, moral reasoning and decision-making processes associated to it hold a strong affective component (Haidt, 2003; Greene et al., 2004), as well as economic decisions (Peters et al., 2006; Heilman et al., 2010), stereotypes and in-group/out-group effects (Esses et al., 1993). Nonetheless, while this view provides a more detailed and complete framework for the description of emotions, it also introduces a higher order of complexity to assess and investigate them in their "pure" form, as they are so intermingled and related to other psychological and cognitive processes. As it has been recently stated, "[…] No one would ever mistake seeing for hearing […], but the same cannot be said for feeling and thinking" (Barrett et al., 2007).

For this reason, valuable current models have attempted to delineate a set of characteristics underlying emotional experiences, in order to isolate and differentiate them from other cognitive occurrences. Indeed, there is a certain agreement on the view of emotions as responses to external or internal events characterized in a positive or negative way that:

1. Implicate a response from several different systems (e.g., behavioral, autonomic; Cacioppo et al., 2000);
2. Have a phasic nature, identifiable objects or triggers, and follow a specific timecourse, contrary to *mood* (Frijda, 1993);
3. Implicate appraisal processes to assess the significance of the stimulus prompting the emotional experience (Scherer et al., 2001).

Starting from the first point of this list, researchers focused on the different systems involved in the experience of an emotion and tried to dissect and describe specifically the different components implicated. Over the years, four of them have been consistently found:

1. *Cognitive*
2. *Motivational*
3. *Subjective*
4. *Somatic*

The *cognitive* component is the conscious or unconscious attribution of meaning or emotional connotation to an object or an event, which usually is translated as the experience of a *feeling*. The *motivational* one is instead related to the motor aspect of emotions, which is what drives the individual to physically and behaviorally react to the affective state he/she is feeling. On the other hand, the *subjective* component is the inner part of an emotion, specific to the individual, and requiring the acknowledgment of all those feelings, beliefs, desires and bodily sensations involved in that affective state and the effort to label it accordingly. Lastly, the *somatic* one is related to the responses of the autonomic and central nervous system associated to that percept. An important aspect is how these components interact, giving rise to various durations and intensity of emotions (Frijda et al., 1992; Frijda, 1993). Indeed, different theories have been trying to investigate these four components at the same time, having a particular difficulty in objectively assess the subjective one. Others instead focused on one of these components in isolation, as for example cognitive theories delineating appraisal and reappraisal processes (Nussbaum, 2003; Solomon, 2007), without considering the context in which the individual is acting. Moreover, for the definition of an affective state, it might be useful to consider what does not qualify as an emotion. Indeed, emotions are related to an external or internal event, and not all "emotional" states have an object of interest. For example, *sadness* refers to a certain negative event that just happened to us, a specific picture or letter describing something unpleasant, while a *depressive* state is not focused on something in particular, but everything feels grey and negative.

It is also important to note that the kinds of stimuli that can cause affective reactions tend to be systematically linked to a certain emotion. As a matter of fact, the objects or events that cause fear, sadness, or happiness, tend to do so fairly consistently in individuals that share the same socio-cultural background, eliciting similar reactions across the four components described above. This evidence is considered as a clear proof of the importance of evolution for the development of emotions (LeDoux, 2012). Indeed, according to several authors, the general purpose of an emotion is to produce a behavioral response and a specific change in the internal state of an individual, useful to react to a certain situation. In light of this, for specific unsafe or advantageous stimuli, evolution has aided us in developing a corresponding answer in the form of an emotion, and this would be the reason why we can find the same responses across individuals and cultures. However, regardless of the biological root of the emotional experience, each individual characteristics, development, culture and society still play an important role in shaping affective states. All these factors together are indeed contributing to the final expression of emotions, the cognitive processes and behavioral acts that follow the unfolding of that state, and the kind of stimuli reliably associated to it.

Furthermore, it is also important to note that the objects or events associated to different emotions cannot be only the ones defined by nature during evolution or available in our early development. Individuals develop and gain emotional experiences from the relationship with different others, animals, events, and by simply acting and interacting with the environment. Therefore, the individual will be able to associate a neutral stimulus to a particular affective state, so that the range of events and objects eliciting an emotion is potentially infinite. In one way or another, most of the things we experience will be perceived as connoted by some emotional parts, which can then provoke a more strong or weak reaction depending on each person characteristics.

In light of all this, it seems clear the prominent role in the human experience that emotions play, affecting the way we see the world, remember events, interact with others and make decisions.

However, the experience of an *emotion* is extremely complex and multifaceted: besides the most basic processes involving the recognition of facial expressions and body postures of someone else, it is also related to the interaction between one's own belief about the world, oneself and other people, the ability to understand context and to imagine possible future consequences and scenarios. These emotional components dynamically interact, demonstrating specific response patterns that spread out over time, highlighting the difficulty in studying their appearance, development and unfolding.

Despite the ease with which these experiences are categorized and labeled, emotional states are often too private, easily influenced and difficult to quantify. Indeed, due to the highly subjective and often confusing internal cascade of events related to affective states, their empirical study is considerably challenging. How emotions are created and represented in the brain might be the most important and interesting question in socio-affective research.

2.1.1.1 Discrete theories of emotions

Discrete theories of affective states postulate the existence of a limited number of emotions, each of them characterized by specific properties shared across individuals and cultures, and involving distinct patterns of response in the physiology, hemodynamic activity and facial expressions of the experiencing human being (Ekman, 1992; Izard, 1993).

In the last thirty years, one of the most influential discrete theory has been the "basic emotion" one, first developed by Paul Ekman (1989, 1992) and extensively revised and appreciated. According to this view, "emotions are specific and consistent collections of physiological responses triggered by certain brain systems when the organism is facing particular objects or situations" (Ekman, 1992). These responses are predetermined by the human genome and are the result of the entire history of human evolution. Hence, the basic emotion theory postulates that humans are equipped with a specific number of discrete emotions with a strong biological root, easily distinguishable one from the other.

Indeed, to be considered "basic" an emotion has to have a set of specific properties, including:

1. Be discrete;
2. Have fixed neurobiological, physiological and bodily responses;
3. Have a fixed association to a specific feeling and following pattern of reaction;
4. Be psychologically primitive.

In light of all this, basic emotions are supposed to be the most primary, culturally universal and with an inherited biological basis, affective states. Overall, if there is a general consensus on the key criteria to consider affective states as basic emotions, there is a greater disagreement on other aspects related. First of all, it is not clear how much basic emotions are actually present in the human life. While some authors sustain that basic emotions are fundamental in early development but then evolve into more complex states in adults (Panksepp, 1982; Izard, 1984; Levenson, 2011), others do not agree and assert that even emotional reactions associated to learning and cultural development might be considered basic (Ekman and Cordaro, 2011).

Furthermore, among the discrete theories of emotions, it has been particularly active the quest for a specific number of them, with the list of included affective states notably varying (Ekman, 1989, 1992; Izard, 1993, 2013; Plutchik and Kellerman, 2013). However, even using different labels, in the majority of discrete emotion theories is almost always included a positive emotion, named happiness, enjoyment or play, and three negative states (i.e., sadness, fear and anger), while there is a greater disagreement on pride (Tracy and Randles, 2011). According to Ekman's theory, the emotion categories that should be considered as basic are happiness, surprise, fear, sadness, anger and disgust.

Moreover, the evolutionary approach views facial expressions as a fundamental part of the emotional response, since they are reliably found across different cultures (Ekman and Friesen, 1971; Ekman et al., 1987; Ekman, 1989), are associated to involuntary muscle actions (Ekman et al., 1987; Rosenberg and Ekman 1995), and are useful to coordinate social interactions through their informative and evocative function (Keltner and Kring, 1998). The claim regarding the universality of facial expressions of discrete emotions is rooted in the notion that the facial anatomy is the same for all individuals (Gray, 1966) and, more importantly, several studies have been consistently found that the facial musculature is activated in emotion specific ways across different cultures (Ekman, 1972; Matsumoto, 1992; Izard, 1994; Ekman et al., 1997; Fritz et al., 2009).

Following the strong biological root of basic emotions, characterized by universal and reliable markers which are difficult to modify or manage, as humans we are bound to the repertory of affective states that evolution equipped us, as we cannot create novel emotions or feelings.

Overall, the basic emotions position has some valuable aspect:

1. It attempts to organize affective phenomena, by highlighting differences between basic emotions and other states;
2. It provides a clear framework useful in empirical research to investigate whether an affective state might be a basic emotion or not;
3. It guided new lines of research which contributed to provide a detailed description of relevant emotions, as the six basic ones (i.e., happiness, surprise, fear, sadness, anger and disgust).

However, this approach has also been criticized as considered too limited and narrow, with the variety of affective states perceived by humans in daily life impossible to be categorized only as basic emotions.

2.1.1.2 Dimensional theories of emotions

In contrast with discrete theories, dimensional theories conceptualize emotions as arising from the combination of more fundamental components and all affective states can be described following these cardinal dimensions.

One of the main dimensional models is the circumplex model of affect (Russell, 1980). According to this, there are just two main largely independent neurophysiological systems: *valence* and *arousal*. The first one describes the polarity of the emotional experience, whether is something positive or negative, while the second one explains the intensity of the perceived affective state and determines how much it is behaviorally activating. Valence is considered to range from highly positive experiences, expressed as joyful or exuberant, to highly negative events, described as wretchedness or desperation (Frijda, 1986; Lang et al., 1993; Russell, 2003). Arousal is instead viewed as the alertness of an organism for action and its level can go from drowsiness and sleep on one side to intense excitement on the other (Lang et al., 1993; Russell, 2003).

From these two cardinal dimensions, we can interpret and give meaning to all our emotional experiences. For example, a fearful state or experience is actually generated from a mixture of negative valence, high arousal and other attributes that are not specific to fear *per se*. Moreover, when subjects are asked to complete self-reports of their emotional experience, vary in how they do so, with some of them more inclined to use discrete categories of emotions, while others describing their state in a more global and broad way. These ratings, when projected into a geometric space, take a circular-like form, or circumplex (Figure 1), indicating that self-reports of emotional experience are heterogeneous and can be decomposed into at least two or more properties (Russell, 1980; Feldman, 1995; Barrett and Russell, 1999).

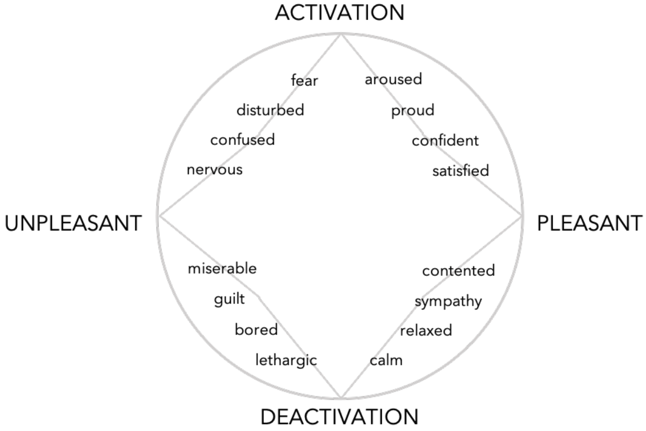


Figure 1. The circumplex model of affect.

As in the case of discrete theories of emotions, there is a general disagreement on the number and types of dimensions needed to represent the variety of affective states. Different models have been developed in this regard, with a varying number and type of dimensions included, as for example *approach* and *withdrawal* (Lang et al., 1990).

The dimensional view is of particular interest since it is parsimonious and generalizable across multiple domains, and it has also been proven to account for a wide range of emotion effects. However, it is still not clear how the cardinal dimensions combine together and with other processes to generate the final emotion percept.

2.1.1.3 Constructivist theories of emotions

Already James and Wundt in the 19th century asserted that it should be adopted a constructionist model, which is in open contrast with the idea that we can identify in the brain distinct categories, as emotions, having their own biological core and a specific physiological pattern. According to the original theory, states, reactions and thoughts can be reduced to common operations or elements. This model has been recently reconsidered and revaluated in the domain of affective experiences. The constructionist approach to emotions assumes that affective states are the result of more basic processes which are called operations, building blocks or ingredients, and that each of these components is associated to a general function.

The building blocks at the basis of emotional experiences are two different sets of psychological operations, one related to inner sensations and the other one to exteroceptive feelings. These two classes of states are then interpreted and give meaning to through other psychological processes, including motivational, physiological and cognitive components, together with each individual characteristic (Barrett, 2009; Lindquist et al., 2012).

Therefore, according to the constructivist point of view, the biological aspect of emotions actually does not exist, since the final percept is built considering past memories, the cultural context in which the individual was born, previous similar events that he experienced and learned from, and the social world surrounding him. In this sense, emotions cannot be inherited and stereotyped across subjects and cultures, instead they assume situation-specific forms and are generated from the interaction of many different components (e.g., cognitive, motor, perceptive) that are needed to properly response to what the individual is facing.

A constructivist approach is valuable since it motivates important shifts in research paradigm and tries to overcome the current limitations related to other models. Indeed, this view allows to describe the core aspects of inner experiences and can also be generalized to other cognitive and affective events, apparently very distant. Nevertheless, this approach also holds some weaknesses, as it neglects relevant neurobiological findings and does not take into account the commonalties between emotional experiences.

2.1.2 The assessment of emotion recognition and behavioral methods to evaluate the subjective emotional experience

As described in the previous section, the emotional response can be divided in four different components, each leading to a specific change in the state of the individual. The responses across these domains have been divided into five output:

1. Overt acts and behavioral reactions, ranging from "survival" kinds (e.g., fight or flight) to more complex ones (e.g., lack of concentration due to anxiety, spacing out after a negative experience);
2. Language expressions, in the sense of emotional communication (e.g., verbal attack, scream);
3. Bodily and facial motor expressions;
4. Self-evaluative reports of feelings and responses experienced by the individual;
5. Physiological reactions which support the emotion experienced (e.g., blushing for embarrassment, heart pounding for fear, sweat for anxiety).

This organization is valuable since it provides a simpler framework to describe the complexity of emotional experiences and tries to reduce it by identifying core components related to it. Of note, it is significantly challenging for researchers to conceive an experimental paradigm able to investigate all components together. Indeed, the central challenge in the investigation of emotions is to provide efficient measurement of each domain and achieve an optimal integration across them. Therefore, the majority of behavioral and neuroimaging studies so far have been exploring one or a combination of two outputs at maximum.

Furthermore, to study emotions in the laboratory setting it is necessary to use stimuli that are able to reliably and consistently evoke congruent responses to the kind of affective state that researchers aim to elicit, similarly to what happens in real life. Across the years, a variety of approaches have been proposed and successfully employed to evoke emotional experiences in experimental subjects, both in healthy and pathological conditions.

One of the most used set of stimuli is composed by a set of colored pictures (IAPS; International Affective Picture System; Lang et al., 1997), sounds (IADS; International Affective Digitized Sounds System; Bradley & Lang, 1999a) and words (ANEW; Affective Norms for English Words; Bradley & Lang, 1999b).

The IAPS includes more than 1,000 images taken as examples of different human experiences, including facial expressions, places and events eliciting a wide variety of affective states. This set of pictures has been extensively employed in empirical investigations since it covers a gamut of emotional percepts, ranging from the most pleasant to the most unpleasant ones. Of note, the IAPS includes events commonly encountered such as landscapes and house furniture as well as rarely experienced or extremely shocking objects as mutilated bodies. After being collected, each image has been rated by a large sample of subjects (n = 100), including males and females, according to the level of valence, arousal and dominance that the picture was evoking. To do so, subjects used a valuable and easy affective rating system, the Self-Assessment Manikin (SAM; Bradley and Lang, 1994). In this rating, subjects are presented with a stylized figure depicting values along the three dimensions on a continuous scale that they can use to report their emotional reactions.

Instead, the IADS is a database of 111 sounds that have been scored according to the affective dimensions of valence and arousal. Sounds included in this database have been selected to be highly emotionally evocative and are sampled from several semantic categories. To complete the two collections of the IAPS and IADS, the Center for the Study of Emotion and Attention also collected the ANEW, which provides a set of normative emotional ratings of a gamut of English words along valence and arousal. This database of stimuli has been successfully and vastly employed across experiments and populations of interest, providing a valuable instrument for researchers to explore the variety of human affective states (Stevenson et al., 2007; Grühn and Scheibe, 2008; Liu et al., 2010; Schneider et al., 2016).

Furthermore, to investigate the processes related to the recognition of emotions in human individuals, different tests have been developed and standardized. One of the most commonly used is the Ekman 60 Test of Facial Affect (EK-60F; Ekman and Friesen, 1976). This task examines the ability to recognize facial expressions of emotions and can be used both in healthy and impaired subjects. The EK-60F has been conceived following the basic emotion theory developed by Ekman, and therefore employs only basic emotions. It is composed by a set of 60 pictures taken from ten different actors, including males and females, trained to reliably express via facial activity happiness, surprise, fear, sadness, anger and disgust.

Along these lines, several other stimuli have been developed, as for example:

- The Karolinska directed emotional faces (KDEF; Lundqvist et al., 1998) is a standardized set of affective stimuli which includes 70 individuals (35 female, 35 male) displaying six basic emotions (i.e., happiness, surprise, fear, sadness, anger and disgust).
- The Penn emotion recognition test (Kohler et al., 2004) comprises 96 color images of several actors coming from different ethnicities, expressing varying degrees of happiness, sadness, anger and fear, together with neutral expressions. This test has been mainly employed to explore affective impairments in psychiatric conditions (Sachs et al., 2004; Weiss et al., 2007) and neurodegenerative disorders (Kohler et al., 2005; Weiss et al., 2008).
- The awareness of social inference test (TASIT; McDonald et al., 2003) is employed to verify the ability to recognize spontaneous emotional expressions and it is composed by vignettes representing everyday social interactions, in which characters depict basic emotions, including neutral states. This test has been developed particularly to explore emotion perception in individuals after a traumatic brain injury (McDonald et al., 2003, 2004).
- The Florida Affect Battery (FAB; Bowers et al., 1998) test the ability to recognize facial and prosodic expressions of emotions with and without semantic distraction. It includes photographs and auditory recordings expressing happiness, sadness, anger, fear and neutral states, providing an easy to administer test particularly valuable in impaired populations (Jacobs et al., 1995; Shaw et al., 1999; Yim et al., 2013).

Of note, all these stimuli focused on the affective information expressed through facial activity. However, it is also of value for research on the emotional experiences to explore signals conveyed via the auditory channel. To this purpose, different auditory stimuli have been developed, as for example the Vocal expressions of 19 emotions across cultures corpus (VENEC; Laukka et al., 2010). This database has been recently developed and it includes recording from 100 professional actors from 5 cultures speaking English (i.e., USA, India, Kenya, Singapore and Australia) who vocally expressed 19 different affective states, ranging from the basics to the secondary ones, including anger, contempt, guilt, sadness and relief among others. Each state is also expressed with three levels of intensity, reaching a total of about 6500 stimuli. Therefore, the VENEC covers a wide gamut of emotions which can be elicited in different kind of populations in experimental conditions. Likewise, also the Montreal Affective Voices set (MAV; Belin et al., 2008) has been developed for this purpose. This corpus gathered 90 nonverbal affect bursts corresponding to the emotions of anger, disgust, fear, pain, sadness, surprise, happiness, pleasure and neutral.

In recent years, a new line of research has been affirmed, asserting the importance to use ecological conditions in the affective domain. Indeed, it is difficult to explore emotions as we experience them in real-life situations when presented with static images that do not correspond to what humans are commonly used to encounter. To solve this issue and try to overcome this limitation, several researchers have put some effort in developing controlled stimuli closely resembling everyday life events. Among these, video stimuli including short movies are an optimal compromise to maintain the experimental rigidity and at the same time to generalize findings to real life experiences.

The movie for the assessment of social cognition (MASC; Dziobek et al., 2006) is an example of this valuable effort. The test is video-based and evaluates at the same time not only skills related to emotions, but also various abilities pertaining to the social cognition domain. It is composed by a video in which characters are either strangers or friends gathered for a dinner. During the course of the dinner, protagonists interact and engage in social exchanges, closely resembling real life situations. This paradigm is relevant for emotions, as in the course of the evening all the characters experience different events that elicit a variety of affective states such as anger, jealousy, embarrassment, or disgust. The use of naturalistic conditions to study emotions and affective occurrences is of particular relevance to explore innerly experienced states. Furthermore, it opens new lines of investigation for the physiological and brain correlates of emotional experiences as they unfold.

2.1.3 Emotions in the brain

It is easily imaginable how all the components partaking in emotional experiences have their central core in the human brain. Indeed, the cognitive, motivational, subjective and somatic mechanisms elicited by affective states could not take place without the intervention of neurobiological processes. Therefore, there is a general consensus in the scientific community that emotions are represented in the brain, yet it is still debated how this representation takes place.

In this regard, the lesion method provides unique insight into how the human brain generates emotion (Feinstein, 2013). Indeed, a lesional approach may offer complementary information to neuroimaging studies by showing if a certain brain region is actually required for the investigated function (Hillis, 2014). However, this research also has some limitations, as the usually gathered sample size of patients suffering stroke and other focal injuries is relatively small. Therefore, it is important to stress the need for a multidisciplinary approach to clear how emotions are represented in the brain, for example involving neuroimaging and neuropsychological researches.

So far, the literature has shown that, following lesions and damage to certain brain structures, patients can show different emotion disturbances, ranging in severity and domain (Bechara, 2004). One region that has been at the center of the attention and considered the core of the "emotional" system, is the insular cortex. In fact, several studies have found a significant role of this brain region in the processing of disgust and, more generally, in the perception of others' and own feelings (Brown et al., 2011; Lamm et al., 2011). In this regard, a patient with a wide lesion involving the left insula was not able to recognize and perceive disgust (Calder et al., 2000). However, other investigations have failed to replicate this observation (Straube et al., 2010; Couto et al., 2013). Moreover, a recent single case study did not find an impairment in the experience of emotions in a patient suffering from extensive damage to bilateral insula (Damasio et al., 2012). Importantly, an extremely rare patient named Roger, provided a valuable opportunity to better examine the role of the insular cortex in affective states. Roger suffered from a severe encephalitis that completely damaged his limbic system, including the insular cortex, and he was still able to feel emotions in his daily life (Feinstein et al., 2010). This case, together with the aforementioned evidence, would show how the processing of affective states is complex and distributed throughout the brain.

On the other hand, also the amygdala has been considered fundamental for the recognition and perception of fear since decades (Davis, 1992). Of note, studies of patients reporting bilateral amygdala lesions have been crucial to establish this knowledge, as the majority of them reported specific deficits in the recognition of fear (Adolphs, 2008). However, so far, mainly basic emotional processes such as recognition and conditioning have been investigated in amygdala lesion patients (Feinstein, 2013). Therefore, very little is known about how the conscious experience of fear is altered following amygdala damage. A case study reported a woman with focal bilateral amygdala lesions and a marked absence of fear, even during exposure to a wide range of fear-provoking stimuli (Feinstein et al., 2011). By contrast, this patient had no difficulty expressing or experiencing a range of other emotions, including happiness, anger, sadness, disgust and surprise. This finding seems to suggest that the amygdala play a relevant role for triggering a state of fear when an individual encounter threatening stimuli in the external environment. Yet, it has been observed that the reliability of fear-specific deficits following amygdala damage is actually low (Adolphs et al., 1999).

Moreover, a particular interest has been devoted to the differential role of the brain right and left hemispheres in emotion perception (Gainotti, 1972). Currently, there are two major hypotheses regarding hemispheric specialization for emotion perception, and studies of unilaterally brain-damaged patients are crucial in testing them (Kucharska-Pietura et al., 2003). The first is the right hemisphere hypothesis, which emphasizes the dominance of the right hemisphere in emotion perception, regardless of valence and processing mode (Adolphs et al., 2000; Borod, 2000). The second is the valence hypothesis, which posits that the right hemisphere is specialized for perceiving negative emotions, while the left is specialized for positive ones, regardless of modality (Silberman & Weingartner, 1986; Borod et al., 1998). Noteworthy, a large sample study investigated sixty brain damaged patients to explore the hemisphere differences in emotion perception, with findings supporting the right hemisphere hypothesis of emotion perception across emotion category, valence and modality (Kucharska-Pietura et al., 2003).

To summarize, while some lesion evidence suggests a certain degree of specificity, other findings highlight the highly distributed nature of emotion processing, with multiple pathways allowing for maximal adaptation and resilience in case of brain injury. For future researches it is fundamental to integrate the diverse set of evidence from multiple methodologies (e.g., human lesion studies, intracranial recording/stimulation, functional neuroimaging), in order to develop more plausible theories of how emotion are represented in the brain.

Over the past decade, neuroimaging techniques have been fundamental to explore how the human emotion system is encoded in the brain. The quest for correlates of each component of the emotional experience has been a prolific line of research, and it is currently ongoing and thriving thanks to the development of more sophisticated analysis techniques.

Particularly, through the use of the functional magnetic resonance imaging (fMRI), it has been possible to explore the hemodynamic activity of brain areas in response to the perception or elicitation of an affective state on a moment-to-moment base. However, different results have been obtained from this method, which have been then interpreted according to the specific emotion theory embraced by researchers. As a result, a vigorous discussion is currently taking place on the organization of the emotion system (Hamann, 2012; Lindquist et al., 2012; Saarimäki et al., 2015; Barrett, 2017; Adolphs, 2017).

In the next section, results obtained from fMRI studies on affective states will be examined and considered in the context of each emotion theory.

2.1.3.1 Discrete emotions in the brain

According to discrete theories of emotions, each affective state has a biological innate root which has a specific correspondent in the brain, so it is possible to find patterns of response associated to basic emotions such as sadness, happiness and fear (Ekman, 1999; Keltner and Ekman, 2000).

In this sense, one of the most renowned examples is constituted by the relationship between the amygdala and the state of fear.

Indeed, human lesion studies provided important evidence supporting the idea that the amygdala is responsible for the perception and recognition of frightful events (Adolphs et al., 2005; Feinstein et al., 2011, 2013; Klumpers et al., 2014). Furthermore, a vast number of neuroimaging researches has reliably found a significant association between increased level of hemodynamic activity of the amygdala and fearful stimuli both in healthy (LaBar et al., 1998; Morris et al., 1999; Phelps et al., 2001; Whalen et al., 2001) and pathological conditions (Gentili et al., 2019. However, the view of the amygdala being the center of fear has also been challenged, with evidence indicating that this area is actually engaged across different aversive and appetitive objects and emotional percepts (Sander et al., 2003; Fusar-Poli et al., 2009; Sabatinelli et al., 2011; Namburi et al., 2016). Moreover, two meta-analyses failed in finding a significant association between activity of the amygdala and the emotion of fear, with one of them reporting it only in 60% of studies examined (Phan et al., 2002) and another one in less than 40% (Murphy et al., 2003).

Nevertheless, a wide number of neuroimaging studies have been conducted to explore the correlates of discrete emotions. So far, these studies have been mainly employing the stimuli described in the previous section, or similar ones. Importantly, a significant share of fMRI investigations does support the idea that at least basic emotions are associated to distinct neural bases, having distinguishable patterns of hemodynamic activity, as measured by this technique (Kragel and LaBar, 2015, 2016; Saarimäki et al., 2015, 2018; Nummenmaa and Saarimäki, 2017). For example, somatomor regions including the premotor cortex, caudate nucleus and posterior insula seem to represent emotions that have a strong behavioral and motivational component (e.g., fear, surprise; Nummenmaa et al., 2008, 2012). Instead, the activity of more anterior regions, such as the anterior prefrontal cortex, seems to be modulated by positive affective states (e.g., happiness; Vytal and Hamann, 2010; Saarimäki et al., 2018).

Of note, in order to assess the extent to which evidence coming from neuroimaging studies are along the lines of basic emotions theory, it has also been employed a meta-analytic approach. This method is of particular value since it can quantitatively assess results coming from a wide sample of studies and provide statistically robust evidence (Fox et al., 1998).

In this regard, results of different meta-analyses supported the idea that discrete basic emotions are characterized by distinct neural representation (Murphy et al., 2003; Vytal and Hamann, 2010). To summarize these evidences, happiness seems to be associated to hemodynamic activity of the rostral anterior cingulate cortex (ACC) and right superior temporal gyrus, while sadness is represented in the middle frontal gyrus (MFG) and subgenual ACC. On the other hand, anger consistently activated the inferior frontal (IFG) and parahippocampal gyri, fear the amygdala and the insula and disgust the IFG and anterior insula. Of note, these results point towards the existence of patterns of brain activity associated to basic affective states and not to a specific region-emotion correspondence. According to basic emotions theory, this evidence is a strong indication of the biological root of discrete affective states, while for constructionist theories, these patterns are actually aspecific and do not constitute a valid fingerprint of basic emotions in the brain (Touroutoglou et al., 2015; Clark-Polner et al., 2017). Furthermore, the pattern of brain regions associated to distinct affective state does not seem to be highly consistent, with different studies reporting dissimilar areas encoding basic emotions (Barrett and Wager, 2006).

Taken together, there is a large part of neuroimaging studies providing results in line with the idea that basic emotions are associated to patterns of brain activity, which are also specific and discernable. However, it should be noted that these researches have explored the construct of emotion through different stimuli (e.g., pictures, short videos) and various types of subjective responses (e.g., recognition, perception). Contrasting point of views have also arisen, and in the next section other theories on the representation of emotions in the brain will be examined.

2.1.3.2 Emotion dimensions in the brain

Dimensional models of emotions assume a different perspective on the neural correlates of affective states compared to discrete theories. For example, the circumplex model of affect (Russell, 1980) postulates that different neural pathways are associated to the cardinal dimensions used to describe emotional experiences (e.g., valence and arousal; Posner et al., 2005), and not to discrete affective states. For this reason, neuroimaging studies trying to support this idea have been employing emotional stimuli rated along the dimensions of interest, usually two. However, the study of cardinal dimensions related to affective experiences poses a substantial challenge, since the dimensions act together during the unfolding of the emotional experience, but it is problematic to assess their individual contribution at the same time. Therefore, a large set of investigations has focused on the two poles of only one dimension, as for example positive against negative valence or high respect to low arousal (Williams et al., 2001; Maddock et al., 2003; Phan et al., 2003; Posner et al., 2009). Across these studies, the hemodynamic activity of the amygdala, the prefrontal cortex (PFC) and the dorsal ACC has been found associated to distinct levels of arousal. On the other hand, the orbitofrontal cortex (OFC), superior temporal sulcus (STS), IFG and precentral sulcus, among other regions, have been reported to be involved in the evaluation of unpleasant stimuli, while the precuneus, posterior cingulate and middle frontal gyrus (MFG) of pleasant ones (Maddock et al., 2003).

However, subsequent investigations have suggested a more complex representation, taking into account the individual contribution of valence and arousal, as well as their interaction. To do so, these neuroimaging studies have been trying to overcome the limitation of considering valence and arousal individually and asked subjects to give an online rating of their affective state across the two (Gerber et al., 2008; Colibazzi et al., 2010). Following this method, researchers have found similar evidence supporting the circumplex model of affect, with the activity of distinct brain regions being associated to subjective ratings of valence and arousal. In particular, the hemodynamic activity of the dorsolateral prefrontal cortex (DLPFC), frontal pole, rostrodorsal ACC, supplementary motor area and inferior parietal seems to be related to the processing of valenced stimuli (Gerber et al., 2008; Colibazzi et al., 2010). On the other hand, emotions rated with a high level of arousal were found to be encoded in subcortical regions as the thalamus, globus pallidus and amygdala, as well as in cortical areas as the medial PFC (Gerber et al., 2008; Colibazzi et al., 2010). Of note, it has also been found that the hemodynamic activity of the medial OFC and putamen was modulated by the interaction between valence and arousal, highlighting the tight relationship between cardinal dimensions.

Importantly, recent neuroimaging studies have employed more ecological stimuli, asking subjects to watch short movies during an fMRI acquisition and to report their emotional state after them, rated along the two dimensions of valence and arousal (Nummenmaa et al., 2012; Kim et al., 2016, 2017). By doing this, it has been possible for researchers to not only extend previous results on the representation of affective states, but also to better characterize brain responses during emotional experiences. Indeed, these studies have confirmed the role of some brain regions in various emotional states elicited by a naturalistic condition (Kim et al., 2016, 2017; Nummenmaa et al., 2012). Furthermore, these studies also unveiled mechanisms related to the sharing of other individuals' emotional states. Indeed, the two cardinal dimensions of valence and arousal seem to be supported by different but partially overlapping networks, and influence in a distinctive manner the synchronization between brain regions of individuals (Nummenmaa et al., 2012). This evidence is of particular value, since it sustains the existence of a dimensional affect framework in the brain and it also shed light on the distinct roles of cardinal dimensions in human interactions.

In conclusion, neuroimaging studies have provided biological plausibility of dimensional models of emotions, demonstrating also the utility to consider this approach in the affective neuroscience domain.

2.1.3.3 Constructivist view on the mapping of emotions in the brain

Constructivist theories of emotions postulate that, as emotions emerge from the interaction of several functions, it is not possible to identify single brain regions associated to specific affective states. On the contrary, it is conceivable to have in the human brain a set of core regions consistently engaged during different kinds of emotional experiences (Oosterwijk et al., 2012; Touroutoglou et al., 2015). Indeed, the conceptual act theory, currently the main constructionist model, postulates that each emotional percept is constructed by the brain from the combination of basic psychological and neural components (Barrett, 2017). Researchers adopting this view strongly criticize the results obtained by neuroimaging studies employing discrete emotions, asserting that they failed to find reliable and consistent features able to discriminate these categories (Barrett, 2006).

On the contrary, constructivist model suggest that the labels assigned to affective states should not be taken in consideration when trying to unveil the structure of emotions in the brain. To do so, neuroimaging studies conducted adopting this approach have been employing a meta-analytic data-driven method (Kober et al., 2008; Lindquist et al., 2012). This method has identified both groups of brain regions functionally active during discrete emotion states, as well as distinct set of areas associated to the various psychological and cognitive processes constructing the emotional experience. In particular, six distributed groups of brain regions have been identified to underlie emotion processing (Kober et al., 2008):

1. Core limbic	Core affect
2. Lateral paralimbic	
3. Medial posterior	
4. Medial prefrontal	Conceptualization
5. Cognitive/motor	
6. Occipital/visual	

According to the interpretation of the authors, the construction process of the affective state begins with the activation of the core affect group, responsible of the emotional reaction, then followed by the conceptualization one. This group, by interacting with a set of brain regions holding attentive, executive and language functions, helps the individual to interpret and give meaning to the emotional percept. Lastly, during the emotional experience, there is also the functional activation of occipital brain regions, dedicated to the visual processing of the stimuli (Kober et al., 2008).

To summarize, the constructivist point of view interpreted the obtained findings as supporting their theory, with different brain regions constituting distinct functional groups in the brain. Of note, components of these groups can also be shared and intervening in more than one aspect of emotion processing, and the combined activity of these networks is what aid in the construction of the emotional experience.

A subsequent meta-analysis aimed to expand these findings and directly compare the discrete emotion approach to the constructivist one (Lindquist et al., 2012). To do so, the authors first considered a set of brain regions associated to the experience of basic affective states and explored whether any of those categories was specifically related to their fluctuations in hemodynamic activity. For example, they found that the amygdala was more activated during fear percepts, but this activation was not functionally specific, as it was shared with other emotion categories (e.g., disgust). The same evidence was valid also for other basic emotions: hemodynamic activity of the insula was not modulated by the state of disgust only, the ACC by sadness, and so on. Indeed, the authors interpreted these findings according to the idea that "emotion categories are not natural kinds that are respected by the brain" (Lindquist et al., 2012). These regions are involved in a variety of cognitive and perceptual tasks, and there is actually no distinction between emotional, linguistic, visual, etc., processes in the brain (Barrett, 2017).

Furthermore, a more recent study, has investigated the conceptual act theory and the basic emotions one employing resting state functional connectivity (Tourotoglou et al., 2015). This method allows the observation of the intrinsic organization of the brain, capturing the dialogue between different brain regions at rest, while individuals are not performing any task (Greicius et al., 2003; Fox and Raichle, 2007). The authors aimed to investigate spontaneous brain activity and its association with affective states, for different relevant reasons. Indeed, the finding of intrinsic brain networks representing discrete emotions would be a valuable evidence to support this kind of theories. Furthermore, these networks are reliably found across cultures and even species, which would again support the basic emotions framework. However, the authors did not find brain networks at rest corresponding to specific affective states, consistently with the constructivist point of view (Tourotoglou et al., 2015). Indeed, using the emotion map of brain regions, general-domain networks arose (e.g., default mode network, salience network, executive control network), supporting the idea that different emotions are built from the interaction of broad systems within the brain (Barrett, 2012; Lindquist and Barrett, 2012; Barrett and Satpute, 2013).

In conclusion, there is a portion of neuroimaging findings which seems to support a constructionist organization of the human brain in the emotion domain and claim the need of revising the affective framework to take into account this model of functioning (Lindquist and Barrett, 2012; Barrett and Satpute, 2013).

In light of all this, it is clear how strong and still ongoing the debate on the organization of the human emotion system is. Indeed, both an alternative perspective and the use of new techniques and more sophisticated methodological analyses are much needed to shed new light on the representation of emotions in the brain.

2.2 Empathy

2.2.1 Definition and perspectives

It is since centuries that philosophers and psychologists have been investigating the fundamental human ability to “feel into”. Indeed, in our daily life it is possible to notice plenty of times in which we find ourselves experiencing some feelings of concern, compassion or different affective states, that are however not related to ourselves. For example, it might happen that a friend tells us that she just lost a dear one and we resonate so much with her that we end up crying, or we see a stranger falling and stuck his hand on a nail and we might perceive shivers running down our spine and feel a very bothering reaction. All these events happen because we are able to perceive what other individuals feel and, most importantly, we tend to allineate our inner state with theirs. According to some authors, the socio-affective abilities responsible for these sensations can all fit under the umbrella term of *empathy*, which includes affective resonance, emotional contagion, sympathy, personal distress, empathic concern, compassion, perspective taking and affective and cognitive empathy. Indeed, it has been reasoned that we can use the term *empathy* to describe “all the processes that emerge from the fact that observers understand others’ states by activating their own personal, neural and mental representations of that state” (de Waal and Preston, 2017; Preston and de Waal, 2002; de Waal, 2008).

However, other descriptions of empathy and emotional contagion sees them as distinct abilities (Decety and Ickes, 2009). In the next section, the two constructs of emotional contagion and empathy in the strict sense will be examined following the latter definition.

2.2.1.1 Emotional contagion

Emotional contagion is defined as: “the tendency to automatically mimic and synchronize expressions, vocalizations, postures and movements with those of another person’s and, consequently, to converge emotionally” (Hatfield et al., 1994).

This process is of fundamental importance in the socio-affective domain of human life. Indeed, the ability to feel ourselves "into" another individual emotions is crucial to build our relationships, it fosters behavioral synchrony and constitutes a pillar for more complex socio-affective abilities. Furthermore, scholars coming from a variety of scientific fields have highlighted the general relevance of emotional contagion to better understand human behavior in cooperative contexts.

Based on its definition, it is clear that emotional contagion has a strong automatic, primitive and subtle component, as often we do not even realize that we are actually mimicking actions and gestures of another person. Instead, we tend to notice only the last part of this process, finding ourselves experiencing the same emotional percept of that individual, without being able to identify the origin of this affective state.

To better characterize emotional contagion, a series of proposition underlying this process have been delineated according to the original model of Hatfield, Cacioppo and Rapson (1992):

1. *Mimicry*

As early as the 16th century, scientists have been observed that people automatically and continuously mimic the emotional expressions of other individuals. In particular, the facial activity of subjects experiencing a certain affective state seems to be mirrored by the observer (Vaughan and Lanzetta, 1980; Dimberg, 1982). Indeed, studies using electromyography, have found that the activity of the muscles involved in the expression of an emotion (e.g., *corrugator supercilii* for anger), was present even when simply observing people feeling that same affective state (Lundqvist and Dimberg, 1995; Lundqvist, 1995; Wild et al., 2003). Furthermore, individuals are so able to track the most subtle moment by moment changes in another person, that this mirroring happens in an almost instantaneous and not always apparent manner. Moreover, together with facial expressions, it seems that also aspects related to interindividual speech undergo mimicry processes, including prosody and pitch (Cappella and Planalp, 1981), and postures and movements (Bernieri, et al., 1991; Hatfield et al., 1994). Indeed, it is easy to think examples of our daily life in which we are subjected to this form of contagion (e.g.,

during a conversation, we might notice that we have crossed our arms after the other person did so).
In sum, it seems that there is a natural tendency in individuals to mimic and synchronize faces, voices, postures, and movements with those around them. They also do this in a very rapid and automatic manner and by integrating at the same time emotional signals coming from various resources.

2. *Feedback*

The components just examined are acting altogether to construct the final emotional experience. Therefore, the feedback coming from each of them is also playing an important role in the process of emotional contagion. In particular, there have been studies showing that when individuals are asked to reproduce a facial expression connoted by a specific emotion, they end up feeling that same affective state (Laird and Bresler, 1992). Furthermore, also patterns of intonation, vocal quality, and rhythm associated to an affective experience, are able to influence the emotions felt by an individual (Hatfield et al., 2011).
Taking into account this, there is relevant evidence that people tend to feel emotions in accordance to the facial, vocal, and postural expressions adopted. However, it is still not clear the role of this type of feedback in the emotional contagion process and further investigations are needed in this regard.

3. *Contagion*

In light of all this, it is believed by researchers that, as a consequence of mimicry and feedback, people tend to capture emotions of other individuals. Indeed, these two processes act together to "change" the affective state of the observer and shifting him/her toward the emotional experience of the other, resulting in the final percept of emotional contagion.

Emotional contagion has been considered so fundamental and lying at the basis of more complex abilities since these different aspects of the process are actually what makes us able to understand the intentions and feelings of other individuals. Indeed, these processes happen on a moment by moment basis, even when we are not explicitly paying attention to them and allow us to properly interact with others in the social world.

However, it is important to note that emotional contagion can also lead to an aversive type of response in the observer, which is translated in a feeling of personal distress. In fact, when we are experiencing contagion, we tend to either act in order to modify the experience and relieve the other person from their suffering, or to avoid the percept and silence our own feeling. Personal distress is guided exactly by the latter process and motivate individuals to take distance from the experience and response in a self-centered manner (Eisenberg and Fabes, 1990).

2.2.1.2 Empathy

Regarding empathy, the characterization of this construct is complex and multifaceted. Indeed, the definition of empathy is undergoing a similar process of the emotions one, with several scholars giving various descriptions of this socio-affective ability. Furthermore, the history of scientific research in the empathy domain is not very long. Even if we can find the first definition of this ability in the ancient Greek with the term *empatheia*, which literally means "in feeling", the scientific investigation of this construct did not take place until recent years, with the work of developmental and social psychologists (e.g., Batson and Shaw, 1991; Eisenberg, 2000; Hoffman, 2008; Batson, 2009). A large part of recent research in the social psychology domain has aimed at investigating the mechanisms lying at the basis of the human capacity of *einfuhlung* (i.e., feel into someone else's shoes).

However, due to the complexity of this socio-affective ability, it has been, and still is, particularly challenging to study empathizing processes in a laboratory setting, using controlled and limited conditions outside the regular context in which it is experienced. In fact, it is only recently that neuroimaging studies have been appeared using *ad hoc* stimuli to investigate this multidimensional phenomenon.

Among the various descriptions of empathy, this construct has been defined for example as the "capacity to understand others and experience their feelings in relation to oneself" (Decety and Jackson, 2004), the "drive to identify another person's emotions and thoughts and to respond to these with an appropriate emotion" (Davis, 2018), the "capacity to identify one's own feelings and needs with those of another person" (MacLean, 1967), and as a “leap of imagination into someone else’s headspace” (Baron-Cohen, 2005). Indeed, empathy does not have a universally accepted definition, and which other cognitive and emotional components act in concert during the unfolding of this socio-affective phenomenon it is still a matter of debate.

Nevertheless, to summarize these descriptions, we can say that empathizing occurs when (De Vignemont and Singer, 2006) we are experiencing an affective reaction which is isomorphic to the one of a different individual. This state has to have been elicited by the observation or imagination of another person in that condition and we should be able to recognize it.

In particular, it is important to stress that empathic processes can take place both when an individual perceives or imagines someone else emotional experience, and this triggers a series of responses leading the observer to feel what the other is feeling (Singer and Lamm, 2009). Moreover, this description implicates a certain type of low-level processing at the basis of more complex components. Indeed, individuals have to first of all recognize that the other is a different person respect to the self (i.e., self-other distinction) and that he/she is the real origin of the emotions we are perceiving when we are empathizing. After this, it is necessary to understand the affective state that the other individual is perceiving, and this process has been identified both as *affective ToM* and as *cognitive empathy* (Eslinger, 1998; Shamay-Tsoory et al., 2009). This latter ability highlights the presence of cognitive capacities in the construct of empathy, which are supposed to aid the individual in recognizing the perspective of a different person. After all this, we should then “resonate” with the same emotional experience that the other is feeling, which has been termed as *affective empathy* (Davis, 2018; Shamay-Tsoory et al., 2009) and implicates some form of sharing between the self and the other.

A recent model, proposed by a renown social neuroscientist, postulates that mentalizing abilities and emotion understanding represent the top-down processes through which human beings are able to produce an empathic response (Decety, 2011). This response, however, requires that such abilities interact with bottom-up processing of affective sharing, elicited by the automatic discrimination of stimuli as pleasant or unpleasant or as threatening or nurturing. The output of these two pathways is then self-regulated based on the perceiver’s motivations and intentions. Overall, this process shapes the quality and the extent of the empathic experience, thus supporting flexible and appropriate behavioral responses in the social domain.

Based on the final experience of empathizing, it is actually challenging to disentangle other similar constructs which may lead us to perceive the same type of resonance. In particular, it is important to make a distinction between empathy and emotional contagion, sympathy, compassion and empathic concern.

Indeed, one characteristic necessary to empathize with a different person, is to recognize that the source of the affective experience resides outside ourselves, while to perceive emotional contagion it is not fundamental. This is also testified by the name given to this process, as a "contagion" is defined as the spreading of a state among a certain number of people, exactly as a disease. Therefore, this spreading will make individuals living a similar experience, which will be then transmitted from person to person and the original individual responsible for the percept will be lost in some sense. One example that has been very common in the history of the human kind and that even had dramatic consequences, regards the spreading of fear, particularly in situations of small spaces occupied by a wide number of individuals (e.g., concerts, squares), where this state is disseminated among all of them, and which may even degenerate in a *mass psychogenic illness*.

Furthermore, during sympathy and compassion, there is usually a strong feeling of concern which is what lead individuals to act and reduce the other person suffering (Klimecki and Singer, 2012). This feeling, however, does not imply also a sharing of the affective state of the other individual, but will probably guide prosocial behavior (i.e., every act that benefits others or the society; Batson et al., 2007), which may not be present in other socio-affective phenomenon as emotional contagion and personal distress. Sympathy and compassion can also be distinguished from empathic concern, which is instead an "emotional and motivational state characterized by the desire to help and promote others' welfare" (Bernhardt and Singer, 2012). Also in this case, the sharing of feelings with a different individual is not necessary in order to promote the wellbeing of this person and therefore, is distinct from empathizing.

As already mentioned above, several models have suggested that empathy is actually composed by cognitive and affective subcomponents (Davis, 1983; Eslinger, 1998; Walter, 2012). In particular, the construct of empathy in the strict sense has been defined as affective or emotional empathy, which can be described as the experience of an affective state elicited by the real or imagined emotional percept of a different individual (Shamay-Tsoory et al., 2009). This phenomenon would represent empathizing, and can be then subdivided into other components, having its origin in emotional contagion, personal distress and empathic concern. On the other hand, the other subcomponent of this construct has been termed as cognitive empathy, which instead refers to "the ability to comprehend others' affective states without necessarily indicating that the empathizer is in an affective state himself" (Walter et al., 2012).

It is important to note that all the constructs under the umbrella term of empathy, are very relevant to understand humans social life, as they are at the basis of fundamental evolutionary factors, as prosociality and altruism. Indeed, emotional contagion, empathy and the other similar socio-affective abilities, are not something that was present at the origins of the human kind, but were instead shaped during years and years of evolutionary history. These phenomena are partly present also in animals (Preston and de Waal, 2002), and have been fundamental for survival, for example by helping individuals in gathering and hunting for food and detecting predators (Plutchik, 1987). Importantly, without social interactions and emotional bonds with others, empathy would not have developed (Decety and Jackson, 2004).

In conclusion, empathizing regards having an appropriate emotional reaction, perceived as an inner experience but triggered by another person affective state. This is much more than simply understand what a different individual is thinking or feeling, and indeed, even psychopaths are able to do so. Instead, what they are missing is the following cascade of events. Indeed, when we observe another person in pain, we don't feel happy or detached, but we are at the mercy of a series of events which makes us perceive concern, compassion, and pain ourselves. Thus, it is easy to imagine the importance of empathy in human relationships.

To summarize, emotional contagion and empathy represent vital phenomena for our everyday communication and survival in the social environment. Therefore, the neuroscientific investigation of these abilities is of fundamental importance to better understand the roots of human behavior.

2.2.2 The assessment of empathic abilities

In light of the multidimensional nature of empathy and the complexity of this construct, it has been and still is, particularly challenging for research to objectively measure and assess empathic processes. Some authors considered empathy as a skill or a set of skills, which therefore vary among individuals, exactly as other cognitive abilities (Chakrabarti et al., 2006). Therefore, it is particularly important to take into account the individual differences in empathizing and different measures have to be developed in order to do so. However, it is also necessary to mainly rely on self-assessment performed by the individual in evaluating his own reactions and behavior, and this might have some limitations in the objective research of empathy. The main issue in using self-assessment and self-reports is actually common to the entire field of human research, where we have to assume that people respond according to how they actually are and not how they would like to appear to fit in social norms.

Regarding empathy, a various number of scales and questionnaires have been constructed to try to assess the complexity of this construct.

One of the first has been the Questionnaire Measure of Emotional Empathy (Mehrabian and Epstein, 1972), which measure the level of easiness of an individual to have an emotional reaction in front of a certain event. However, this measure has been revised, and a newer version was developed by the authors, the Balanced Emotional Empathy Scale (Mehrabian, 2000). This questionnaire, instead, is supposed to assess specifically only affective empathy, with items like "I cannot easily empathize with the hopes and aspirations of strangers/I easily get carried away by the lyrics of a love song". The main issue with this type of questionnaire is whether it is truly assessing only empathic abilities.

Currently, the main instrument used to evaluate empathizing skills in healthy and pathological conditions, is the Interpersonal Reactivity Scale (IRI; Davis, 1980). This is a self-administered questionnaire, which in case of an individual affected by a disorder, can be completed by the caregiver of that person instead. The test is divided in four subscales: perspective taking, empathic concern, fantasy and personal distress. Each of this component is supposed to tackle different aspects of the general construct of empathy. Indeed, perspective taking items detect “the tendency to spontaneously adopt the psychological point of view of others”, with statements like “I sometimes try to understand my friends better by imagining how things look from their perspective” (Davis, 1980). The subscale of fantasy, instead, measure the “respondents’ tendencies to transpose themselves imaginatively into the feelings and actions of fictitious characters in books, movies, and plays”, as for example “When I am reading an interesting story or novel I imagine how I would feel if the events in the story were happening to me” (Davis, 1980). The affective and emotional components of empathy is revealed by the empathic concern items, which “assesses other-oriented feelings of sympathy and concern for unfortunate others”, like “I often have tender, concerned feelings for people less fortunate than me”, and by the personal distress ones (Davis, 1980). The latter measure “self-oriented feelings of personal anxiety and unease in tense interpersonal settings”, as “Being in a tense emotional situation scares me” (Davis, 1980). This instrument, however, has also been criticized on its capacity to again unveil empathy in itself (Baron-Cohen and Wheelwright, 2004).

Moreover, the Empathy Quotient (Baron-Cohen and Wheelwright, 2004) is a scale that has been originally designed to be useful in the clinical setting and to be particularly sensitive to the lack of empathic abilities in psychopathologies. The target population of this test was individuals affected by autistic spectrum disorders, but it has also been used in normal conditions (e.g., Preti et al., 2011).

More recently, neuroimaging studies have been employing dynamic or static stimuli with videos or pictures of people feeling pain (Singer et al., 2004; Benuzzi et al., 2008), disgust (Benuzzi et al., 2008), anger (de Greck et al., 2012), interacting with each other and feeling social exclusion (Masten et al., 2011) or embarrassment (Krach et al., 2011). All these stimuli have been successfully used to elicit empathizing processes in observing subjects and to record patterns of brain activity associated to them.

For example, the Socio-affective Video Task (Klimecki et al., 2013) is constituted by silent videos showing people in situation evoking high distress, or while they are performing daily life activities.

A very recent paradigm, the EmpaToM, has been developed by Kanske and colleagues (2015). This task allows the simultaneous investigation of empathic responding and ToM processes, which are particularly related, as it has been described in the previous section. The EmpaToM is of particular value since it "implements an orthogonal manipulation of empathy and ToM during an ongoing realistically complex and demanding situation requiring social understanding of others" (Kanske et al., 2015). Indeed, the task is composed by different videos depicting actors reporting an autobiographic narrative which can be negative, as experiences of loss and threat, or neutral. Importantly, the EmpaToM has been found to be able to assess both neural and behavioral markers of empathy, compassion and ToM ability at the same time.

To conclude, also in the empathic and emotional contagion domain it is now occurring a shift toward more naturalistic investigations of these abilities, in order to properly capture their multidimensional and complex nature. However, further studies are needed to explore empathizing processes in an ecological manner.

2.2.3 Empathy in the brain

The use of neuroimaging techniques in vivo has led researchers to look for the existence of brain regions involved in various social cognition abilities, including empathizing processes. Indeed, it has been observed a consistent recruitment of certain brain areas in tasks requiring observing, predicting and responding to social stimuli, and therefore it has been postulated the existence of a "social brain" (Frith and Frith, 2007). For example, regions as the amygdala and the ACC have been identified as two fundamental hubs in the network of brain areas dedicated to empathy and emotional contagion. Indeed, an increase in the hemodynamic activity of these regions has been observed in a wide range of empathic responses (Bernhardt and Singer, 2012).

However, due to the complexity of the two constructs of emotional contagion and empathy (Decety and Ickes, 2009), different models have been proposed to describe their representation in the human brain.

In a seminal article, it has been postulated that "the observation and imagination of others in a given emotional state automatically activates a corresponding representation in the observer, along with its associated autonomic and somatic responses" (Preston and de Waal, 2002). The authors developed a specific framework for empathic abilities, which they named PAM: perception-action mechanism. According to this model, attending to another individual distress will activate the observer's own representations of that situation, retrieving feelings, memories and associations. Indeed, the PAM is a particularly broad model, since it includes in the empathizing responses the unique past experience of each individual. For example, a woman feeling pain for menstrual cramps will elicit a stronger empathic reaction in another female, which will understand better the "kind" of pain she is experiencing. This is obviously true for males as well, as observing a man getting hit below the waist will elicit a different type of emotional response in other males, compared to a woman observing the same scene. Therefore, it will be possible to observe in the brain a shared and distributed neural representation for the self and the other, which will be different among individuals and specific for each one's own past experience.

On the other hand, almost concurrently to these investigations of the nervous system aspects related to empathizing, there has been an important discovery in the scientific community: mirror neurons. This class of neurons found in the monkey for the first time, has been observed to respond both when the monkey was executing an action and when it was simply observing another one performing the same action (Rizzolatti et al., 2001; Gallese et al., 2004). This discovery opened an entire field of research investigating a whole gamut of processes, including motor, social, and somatosensory functions, and used as a model to explain pathological conditions (e.g., autism; Williams et al., 2001). As it can be easily imagined, also in the socio-affective domain mirror neurons have been extensively investigated trying to disentangle their role in our emotional percepts (for a review see Gallese, 2001 and Iacoboni, 2009). Indeed, several researchers embrace a "mirror mechanism" as the basis of socio-affective abilities, with the experience of empathy originating through a simulation of the mental states of other people, made possible thanks to the mirror neurons system (Rizzolatti and Caruana, 2017; Iacoboni, 2009). According to this view, it is exclusively the mirror system to supports the representation and imitation of emotional expressions, empathy and empathic pain.

Despite the differences in these two models, a gamut of neuroimaging studies has been conducted to explore the association between empathic responses and brain hemodynamic activity elicited during various conditions.

In particular, one of the main states that has been investigated is indeed pain and the empathic responses associated to it. In fact, it is well known that the pain experience is characterized by different factors:

1. It is generally aversive;
2. It motivates behavioral responses to reduce the harmful stimulation (Price, 2000);
3. Can induce forms of warning communication to conspecifics (Craig, 2004).
4. When humans simply observe other individuals suffering, but are in no way involved in this event, they enact helping behavior and later report it as an aching experience for themselves (Hein et al., 2010).

A seminal study investigated a group of females and their respective partners during a painful stimulation in the fMRI scanner (Singer et al., 2004). The authors found an increase in the hemodynamic activity of the anterior insula (AI), dorsal ACC (dACC), brain stem and cerebellum both when the female subjects received an electric shock and when they observed the hand of their partner receiving the same stimuli. Importantly, this group of brain regions has been found involved in firsthand pain experiences as well, along with other areas as the periacqueductal gray and primary and secondary somatosensory cortices (S1, S2). Indeed, a meta-analysis performed on 32 fMRI studies reported a consistent network of brain regions involved in empathic pain processes, including the AI, dACC and somatosensory regions (Lamm et al., 2011).

Furthermore, several subsequent studies found the same results also when simply presenting to observers facial expressions of others in pain (Lamm et al., 2007) or body parts subjected to painful stimulation (Jackson et al., 2006; Lamm et al., 2007). A recent neuroimaging study employed multivoxel pattern analysis to unveil a distributed cortical network associated to aversive events, regardless of the subject affected (Corradi-Dell'Acqua et al., 2011).

Altogether, this evidence is in accordance with the seminal work of Preston and de Waal (2002), in which they postulated that emotional contagion and empathic processing is based on the corresponding representation of the state of an individual in the observer.

Several additional neuroimaging studies have been conducted exploring other affective states respect to pain, including anger, disgust, anxiety, and social exclusion. In particular, for the experience of disgust, different researches have used a direct stimulation with disgusting odorants and bitter liquids and found an involvement of the AI and ACC both when subjects inhaled/tasted them and when they simply observed the facial expression of other people doing so (Wicker et al., 2003; Jabbi et al., 2008). Interestingly, a recent meta-analysis highlighted the association between hemodynamic activity of these two regions in a variety of empathic domains, such as subjects observing fear, happiness, disgust, and anxiety expressed by other individuals (Fan et al., 2011). In this regard, a recent fMRI study investigated the neural correlates of empathizing to individuals experiencing happy, sad, and anxious events. The results highlighted the existence of a set of brain regions supporting empathic processes, including the right temporo-parietal junction (TPJ; Morelli and Lieberman, 2013), an area found to be involved in different socio-affective abilities (Saxe and Kanwisher, 2003; Skerry and Saxe, 2015). Importantly, this finding would suggest that the TPJ, already known to be fundamental for mentalizing and emotion processing, may be critical also for generating empathic responses. This would highlight the tight relationship between the different socio-affective abilities, and how these might be represented in the brain in a similar manner.

Of note, as empathy can be considered as a set of skills, which are present in variable levels among individuals, some studies have investigated how these differences affect empathic responses. Indeed, using the IRI, it has been observed a significant modulation of the empathic response according to the traits measured by this questionnaire (Singer et al., 2004; Jabbi et al., 2007). Moreover, a various number of neuroimaging studies has been conducted exploring whether empathizing might be modulated even by the information available to the observer. Indeed, both implicit and explicit information regarding the reasons and the context behind another person suffering pain, seems to influence empathic responses (Singer et al., 2006; Lamm et al., 2007; Hein et al., 2010).

Lastly, it is important to note that very recently it has been suggested an integrative approach to consider altogether the different facets of empathy and not in isolation (Lamm et al., 2017).

In conclusion, neuroimaging studies have provided valuable evidence in favor of the existence of shared representations in the brain involved both in firsthand and vicarious experiences of sensations and emotions. Furthermore, it seems that these empathic responses in the brain can be dynamically modulated by the context or the person characteristics. For example, these processes may engage other neural networks if other socio-affective abilities are more needed, according to the information available to the observer. In this regard, how socio-affective abilities interact during naturalistic situations and how this dialogue is encoded in the brain is still unclear.

2.3 Theory of mind

2.3.1 Definition and perspectives

To properly act in the social world, it is necessary to understand what a different person is thinking or feeling. Indeed, this is an essential ability to interpret and give meaning to the behavior of individuals and produce appropriate responses.

Premack and Woodruff in their seminal work "Does the chimpanzee have a theory of mind?", first delineated the existence of a process by which "an individual imputes mental states to himself and others": a *Theory of Mind* (*ToM*; Premack and Woodruff, 1978). According to the authors, to hold a ToM means that a person has successfully developed an internal model of what mental states are, allowing the attribution of these to others. Moreover, it is possible to distinguish a *first-order* ToM, which requires to understand the intentions or beliefs of an individual (e.g., "he thinks that"), and a *second-order* ToM, which instead is related to the inferral of someone's beliefs about the thoughts of a different person (e.g., "he thinks that she thinks"; Baron-Cohen et al., 1985; Baron-Cohen, 1989). It is commonly assumed that the second-order ToM requires a higher cognitive load and actually appears later in the development, whereas the first-order can be observed in children already at 3 years old (Hughes and Leekam, 2004).

The fundamental socio-affective ability of ToM has also been referred to as *mentalizing* or *mindreading* interchangeably, as the act *per se* might indeed be considered as "reading the mind" of different individuals. However, this has also created more confusion in the empirical research, with different scholars having various definition of the ToM construct in mind, leading to a high level of heterogeneity in findings (Schaafsma et al., 2015).

Furthermore, it has been suggested that mentalizing abilities can be decomposed in many different processes. In this regard, ToM is currently recognized as requiring the integration of several components including the skill of attributing both intentions (i.e., *cognitive ToM*) and emotions (i.e., *affective ToM*) to others (Brothers and Ring, 1992; Shamay-Tsoory et al., 2006; Völlm et al., 2006; Shamay-Tsoory and Aharon-Peretz, 2007). This distinction is

oftentimes also referred to as the *hot* (i.e., reasoning about feelings and emotions) and *cold* (i.e., reasoning about beliefs and intentions) aspects of ToM. Moreover, different models have been developed based on the psychological processes underlying the two different dimensions of ToM. For example, the ability of an individual to construct a scientific theory, engaged in understanding others, seems to be at the basis of the cold dimension of ToM (i.e., *theory-theory*; Gopnik and Wellman, 1994). On the other hand, the hot facet of ToM is more related to the skills needed to simulate the experience of different individuals (i.e., *simulation theory*; Gallese and Goldman, 1998; Gordon, 1986).
It is important to note that the different components of ToM, the recognition of emotion, the empathic concern, are all interrelated processes and are usually acting at the same time in an individual. To consider an example in this regard, when interacting with a different person and having a conversation, we read signals coming from their body, face, voice, we interpret what they are saying, we understand when they might be deceptive, we can feel compassion for them or a sense of internal sadness, and all of this happens in a matter of seconds, oftentimes without us realizing it. Indeed, all the socio-affective abilities examined in this *Introduction* chapter actually act in concert, involving also different cognitive and executive components, and the distinction among them during their unfolding is ambiguous and difficult to fully unravel.
However, it is also important to note that "to understand the intentions/desires of someone else does not guarantee emotion understanding, emotion understanding does not guarantee empathy and empathy does not guarantee sympathy" (Dvash and Shamay-Tsoory, 2014). In fact, a breakdown in one of these skills might involve also the others and impair the entire functioning of an individual.

In light of all this, further studies on mentalizing processes and their relationship with other socio-affective abilities are needed to shed new light on human's social behavior.

2.3.2 The assessment of mentalizing abilities

Over the past decades, a variety of investigations explored ToM and its different components, focusing on either cognitive or affective dimensions, both in healthy and pathological conditions. In particular, research in the mentalizing domain has focused on children, how they develop a ToM, and what happens when they fail to do so (e.g., autism spectrum disorder; Wellman, 1992; Baron-Cohen, 1997; Wellman et al., 2001; Frith, 2003). In light of this, a variety of verbal and non-verbal stimuli, investigating first or second order mentalizing processes have been employed in the empirical research.

For a fully developed ToM, it is necessary to hold representations of objects in the world, to understand that these might change in response to an event, and to project this skill when another person is involved. In particular, the importance of mentalizing abilities is evident when a person is misinformed, that is when he/she has a false belief. Indeed, ToM has been investigated mainly through the understanding of false-beliefs.

The false-belief task was originally developed for pre-schooled children and is now the most common used and known stimulus to explore the cognitive dimension of ToM, including first- and second-order processes.

The first and most famous false-belief condition, employed to assess ToM in children, is the Sally-Ann task (Wimmer and Perner, 1983; Baron-Cohen et al., 1985). This task has many different versions, but it always presents some constraints. The basic design of this task involves a protagonist that puts an object in a certain location A and leaves, and while he/she is outside another character moves the object to a different place B. The child is then asked to answer the question of where the protagonist will look for the object once he / she comes back. Usually, before the development of ToM, children answer with B, where the object actually is, while at 5-6 years old they will indicate location A, where the protagonist last saw it.

The false-belief task is very useful since it tests the ability to attribute mental states (e.g., intentions, beliefs, desires) to others and understand that these might be different from one's own. For these reasons, it has been employed in a variety of conditions and populations of interest (Apperly et al., 2004; Buttelmann et al., 2009; Lind and Bowler, 2009).

Instead, to explore the affective component of ToM it is necessary to require subjects to infer the emotions and feelings of other individuals. One of the most used tasks in this regard, is the Reading the Mind in the Eyes (Baron-Cohen et al, 1997). This stimulation is composed by a set of photographs representing only the eye region of different actors expressing an affective state. The individual is then asked to infer which is the emotional state that the actor is perceiving, based on the eye gaze of that person (Baron-Cohen et al, 2001). Also this task is highly valuable since it can be easily administered to healthy and pathological conditions, ranging from autism (Baron-Cohen et al., 2001; Rutherford et al., 2002) to psychopathy (Richell et al., 2003).

The majority of studies focused on the cognitive or the affective component of ToM, employing specifically developed tasks. Therefore, the “hot and cold” differentiation of ToM has been neglected by a vast part of the empirical research, limiting the generalizability of findings. Indeed, the use of two separate ToM tasks which differ in their demands on other cognitive abilities might have confounded the cognitive-affective dissociation of ToM. To overcome this limitation, new paradigms have been developed to assess at the same time the hot and cold dimensions of mentalizing processes.

The Yoni task (Shamay-Tsoory and Aharon-Peretz, 2007) requires subjects the ability to infer mental states based on verbal and eye gaze cues. The stimulus is composed by a cartoon outline of a face, which is called *Yoni*, and four colored pictures of objects belonging to a single category (e.g., fruits, chairs) or faces. Each of this image will appear in one corner of the screen, while the Yoni face will be at the center. After that, a short sentence will appear, and subjects are then required to indicate to which image Yoni is referring. To properly answer this question, the individual will need to use the facial expression, the eye gaze of Yoni character, and the facial expression of the face to which Yoni is referring. Furthermore, this task will involve either the evaluation of cognitive aspects (i.e., intentions, desires, beliefs), or of affective ones (i.e., emotions, feelings).

Moreover, as it has been described in the previous section, a very recent paradigm, the EmpaToM, has been developed by Kanske

and colleagues (2015). This task allows the simultaneous investigation of empathic responding and ToM processes. The EmpaToM is of particular value since it “implements an orthogonal manipulation of empathy and ToM during an ongoing realistically complex and demanding situation requiring social understanding of others” (Kanske et al., 2015). Indeed, the task is composed by different videos depicting actors reporting an autobiographic narrative which can be negative, as experiences of loss and threat, or neutral. Importantly, the EmpaToM has been found to be able to assess both neural and behavioral markers of empathy, compassion and mentalizing abilities at the same time.
In this regard, very recently, more naturalistic stimulations have started to be implemented in ToM research, involving short film developed for the general public (Richardson et al., 2018). This new approach might provide valuable insights on the ecological experience of emotional and affective states and allow a better understanding of the unfolding of these processes in daily life.

In conclusion, empirical research is still in need of comprehensive and ecological tasks for the evaluation of mentalizing abilities and their interaction with other socio-affective processes.

2.3.3 Theory of mind in the brain

Several functional neuroimaging studies have investigated the brain correlates of mentalizing abilities in healthy individuals (Van Overwalle and Baetens, 2009; Sebastian et al., 2011; Bodden et al., 2013), as well as in different pathological conditions including brain damage (Shamay-Tsoory and Aharon-Peretz, 2007; Shamay-Tsoory et al., 2009; Dal Monte et al., 2014), neurodegenerative diseases (Bodden et al., 2010; Elamin et al., 2012), and psychiatric disorders (Brüne, 2005; Wang et al., 2008).

Various models have been built to explain how ToM abilities are represented in the brain. Among the most relevant ones, we can find modularity and simulation theories. The first postulates the existence of a biologically rooted ToM network specifically dedicated to mentalizing abilities (Leslie et al., 2004; Baron-Cohen, 1998). According to this model, the innate network of ToM is composed by brain regions consistently activated whenever individuals engage in mental state reasoning. Moreover, this pattern of activation would be consistently found across the entire life of an individual.

On the other hand, according to simulation theories, it is fundamental the role of the mirror neuron system (MNS; Rizzolatti et al., 2001; Gallese et al., 2004, Goldman, 2006). This theory postulates the existence of a set of brain regions involved both in first-person and during the passive viewing of experiences. In this regard, various studies reported the activation of the premotor cortex and the inferior parietal lobule in the deduction of intentions from actions (Iacoboni et al., 2005) and in ToM processes (Van Overwalle and Baetens, 2009; Sebastian et al., 2011). The involvement of the MNS has been considered fundamental for the attribution of mental states, since it allows individuals to project into others' situations and better understand them (Gallese and Goldman, 1998; Goldman, 2006).

More specifically, a set of brain regions has been consistently and reliably found across different studies and conditions, playing a relevant role in the inference of mental states, including the right TPJ, STS, precuneus and middle prefrontal cortex (MPFC; for a review see Frith and Frith, 2012). Furthermore, a vast amount of neuroimaging studies has investigated how the processing of cognitive and affective ToM takes place, exploring brain hemodynamic activity changes in response to cold (Sommer et al., 2007; Sebastian et al., 2011; Bodden et al., 2013) and hot ToM components (Shamay-Tsoory and Aharon-Peretz, 2007; Sebastian et al., 2011; Bodden et al., 2013).

Different networks of brain areas have been associated to the two dimensions pertaining to mentalizing abilities. Regarding the affective component of ToM, regions as the vmPFC, OFC and pSTS/TPJ (Shamay-Tsoory and Aharon-Peretz, 2007; Corradi Dell'Acqua et al., 2014; Skerry and Saxe, 2015) seems to be involved in the attribution of emotions to others. On the other hand, it has been argued that the dorsal MPFC, DLPFC and TPJ are specifically involved in the cold dimension of mentalizing (Sommer et al., 2007; Schurz et al., 2014). Over these results, an ongoing debate took place, with different scholars trying to indicate the core regions necessary for ToM (e.g., TPJ for Saxe and Kanwisher, 2003; anterior paracingulate cortex for Gallagher and Frith, 2003).

However, despite the various brain regions that have been involved in this discussion, there is strong evidence for the general engagement of posterior temporal regions in ToM tasks (Saxe and Kanwisher, 2003; Saxe and Wexler, 2005; Van Overwalle and Baetens, 2009; Abu-Akel and Shamay-Tsoory, 2011; Corradi-Dell'Acqua et al., 2014; Skerry and Saxe, 2015). Indeed, these results suggest that a system of temporo-parietal brain structures which mediates perspective-taking and detection of social signals in general (e.g., gaze direction, biological motion; Pelphrey et al., 2004) is involved in ToM processes irrespectively of the type of information being treated. In this regard, the TPJ region seems to play a fundamental role in the attribution of both mental and affective states to others (Saxe and Kanwisher, 2003; Van Overwalle, 2009; Schurz et al., 2014; Skerry and Saxe, 2015), likely underpinning an attentional switch between different perspectives (Corbetta and Shulman, 2002).

At the same time, as highlighted by a comprehensive meta-analysis (Kober et al., 2008), right pSTS/TPJ is consistently activated when experiencing and perceiving emotions (Burnett and Blakemore, 2009; Nummenmaa et al., 2012). Indeed, activity of this brain area has been linked to the comprehension (Mano et al., 2009) and understanding of prosody (Hervé et al., 2012, 2013) of emotional narratives, to emotional contagion (Lee et al., 2007;

Nummenmaa et al., 2008) and empathy (Morelli et al., 2012; Morelli and Lieberman, 2013), and to the processing of emotionally-charged facial expressions (Srinivasan et al., 2016; Spunt and Adolphs, 2017). These findings highlight the relevance of the pSTS/TPJ in a variety of socio-affective abilities, likely coordinating the concert response of these processes during an emotional or social experience. However, further studies are needed to better understand the role and organization of this brain region.

Thus, it is still unclear how mentalizing abilities are represented in the brain and especially how the interaction between different socio-affective skills take place. Therefore, neuroimaging studies are in need of new more naturalistic paradigms to properly explore ToM and the different dimensions pertaining to it.

2.4 The current studies

In the current studies, we focused on the subjective emotional experience associated to a naturalistic stimulation. As it has been discussed in this chapter, together with the experience of emotions, several socio-affective abilities are likely to intervene to generate the final percept. In our studies, we used continuous behavioral ratings as a proxy of the subjective affective experience.

Humans use emotions to decipher complex cascades of internal events. However, which mechanisms link descriptions of affective states to brain activity is unclear, with evidence supporting either local or distributed processing. A biologically favorable alternative is provided by the notion of gradient, which postulates the isomorphism between functional representations of stimulus features and cortical distance. Therefore, in the first study we used fMRI activity evoked by an emotionally charged movie and continuous ratings of the perceived emotion intensity to reveal the topographic organization of affective states.

Moreover, emotions can rapidly change over time, as affective states are the consequence of an active interplay between an individual and the environment. Different models have been proposed to delineate the temporal characteristics of emotions (e.g., duration, onset, resurgence) and how these are represented in the brain, with the dynamics between specific regions related to different emotional experiences. However, the majority of neuroimaging studies so far have been employing brief or static stimuli to investigate emotion perception, which do not take into account the dynamism of affective states and their unfolding over time. In light of this, in the second study we aimed to overcome this limitation and explore the dynamic interplay between different brain regions throughout naturalistic situations. To do so, we tracked changes in functional connectivity among distinct brain regions during the watching of an emotionally charged movie. Continuous behavioral ratings of the perceived intensity of emotional states in an independent sample of subjects were then related to functional connectivity dynamics.

3. Emotionotopy in the human right temporo-parietal cortex

3.1 Introduction

Emotions promptly translate inner experiences into specific patterns of interpretable behaviors. To understand our own affective states and to infer those of others represent crucial aspects both when humans directly relate to each other and when they simply observe social interactions. Through years, the relevance of such abilities motivated the quest for models that optimally associate behavioral responses to emotional experiences.

In this regard, seminal works pointed toward the existence of discrete basic emotions characterized by distinctive and culturally stable facial expressions (Panksepp, 1982), patterns of autonomous nervous system activity (Kreibig, 2010; Stephens et al., 2010), and bodily sensations (Nummenmaa et al., 2014). Happiness, surprise, fear, sadness, anger and disgust represent the most frequently identified set of basic emotions (Ekman, 1992), though alternative models propose that other emotions, such as pride or contempt, should be included for their social and biological relevance (Tracy and Randles, 2011). To prove the neurobiological validity of these models, neuroscientists investigated whether basic emotions are consistently associated with specific patterns of brain responses across subjects. Findings show that activity in amygdala, medial prefrontal, anterior cingulate, insular, middle/inferior frontal, and posterior superior temporal cortex, is associated to the perceived intensity of emotions and supports their recognition (Vytal and Hamann, 2010; Peelen et al., 2010; Saarimäki et al., 2015; Kragel and LaBar, 2015). However, this perspective has been challenged by other studies, which failed to demonstrate significant associations between single emotions and activity within distinct cortical areas or networks (Kober et al., 2008; Lindquist et al., 2012; Touroutoglou et al., 2015).

Dimensional theories propose instead that behavioral, physiological and subjective characteristics of emotions would be more appropriately described along a number of continuous cardinal dimensions (Russell, 1980; Smith and Ellsworth, 1985), generally one governing pleasure versus displeasure (i.e., valence) and another one the strength of the experience (i.e., arousal). While these two dimensions have been reliably and consistently described, alternative models propose that additional dimensions, such as dominance or unpredictability, are needed to adequately explain affective states (Mehrabian and Russell, 1974; Fontaine et al., 2007). Neuroimaging studies also demonstrated that stimuli varying in valence and arousal elicit specific and reliable brain responses (Anderson et al., 2003; Wager et al., 2003), which have been recently employed to decode emotional experiences (Kassam et al., 2013). Activity recorded in insula, amygdala, ventral striatum, anterior cingulate, ventromedial prefrontal and posterior territories of the superior temporal cortex, is associated to transitions between positive and negative valence and fluctuations in arousal (Nummenmaa et al., 2012; Lindquist et al., 2015).

Of note, other than in the ventromedial prefrontal regions, studies using either discrete emotion categories (Kober et al., 2008; Peelen et al., 2010; Kragel and LaBar, 2015) or emotion dimensions (Mourao-Miranda et al., 2003; Kensinger and Schacter, 2006; Mather et al., 2006; Nummenmaa et al., 2012) have shown responses in the posterior portion of the superior temporal cortex, extending to temporo-parietal territories. Furthermore, these temporo-parietal regions are fundamental for social cognition, as they support empathic processing (Morelli et al., 2012; Morelli and Lieberman, 2013) and the attribution of intentions, beliefs and emotions to others (Saxe and Kanwisher, 2003; Skerry and Saxe, 2015).

However, despite this large body of evidence, it remains to be determined whether emotional experiences are better described through discrete basic emotions or emotion dimensions. Moreover, regardless of the adopted model, it is still debated how emotion features are spatially encoded in the brain (Barrett and Wager, 2006; Lindquist et al., 2012; Touroutoglou et al., 2015; Saarimäki et al., 2015; Clark-Polner, et al., 2017; Saarimäki et al., 2018). As a matter of fact, while findings support the role of distinct regions (Vytal and Hamann, 2010), others indicate the recruitment of distributed networks in relation to specific affective states (Wager et al., 2015).

An alternative and biologically favorable perspective may be provided by the notion of gradient. Gradients have been proven a fundamental organizing principle through which the brain efficiently represents and integrates stimuli coming from the external world. For instance, the location of a stimulus in the visual field is easily described through two orthogonal spatially overlapping gradients in primary visual cortex: rostrocaudal for eccentricity and dorsoventral for polar angle (Sereno et al., 1995). Thus, using fMRI and retinotopic mapping, one can easily predict the location of a stimulus in the visual field considering the spatial arrangement of recruited voxels with respect to these orthogonal gradients. Crucially, recent investigations revealed that gradients support the representation of higher-order information as well (Huth et al., 2016; Margulies et al., 2016; Huntenburg et al., 2018), with features as animacy or numerosity being topographically arranged onto the cortical mantle (Harvey et al., 2013; Sha et al., 2015; Huth et al., 2016).

Following this view, we hypothesize that affective states are encoded in a gradient-like manner in the human brain. Specifically, different affective states would be mapped onto the cortical mantle through spatially overlapping gradients, which would code either the intensity of discrete emotions (e.g., weak to strong sadness) or, alternatively, the smooth transitions along cardinal dimensions (e.g., negative to positive valence). In either case, the pattern of brain activity could be interpreted according to emotion gradients to predict the subjective affective state.

Here, we tested this hypothesis using moment-by-moment ratings of the perceived intensity of emotions elicited by an emotionally charged movie. To unveil cortical regions involved in emotion processing, behavioral ratings were used as predictors of fMRI activity in an independent sample of subjects exposed to the same movie. The correspondence between functional characteristics and the relative spatial arrangement of distinct patches of cortex was then tested to reveal the existence of emotion gradients.

3.2 Materials and Methods

In the present study, we took advantage of a high quality publicly available dataset, part of the studyforrest project (Hanke et al., 2016; http://studyforrest.org), to demonstrate the existence of a gradient-like organization in brain regions coding emotion ratings. Particularly, we used moment-by-moment scores of the perceived intensity of six basic emotions elicited by an emotionally charged movie (Forrest Gump; R. Zemeckis, Paramount Pictures, 1994), as predictors of fMRI activity in an independent sample. We then tested the correspondence between the fitting of the emotion rating model in the peak of association with brain activity (TPJ) and the relative spatial arrangement of voxels in this region to reveal the existence of orthogonal spatially overlapping gradients.

3.2.1 Behavioral Experiment

3.2.1.1 Participants. To obtain moment-by-moment emotion ratings during the Forrest Gump movie, we enrolled 12 healthy Italian native speakers (5F; mean age 26.6 years, range 24-34). None of the participants reported to have watched the movie in one-year period prior to the experiment, while two of them watched it in their infancy and had no detailed memories of it. Subjects signed an informed consent to participate in the study, had the right to withdraw at any time and received a small monetary compensation for their participation. The study was conducted in accordance with the Declaration of Helsinki and was approved by the local IRB (Protocol N°1485/2017).

3.2.1.2 Experimental setup. We started from the Italian dubbed version of the movie, edited following the exact same description reported in the studyforrest project (eight movie segments ranging from a duration of 11 to 18 minutes). The movie was presented in a setting free from distractions using a 24" monitor with a resolution of 1920x1080 pixels connected to a MacBook™ Pro running Psychtoolbox v3.0.14 (Kleiner et al., 2007). Participants wore headphones in a noiseless environment (Sennheiser™ HD201; 21-18,000 Hz; Maximum SPL 108dB) and were instructed to continuously rate the subjective perceived intensity (on a scale ranging from 0 to 100) of six basic emotions throughout the entire movie: happiness, surprise, fear, sadness, anger and disgust (Ekman, 1992). Specific buttons on the keyboard placed in front of subjects mapped the increase and decrease in intensity of each emotion and they were instructed to represent their inner experience by freely adjusting or maintaining the level of intensity. Moreover, the six emotion labels were always visible on the screen so that subjects had a continuous visual feedback of what they were rating. Participants were allowed to report more than one emotion at the same time and ratings were continuously recorded with a 10Hz-sampling rate. Subjects were presented with the same eight movie segments employed in the fMRI study one after the other, for an overall duration of 120 minutes. Further, before starting the actual emotion rating experiment, all participants performed a 20 minutes training session to familiarize with the experimental procedure. Specifically, they had to reproduce various levels of intensity for random combinations of emotions that appeared on the screen every ten seconds. Lastly, the experimenter was present in the room with the participants but stayed out of their sight.

Unfortunately, no objective psychophysiological measures were recorded during the experiment. Future studies might address this issue and use the recollection of these data for further analysis.

3.2.1.3 Behavioral data pre-processing. For each subject, we recorded six timeseries representing the moment-by-moment perceived intensity of basic emotions. First, we downsampled timeseries to match the fMRI temporal resolution (2s) and, afterwards, we introduced a lag of 2s to account for the delay in hemodynamic activity. The resulting timeseries were then temporally smoothed using a moving average procedure (10s window). This method allowed us to further account for the uncertainty of the temporal relationship between the actual onset of emotions and the time required to report the emotional state.

3.2.1.4 Agreement across subjects of the six basic emotions. To verify the consistency in the occurrence of affective states while watching the Forrest Gump movie, we computed the Spearman's ρ correlation coefficient across subjects for each of the six ratings.

Statistical significance of the agreement was assessed by generating a null distribution of random ratings using the IAAFT procedure (Iterative Amplitude Adjusted Fourier Transform, Schreiber and Schmitz, 1996; Chaotic System Toolbox), which provided surrogate data with the same spectral density and temporal autocorrelation of the averaged ratings across subjects (1,000 surrogates).

Basic emotion and emotion dimension models. Preprocessed and temporally smoothed single-subject emotion ratings were averaged to obtain six group-level timeseries representing the basic emotion model. After measuring the Spearman's ρ between pairings of basic emotions, we performed principal component (PC) analysis and identified six orthogonal components, which constituted the emotion dimension model.

3.2.1.5 Agreement across subjects of the emotion dimensions. To verify the consistency across subjects of the PCs, we computed the agreement of the six components by means of a leave-one-subject-out cross validation procedure. Specifically, for each iteration, we performed PC analysis on the left-out subject behavioral ratings and on the averaged ratings of all the other participants. The six components obtained from each left-out subject were rotated (Procrustes analysis, reflection and orthogonal rotation only) to match those derived from all the other participants. This procedure generated for each iteration (i.e., for each of the left-out subjects) six components, which were then compared across individuals using Spearman's ρ similarly to what has been done for the six basic emotions. To assess the statistical significance, we created a null distribution of PCs from the generated surrogate data of the behavioral ratings, as described above (1,000 surrogates).

3.2.1.6 Richness of the reported emotional experience. Although subjects were asked to report their inner experience using six emotion categories, their ratings were not limited to binary choices. Indeed, at each timepoint raters could specify the perceived intensity of more than one emotion, leading to the definition of more complex affective states as compared to the basic ones. To further highlight this aspect, we performed dimensionality reduction and clustering analyses on emotion timeseries. Starting from emotion ratings averaged across participants, we selected timepoints characterized by the highest intensity (i.e., by summing the six basic emotions and setting the threshold to the 50th percentile) and applied Barnes-Hut t-distributed stochastic neighbor embedding (Maaten and Hinton, 2008; Cowen and Keltner, 2017; t-SNE; perplexity = 30; theta = 0.05). The algorithm measures the distances between timepoints in the six-dimensional space defined by the basic emotions as joint probabilities according to a Gaussian distribution. These distances are projected onto a two-dimensional embedding space using a Student's t probability distribution and by minimizing the Kullback–Leibler divergence. To further describe the variety of affective states elicited by the movie, we then applied k-means clustering analysis to the projection of timepoints in the t-SNE manifold and determined the number of clusters using the silhouette criterion (Rousseeuw and Kaufman, 1990).

3.2.2 fMRI Experiment

We selected data from the phase II of the studyforrest project, in which 15 German mother tongue subjects watched an edited version of the Forrest Gump movie during the fMRI acquisition. Participants underwent two 1-hour sessions of fMRI scanning (3T, TR 2s, TE 30ms, FA 90°, 3mm ISO, FoV 240mm, 3599 tps), with an overall duration of the experiment of 2h across eight runs. Subjects were instructed to inhibit any movement and simply enjoy the movie (for further details40). We included in our study all participants that underwent the fMRI acquisition and had the complete recordings of the physiological parameters (i.e., cardiac trace) throughout the scanning time (14 subjects; 6F; mean age 29.4 years, range 20-40 years).

3.2.2.1 fMRI data preprocessing. We employed ANTs (Avants et al., 2009) and AFNI v.17.2.00 (Cox, 1996) to preprocess MRI data. First, structural images were brain extracted (antsBrainExtraction.sh) and non-linearly transformed to match the MNI152 template (3dQwarp). The estimated deformation field was subsequently used to bring single-subject activation maps from the original to the standard space. Functional data were corrected for intensity spikes (3dDespike) and adjusted for slice timing acquisition (3dTshift). We also compensated head movements by registering each volume to the most stable timepoint (3dvolreg). In this regard, a rigid body transformation was adopted, and the six estimated motion parameters were included as confounds in further analyses. The transformation matrices were also used to compute an aggregated measure - framewise displacement87 - that highlighted timepoints affected by excessive motion. Functional data were linearly (align_epi_anat.py) and non-linearly registered to the T1w images, also correcting for phase distortion, and warped to match the MNI152 template using the already computed deformation field (3dNwarpApply). Furthermore, timeseries were smoothed until they reached a full width at half maximum of 6mm (Gaussian kernel). In this regard, we did not simply apply a 6mm smoothing filter to the original data, rather we adopted the AFNI's 3dBlurToFWHM tool, which estimates and iteratively increases the smoothness of data until a specific FWHM level is reached. Lastly, we ruled out the effects of signal drifts, head motion and heartbeat (3dDeconvolve) to obtain timeseries of brain activity cleaned from these nuisance regressors.

Following the same procedure adopted for the behavioral processing, single-subject preprocessed fMRI data were averaged to obtain group-level hemodynamic activity and for each voxel the same windowing procedure was employed to temporally smooth data (moving average: 10s window). From the obtained aggregated and smoothed timeseries, the timecourse of low-level acoustic (i.e., volume energy - RMS of the signal) and visual (i.e., Gabor contrast energy for 0.5 and 8 cyc/deg spatial frequencies for each frame) features of the movie were regressed out to mitigate the possible collinearities between emotion ratings and psychophysical properties of the stimulus (e.g., fearful events might be associated to sudden volume increases). Specifically, the RMS value was estimated on 2s non-overlapping windows (Lahnakoski et al., 2012) matching the TR of the fMRI scan. For the low-level visual features instead, we modeled the canonical response of area V1(Hubel and Wiesel, 1962). Each movie frame was filtered with a set of oriented Gabor filters encompassing the lowest and highest limits of V1 spatial frequency selectivity (0.5 and 8 cyc/deg), as found by cell recordings in non-human (Foster et al., 1985) and by fMRI in humans' primates (Kay et al., 2008).

Filters response was averaged across four orientations (i.e., 0, 45, 90, 135 deg) and all pixels, to obtain a global descriptor for each frequency in each frame. Visual features were then temporally averaged across frames, delayed and smoothed in time to match the temporal resolution of fMRI data.

Overall, low-level features modelling generated three regressors of no interest (i.e., low and high spatial frequencies of movie frames and RMS of the audio track) that were regressed out from brain activity using a multiple regression analysis. The obtained regression residuals, consisting of 3,595 timepoints, were considered as the dependent variable in the encoding analysis having emotional ratings as predictors.

3.2.3 Encoding Analysis

Voxel-wise encoding (Naselaris et al., 2011; Handjaras et al., 2017) was performed using a multiple linear regression approach to measure the association between brain activity and the emotion ratings, constituted by the six principal components. We believe that our sample size allowed for this type of analysis, as, considering the total amount of collected data (i.e., degrees of freedom), our final sample was composed by 3,595 timepoints acquired in 14 individuals. Moreover, we are confident that the differences between the two groups are limited, as all participants pertain to the same macro-culture, allowing us to test the association between German subjects' brain activity and Italian's emotion ratings. Lastly, an encoding analysis performed using two different samples should provide even more robust results and corroborate the possible conclusions.

Of note, performing a least square linear regression using either the six principal components or the six basic emotion ratings yields the same overall fitting (i.e., full model R^2), even though the coefficient of each column could vary among the two predictor sets. To reduce the computational effort, we limited the regression procedure to gray matter voxels only (~44k with an isotropic voxel resolution of 3mm). We assessed the statistical significance of the R^2 fitting of the model for each voxel using a permutation approach, by generating 10,000 null encoding models. Null models were obtained by measuring the association between brain activity and surrogate data having the same spectral density and temporal autocorrelation of the original six PCs. This procedure provided a null distribution of R^2 coefficients, against which the actual association was tested. The resulting p-values were corrected for multiple comparisons through the False Discovery Rate method ($q < 0.01$; Benjamini and Hochberg, 1995; Supplementary Figure 1 and Supplementary Table 1 in the Appendix). R^2 standard error was calculated through a bootstrapping procedure (1,000 iterations). Moreover, we conducted a noise-ceiling analysis for right TPJ data, similarly to what has been done by Ejaz and colleagues, 2015 (please see the Appendix).

3.2.4 Emotion Gradients in right TPJ

3.2.4.1 Estimation of right TPJ topography. We tested the existence of emotion gradients by measuring the topographic arrangement of the multiple regression coefficients in regions lying close to the peak of fitting for the encoding procedure (i.e., right pSTS/TPJ). To avoid any circularity in the analysis (Kriegeskorte et al., 2009), we first delineated a region of interest (ROI) in the right pSTS/TPJ territories using an unbiased procedure based on the NeuroSynth74 database v0.6 (i.e., reverse inference probability associated to the term "TPJ"). Specifically, we started from the peak of the "TPJ" NeuroSynth reverse inference meta-analytic map to draw a series of cortical ROIs, with a radius ranging from 9 to 27 mm. Afterwards, to identify the radius showing the highest significant association, for each spherical ROI we tested the relationship between anatomical and functional distance (Yarrow et al., 2014; Supplementary Table 2 in the Appendix). This procedure was performed using either multiple regression coefficients obtained from the three emotion dimensions or from the four basic emotions stable across all subjects. We built for each radius two dissimilarity matrices: one using the Euclidean distance of voxel coordinates, and the other one using the Euclidean distance of the fitting coefficients (i.e., ϱ values) of either the three emotion dimensions or the four basic emotions (Supplementary Figure 2 in the Appendix). The rationale behind the existence of a gradient-like organization is that voxels with similar functional behavior (i.e., lower functional distance) would also be spatially arranged close to each other on the cortex75 (i.e., lower physical distance). The functional and anatomical dissimilarity matrices were compared using the Spearman's ϱ coefficient. To properly address the significance of the anatomo-functional association, we built an ad hoc procedure that maintained the same spatial autocorrelation structure of TPJ in the null distribution. Specifically, we generated 1,000 IAAFT-based null models for the emotion dimension and the basic emotion data, respectively. These null models represented the predictors in a multiple regression analysis and generated a set of null ϱ regression coefficients. Starting from these coefficients we built a set of functional dissimilarity matrices that have been correlated to the anatomical distance and provided 1,000 null Spearman's ϱ coefficients, against which the actual anatomo-functional relationship was tested. Confidence intervals (CI, 2.5 and 97.5 percentile) for the obtained correlation values were calculated employing a bootstrap procedure (1,000 iterations). We also tested the existence of gradients in other brain regions encoding emotion ratings using a data-driven searchlight analysis. Results and details of this procedure are reported in the Appendix.

3.2.4.2 Impact of spatial smoothing on emotion topography. To estimate the significance of right TPJ gradients we used null models built on emotion ratings, leaving untouched the spatial and temporal structure of brain activity. However, as spatial smoothness may still affect the estimation of gradients, we tested right TPJ topography using the group-average unfiltered data. In brief, all the steps described in the fMRI data pre-processing section were applied, with the only exception of spatial filtering. Following this procedure, the estimated smoothness of the right TPJ region was 4.5x4.2x3.6mm (3dFWHMx tool). Using these data and the same procedure described in the Estimation of right TPJ topography paragraph, we measured the significance of emotion gradients. Results are detailed in Supplementary Table 6 and Supplementary Figure 7 in the Appendix.

3.2.4.3 Impact of cortical folding on emotion topography. The Euclidean metric does not take into account cortical folding. Indeed, because of the morphological characteristics of TPJ, which include a substantial portion of STS sulcal walls, the estimation of emotion gradients would benefit from the use of a metric respectful of cortical topology.

For this reason, we ran the Freesurfer recon-all analysis pipeline (Reuter et al., 2012) on the standard space template (Fonov et al., 2009) used as reference for the nonlinear alignment of single-subject data. We then transformed the obtained files in AFNI-compatible format (@SUMA_Make_Spec_FS). This procedure provided a reconstruction of the cortical ribbon (i.e., the space between pial surface and gray-to-white matter boundary), which has been used to measure the anatomical distance. In this regard, we particularly employed the Dijkstra algorithm as it represents a computationally efficient method to estimate cortical distance based on folding (Fischl et al., 1999; Van Essen et al., 2011). The single-subject unsmoothed timeseries were then transformed into the standard space, averaged across individuals and projected onto the cortical surface (AFNI 3dVol2Surf, map function: average, 15 steps). Afterwards, we performed a multiple linear regression analysis using principal components derived from emotion ratings as predictors of the unsmoothed functional data. This analysis was carried out within a cortical patch that well approximated the size of the 3D-sphere used in the original volumetric pipeline and centered at the closest cortical point with respect to the Neurosynth "TPJ" peak. Thus, for each regressor of interest, we obtained unsmoothed β values projected onto the cortical mantle. We then tested the existence of a gradient-like organization for each predictor, using the Dijkstra algorithm and the same procedure described above (i.e., Estimation of right TPJ topography paragraph). Results are detailed in Supplementary Table 6 in the Appendix and Figure 4.

3.2.5 Right temporo-parietal gradients and portrayed emotions

We tested whether the gradient-like organization of right TPJ reflects portrayed emotions. Thus, we took advantage of publicly available emotion tagging data of the same movie, provided by an independent group (Labs et al., 2015). Differently from our behavioral task, raters were asked to indicate the portrayed emotion of each character (e.g., Forrest Gump, Jenny) in 205 movie segments (average duration ~35s) presented in random order and labeled over the course of approximately three weeks. As also suggested by the authors (Labs et al., 2015), this particular procedure minimizes carry-over effects and help observers to exclusively focus on indicators of portrayed emotions. Importantly, in their editing, the authors respected the narrative of movie scenes (e.g., Forrest in front of Jenny's grave is a single cut with a duration of ~131s), so that the raters could have a clear understanding of what was shown on the screen. In addition, the possibility to tag emotions independently in each movie segment and to watch each scene more than once, allowed subjects to choose among a larger number of emotion categories (Ortony et al., 1990; N = 22), as compared to our set of emotions. Moreover, each observer was instructed to report with a binary label whether the portrayed emotion was directed toward the character itself (self-directed; e.g., Forrest feeling sad) or toward another character (other-directed; e.g., Forrest feeling happy for Jenny). These two descriptions served as third-person emotion attribution models and underwent the exact same processing steps (i.e., 2s lagging and temporal smoothing), which have been applied to our subjective emotion rating model. As the two third-person emotion attribution models included the four basic emotions found to be consistent across observers in our experiment (i.e., happiness, fear, sadness and anger), we have been able to directly assess the correlation for these ratings using Spearman's ϱ.

We then measured the extent to which the two third-person emotion attribution models explained brain activity in right TPJ following the method described in the Encoding analysis section. As these two descriptions are higher in dimensionality as compared to our subjective emotion rating model, we assessed the significance of fitting using three different procedures: (A) matching the dimensionality across models by selecting the first six principal components only; (B) matching the emotion categories in ratings, by performing PCA on the four basic emotions shared across models (i.e., happiness, fear, sadness and anger); (C) using the full model regardless of the dimensionality (i.e., six components for our subjective emotion rating model and 22 for each of the two emotion attribution models). In addition, to allow a direct and unbiased comparison between R^2 values obtained from different models, we performed cross-validation using a half-run split method (for further details please refer to the Appendix).

Lastly, we tested whether right TPJ gradients encode emotion attribution models. Specifically, we evaluated two different scenarios: (1) the existence of right TPJ gradients encoding the 22 components of each emotion attribution model; (2) the possibility to identify emotion gradients following the multidimensional alignment81 (i.e., canonical correlation analysis) of the 22-dimensional emotion attribution space to the 6-dimensional space defined by subjective ratings. These alternative procedures relate to two different questions: (1) whether the process of emotion attribution is associated to emotion gradients in right TPJ and (2) whether starting from a third-person complex description of portrayed emotions, one can reconstruct the subjective report of our raters. Results for these two procedures are detailed in Supplementary Table 5 in the Appendix.

3.2.6 Characterization of emotion gradients in right TPJ

3.2.6.1 Principal orientation of right TPJ gradients. Once the optimal ROI radius was identified, we tested the gradient-like organization of right TPJ for each individual emotion dimension and basic emotion (Supplementary Table 3 in the Appendix), using the same procedure described above (i.e., Estimation of right TPJ topography section). We calculated the numerical gradient of each voxel using ϱ values. This numerical gradient estimates the partial derivatives in each spatial dimension (x, y, z) and voxel, and can be interpreted as a vector field pointing in the physical direction of increasing ϱ values. Afterwards, to characterize the main direction of each gradient, rather than calculating its divergence (i.e., Laplacian of the original data, Glasser et al., 2016; Haak et al., 2017), we computed the sum of all vectors in the field. This procedure is particularly useful to reveal the principal direction of linear gradients and provides the opportunity to represent this direction as the orientation of the symmetry axis of the selected ROI. The above-mentioned procedure was also adopted to assess the reliability of the emotion gradients in each subject. Results and details of this procedure are reported in the Appendix. Furthermore, since gradients built on ϱ coefficients could reflect positive or negative changes in hemodynamic signal depending on the sign of the predictor, we represented the average TPJ activity during movie scenes characterized by specific affective states.

<u>*3.2.6.2 Population receptive field estimates in right TPJ.*</u> We investigated whether distinct populations of voxels are selective for specific affective states. To this aim, we employed the population receptive field method (pRF; Dumoulin and Wandell, 2008) and estimated the tuning curve of right TPJ voxels for each predictor found to be topographically encoded within this region. We modeled the tuning curve of each voxel as a Gaussian distribution, in which μ represented the preferred score of the predictor and σ the width of the response. The optimal combination of tuning parameters was selected among ~5k plausible values of μ (5th-95th percentile of the scores of each predictor - 0.5 step) and σ (ranging from 1 to 12 - 0.25 step), sampled on a regular grid. Each emotion timeseries was then filtered using these ~5k Gaussian distributions and fitted in brain activity using a linear regression approach. This produced t-values (i.e., $\beta / SE\ \beta$) expressing the goodness of fit of μ and σ combinations, for each right TPJ voxel. The principal tuning of voxels was then obtained by selecting the combination characterized by the highest t-value across the ~5k samples.

To estimate the similarity between tunings (i.e., μ parameters) obtained from the pRF approach and our original results (i.e., β coefficients of the gradient estimation), we computed Spearman's ϱ across right TPJ voxels. The significance of such an association was tested against a null distribution of β coefficients obtained through the IAAFT procedure (N = 1,000).

Lastly, we further characterized the prototypical responses of populations of voxels as function of affective states. To do so, we used the non-negative matrix factorization84 and decomposed the multivariate pRF data (i.e., voxels t-values for each μ and σ) into an approximated matrix of lower rank (i.e., 10, retaining at least 90% of the total variance). This method allows parts-based representations, as the tuning of right TPJ voxels is computed as a linear summation of non-negative basis responses. The results of this procedure are summarized in Supplementary Figure 9 in the Appendix. All the analyses were performed using MATLAB R2016b (MathWorks Inc., Natick, MA, USA).

3.3 Results

3.3.1 Emotion Ratings

A group of Italian native speakers continuously rated the perceived intensity of six basic emotions (i.e., happiness, surprise, fear, sadness, anger and disgust; Ekman, 1992) while watching an edited version of the Forrest Gump movie (R. Zemeckis, Paramount Pictures, 1994). We first assessed how much each basic emotion contributed to the behavioral ratings and found that happiness and sadness explained 28% and 36% of the total variance, respectively. Altogether, fear (18%), surprise (8%), anger (7%), and disgust (3%) explained the remaining one-third of the total variance. We also evaluated the agreement in ratings of the six basic emotions (Figure 2A), and found that happiness (Spearman's $\varrho = 0.476 \pm 0.102$, range 0.202 - 0.717), fear ($\varrho = 0.522 \pm 0.134$, range 0.243 - 0.793), sadness ($\varrho = 0.509 \pm 0.084$, range 0.253 - 0.670), and anger ($\varrho = 0.390 \pm 0.072$, range 0.199 - 0.627) were consistent across all the subjects, whereas surprise ($\varrho = 0.236 \pm 0.099$, range 0.010 - 0.436) and disgust ($\varrho = 0.269 \pm 0.115$, range 0.010 - 0.549) were not. Nonetheless, ratings for these latter emotions were on average significantly different from a null distribution of randomly assigned emotion ratings (p-value < 0.05).

To reveal emotion dimensions, we averaged across subjects the ratings of the six basic emotions, measured their collinearity (Figure 2B) and performed PCA (Figure 2C). The first component reflected a measure of polarity (PC1: 45% explained variance) as positive and negative emotions demonstrated opposite loadings. The second component was interpreted as a measure of complexity (PC2: 24% explained variance) of the perceived affective state, ranging from a positive pole where happiness and sadness together denoted inner conflict and ambivalence, to a negative pole mainly representing fearful events. The third component was a measure of intensity (PC3: 16% explained variance), since all the six basic emotions showed positive loadings (Figure 2C). Altogether, the first three components explained approximately 85% of the total variance. We further assessed the stability of the PCs and found that only these first three components (polarity: $\varrho = 0.610 \pm 0.089$, range 0.384 - 0.757; complexity: $\varrho = 0.453 \pm 0.089$, range 0.227 - 0.645; intensity: $\varrho = 0.431 \pm 0.071$, range 0.258 - 0.606), hereinafter emotion dimensions, were consistent across all the subjects (Figure 2D). The fourth PC described movie segments during which participants experienced anger and disgust at the same time (PC4: 8% explained variance, $\varrho = 0.329 \pm 0.128$, range -0.003 - 0.529), whereas the fifth PC was mainly related to surprise (PC5: 6% explained variance, $\varrho = 0.214 \pm 0.090$, range 0.028 - 0.397). Notably, these two PCs were not consistent across all the subjects, even though their scores were on average significantly different from a null distribution ($p\text{-value} < 0.05$). Scores of the sixth PC were not significantly consistent across subjects (PC6: 1% explained variance, $p\text{-value} > 0.05$).

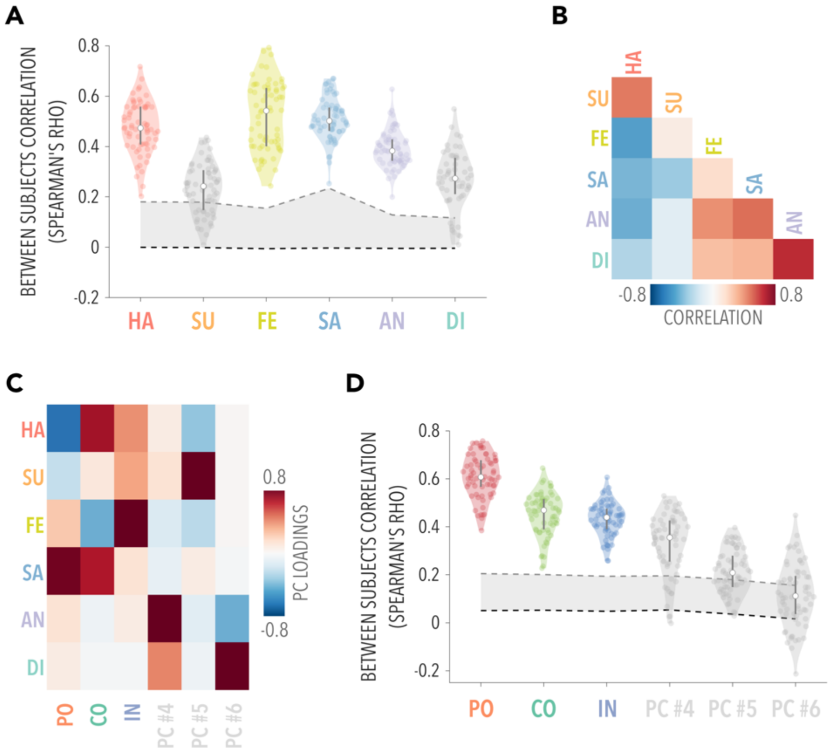


Figure 2 Emotion ratings - **A**. Violin plots show the agreement (Spearman's ϱ coefficient) of the six basic emotions across subjects. Gray shaded area represents the null distribution of behavioral ratings and dashed lines the mean and 95th percentile of the null distribution. **B**. Correlation matrix showing Spearman's ϱ values for pairings of basic emotions. **C**. Principal Component Analysis: loadings of the six principal components. Explained variance was 45% for polarity, 24% for complexity and 16% for intensity. **D**. Violin plots show the agreement (Spearman's ϱ coefficient) of the six principal components across subjects. Gray shaded area represents the null distribution of behavioral ratings and dashed lines the mean and 95th percentile of the null distribution. HA = Happiness, SU = Surprise, FE = Fear, SA = Sadness, AN = Anger, DI = Disgust, PC = Principal component, PO = Polarity, CO = Complexity, IN = Intensity.

3.3.2 Richness of the Reported Emotional Experience

In our behavioral experiment, participants were allowed to report the perceived intensity of more than one emotion at a time. Thus, the final number of elicited emotional states might be greater than the original six emotion categories. To measure the richness of affective states reported by our participants, we performed dimensionality reduction and clustering analyses on group-averaged behavioral ratings. Results revealed the existence of 15 distinct affective states throughout the movie (Figure 3). Among these, some states were characterized by a single basic emotion, whereas others by a peculiar mixture of them. Combinations of distinct emotions likely expressed secondary affective states, as ambivalence (i.e., cluster j depicting movie scenes in which happiness and sadness are simultaneously experienced) or resentment (i.e., cluster i representing movie segments in which a mixture of sadness, anger and disgust is perceived). Of note, this evidence is supported also by single-subject reports, in which the 38% (SE: ±2.3%) of timepoints were associated to a single emotion, the 29% (SE: ±3.5%) to two basic emotions and the 6% (SE: ±1.4%) to the concurrent experience of three distinct emotions. Altogether, these results show that the Forrest Gump movie evoked complex and multifaceted experiences, which cannot be reduced to the original six categories.

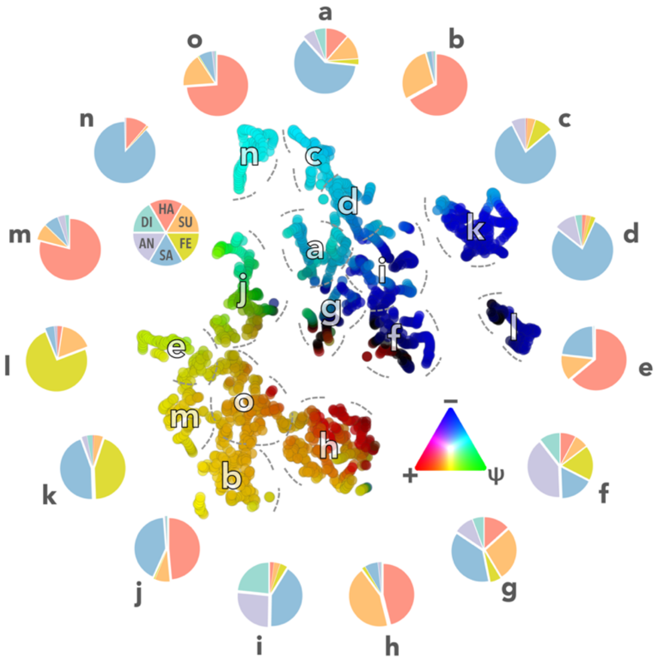


Figure 3 Richness of the emotional experience - Results of the dimensionality reduction (t-SNE) and clustering analyses (k-means) on the group-averaged behavioral ratings showing the existence of 15 distinct affective states throughout the movie (clusters denoted with alphabetical letters). Each element represents a specific timepoint in the movie and the distance between elements depends on the statistical similarity of emotion ratings. Element color reflects the scores of the polarity and complexity dimensions: positive (+) and negative (-) events (i.e., polarity) are associated respectively to the red and blue channels, whereas complexity (Ψ) scores modulate the green channel. Pie charts show the relative contribution of the six basic emotions to each of the 15 identified clusters. HA = Happiness, SU = Surprise, FE = Fear, SA = Sadness, AN = Anger, DI = Disgust.

3.3.3 Brain Regions encoding Emotion Ratings

Emotion ratings obtained from the behavioral experiment were used as predictors of brain activity in independent subjects exposed to the same movie (Hanke et al., 2016; http://studyforrest.org). The model significantly explained activity in right inferior frontal gyrus (IFG), rostral middle frontal gyrus (rMFG), medial superior frontal gyrus (mSFG), occipitotemporal sulcus (OTS), precentral sulcus (preCS), posterior part of the superior temporal sulcus/temporoparietal junction (pSTS/TPJ), middle occipital gyrus (MOG) and posterior middle temporal gyrus (pMTG). We also observed significant results in the left supramarginal gyrus (SMG) and pMTG ($q < 0.01$ FDR corrected and cluster size > 10; Figure 4A). Notably, the peak of association between emotion ratings and brain activity was located in the right pSTS/TPJ, an important region for social cognition (Saxe and Kanwisher, 2003; Kober et al., 2008; Nummenmaa et al., 2012; Morelli et al., 2012; Skerry and Saxe, 2015; $R^2 = 0.07 \pm SE = 0.009$; Center of Gravity – CoG: x = 61, y = -40, z = 19; noise ceiling lower bound 0.13, upper bound 0.23; Figure 4B). The peak of association was also located in proximity (11 mm displacement) of the reverse inference peak for the term "TPJ" (CoG: x = 58, y = -50, z = 16) as reported in the NeuroSynth database (neurosynth.org; Figure 4B).

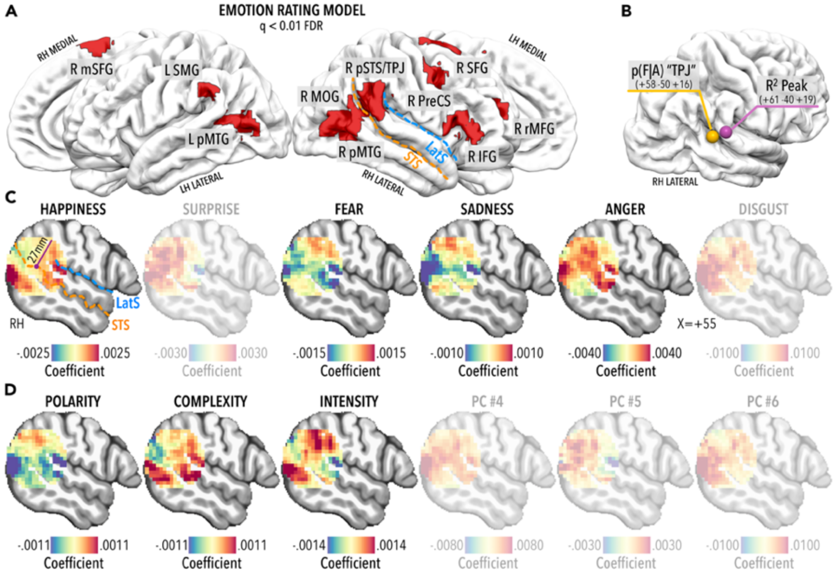


Figure 4 Encoding of emotion ratings - **A**. Brain regions encoding emotion ratings corrected for multiple comparisons through the False Discovery Rate method ($q < 0.01$). **B**. Peak of association between emotion ratings and brain activity (purple sphere) and reverse inference peak for the term "TPJ" as reported in the NeuroSynth database (yellow sphere). Coordinates represent the center of

gravity in MNI152 space. **C**. β coefficients associated to basic emotions in a spherical region of interest (27mm radius) located at the reverse inference peak for the term "TPJ". Maps for emotions not consistent across all the subjects (i.e., surprise and disgust) are faded (see the Agreement across subjects of the six basic emotions section). **D**. β coefficients associated to emotion dimensions in a spherical region of interest (27mm radius) located at the reverse inference peak for the term "TPJ". Maps for components not consistent across all the subjects (i.e., PC4, PC5 and PC6) are faded (see the Agreement across subjects of the Emotion Dimensions section). IFG = Inferior Frontal Gyrus, rMFG = rostral Middle Frontal Gyrus, mSFG = Medial Superior Frontal Gyrus, preCS = Precentral Sulcus, pSTS/TPJ = posterior part of the Superior Temporal Sulcus/Temporoparietal Junction, MOG = Middle Occipital Gyrus, pMTG = posterior Middle Temporal Gyrus, SMG = Supramarginal Gyrus, LatS = Lateral Sulcus, STS = Superior Temporal Sulcus.

3.3.4 Emotion gradients in right temporo-parietal territories

To avoid double-dipping, we tested the existence of either basic emotion or emotion dimension gradients in a spherical region of interest located at the reverse inference peak for the term "TPJ". This analysis was conducted on behavioral ratings consistent across all the subjects: happiness, sadness, fear and anger for basic emotions and polarity, complexity and intensity for emotion dimensions.

Using ϱ coefficients obtained from the encoding analysis, we observed that, within right TPJ, voxels appeared to encode happiness in an anterior to posterior arrangement, fear and sadness in an inferior to superior manner, while anger showed a patchier organization (Figure 4C). With respect to emotion dimensions, voxels seemed to encode polarity and intensity in a more inferior to superior fashion, whereas complexity in a more posterior to anterior direction (Figure 4D).

To prove the existence and precisely characterize the orientation of these gradients, we tested the association between physical distance and functional characteristics of right TPJ voxels (Supplementary Figure 2 in the Appendix). Results demonstrated that within a 15 mm radius sphere, the relative spatial arrangement and functional features of right TPJ were significantly and maximally correlated, either considering the basic emotion model ($\varrho = 0.352$, p-value = 0.004, 95% Confidence Interval - CI: 0.346 to 0.357) or the emotion dimension one ($\varrho =$

0.399, p-value < 0.001, 95% CI: 0.393 to 0.404; for alternative definitions of the right TPJ region see Supplementary Table 2 in the Appendix).

Crucially, when focusing on each emotion dimension, results revealed the existence of three orthogonal and spatially overlapping gradients: polarity ($\varrho = 0.241$, p-value = 0.041, 95% CI: 0.235 to 0.247), complexity ($\varrho = 0.271$, p-value = 0.013, 95% CI: 0.265 to 0.277) and intensity ($\varrho = 0.229$, p-value = 0.049, 95% CI: 0.223 to 0.235; Figure 5 and Supplementary Table 3 in the Appendix). On the contrary, happiness ($\varrho = 0.275$, p-value = 0.013, 95% CI: 0.269 to 0.281), but not other basic emotions (fear: $\varrho = 0.197$, p-value = 0.091; sadness: $\varrho = 0.182$, p-value = 0.160; anger: $\varrho = 0.141$, p-value = 0.379; Supplementary Table 3 in the Appendix), retained a gradient-like organization. Of note, the peculiar arrangement of group-level emotion dimension gradients (Figure 5) was also identified using single-subject fMRI data (Supplementary Figure 3 and Supplementary Table 4 in the Appendix).

As any orthogonal rotation applied to the emotion dimensions would result into different gradients, we measured to what extent rotated solutions explained the topography of right TPJ. Therefore, we tested the correspondence between anatomical distance and the fitting of ~70,000 rotated versions of polarity, complexity and intensity (see the Appendix for a comprehensive description). Results showed that the original unrotated emotion dimensions represented the optimal solution to explain the gradient-like organization of right temporo-parietal territories (Supplementary Figure 4 in the Appendix).

Further, we performed a data-driven searchlight analysis to test whether right TPJ was the only region significantly encoding all the three emotion dimension gradients (please refer to the Appendix for details). Results obtained from the meta-analytic definition of right TPJ were confirmed using this alternative approach ($q < 0.05$ FDR corrected and cluster size > 10; CoG: $x = 58$, $y = -53$, $z = 21$; Supplementary Figure 5 in the Appendix), as no other region encoded the combination of polarity, complexity and intensity in a topographic manner.

Moreover, we conducted three separate searchlight analyses to characterize the spatial arrangement of single emotion dimension gradients (for details please see the Appendix). Polarity, complexity and intensity maps revealed specific topographies: regions as the right preCS represented the three emotion dimensions in distinct - yet adjoining - subregions, whereas the right OTS encoded overlapping gradients of complexity and intensity (Supplementary Figure 6 in the Appendix).

When we explored whether the left hemisphere homologue of TPJ (CoG: x = -59, y = -56, z = 19) showed a similar gradient-like organization, we did not find significant associations between spatial and functional characteristics either for the basic emotion model ($\varrho = 0.208$, p-value = 0.356) or the emotion dimension one ($\varrho = 0.251$, p-value = 0.144; Supplementary Table 2 in the Appendix). Specifically, neither any of the emotion dimensions (polarity: $\varrho = 0.132$, p-value = 0.354; complexity: $\varrho = 0.157$, p-value = 0.222; intensity: $\varrho = 0.149$, p-value = 0.257) nor any of the basic emotions showed a gradient-like organization in left TPJ (happiness: $\varrho = 0.158$, p-value = 0.216; fear: $\varrho = 0.142$, p-value = 0.293; sadness: $\varrho = 0.156$, p-value = 0.213; anger: $\varrho = 0.073$, p-value = 0.733; Supplementary Table 3 in the Appendix).

Lastly, as spatial smoothness of functional data and cortical folding may affect the estimation of gradients, we performed additional analyses considering the unfiltered version of group-average brain activity and obtaining a measure of the anatomical distance respectful of cortical topology. Results showed that the topographic arrangement of emotion dimensions in right temporo-parietal territories was not affected by smoothing (Supplementary Figure 7 in the Appendix) and respected the cortical folding (polarity: $\varrho = 0.248$, p-value = 0.026, CI: 0.238-0.257; complexity: $\varrho = 0.314$, p-value = 0.001, CI: 0.304-0.323; intensity: $\varrho = 0.249$, p-value = 0.013, CI: 0.239-0.258). For details about this procedure and a comprehensive description of the results please refer to the Appendix.

To summarize, polarity, complexity and intensity dimensions were highly consistent across individuals, explained the majority of the variance in behavioral ratings (85%) and were mapped in a gradient-like manner in right (but not left) TPJ. Happiness (28% of the total variance in behavioral ratings) was the only basic emotion to be consistent across subjects and to be represented in right TPJ. Importantly, though, happiness and complexity demonstrated high similarity both in behavioral ratings ($\varrho = 0.552$) and in brain activity patterns ($\varrho = 0.878$). Taken together, these pieces of evidence indicate the existence of emotion dimension gradients in right temporo-parietal territories, rather than the presence of discrete emotion topographies.

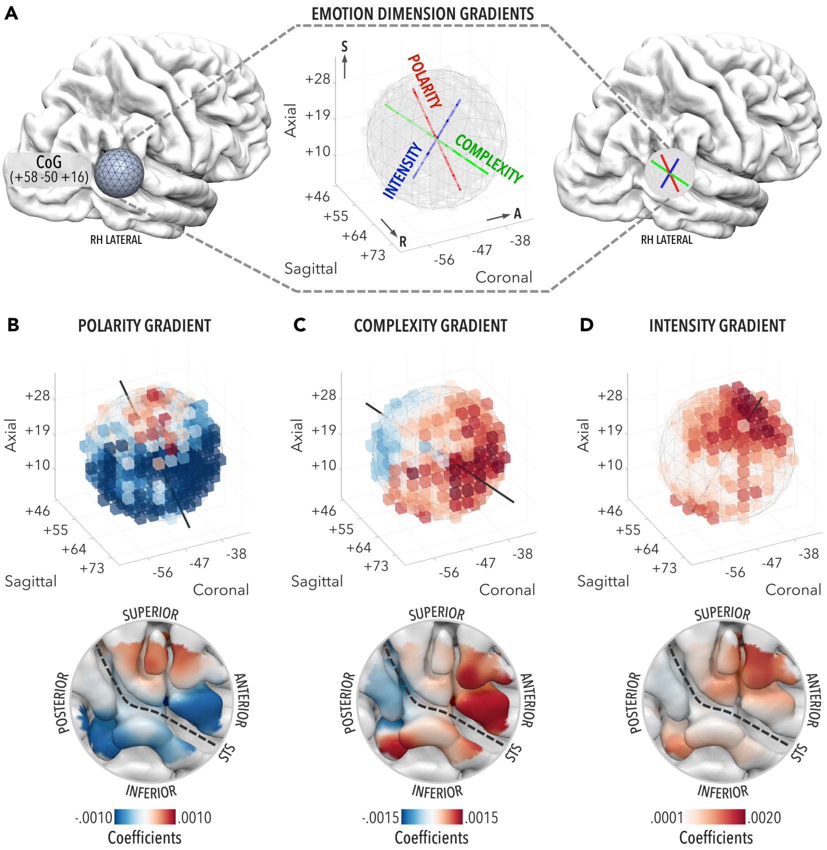


Figure 5 Emotion gradients in right TPJ - A. We revealed three orthogonal and spatially overlapping emotion dimension gradients (polarity, complexity and intensity) within a region of interest located at the reverse inference peak for the term “TPJ” (15mm radius sphere). Symmetry axis of the region of interest represents the main direction of the three gradients **B**. β coefficients of the polarity dimension are mapped through a gradient with an inferior to superior direction. **C**. β coefficients of the complexity dimension are mapped through a gradient with a posterior to anterior direction. **D**. β coefficients of the intensity dimension are mapped through a gradient with an inferior to superior direction. For single-subjects results please refer to Supplementary Figure 6 in the

Appendix. Lowermost row depicts the arrangement of the emotion dimension gradients in surface space. CoG = Center of Gravity, STS = Superior Temporal Sulcus.

3.3.5 Do emotion dimension gradients simply encode portrayed emotions?

In movie watching, actions and dialogues are not usually directed toward the observer and, the reported subjective experience is very likely influenced by character emotions, intentions and beliefs. Therefore, we tested whether the gradient-like organization of right TPJ can be explained simply considering emotions of movie characters. To this aim, we took advantage of publicly available tagging data of Forrest Gump (Labs et al., 2015), in which participants indicated the portrayed emotion of each character and whether such an emotion was directed toward the character itself (self-directed; e.g., Forrest feeling sad) or toward another one (other-directed; e.g., Forrest feeling happy for Jenny). As in Labs and colleagues, these reports constituted two third-person emotion attribution descriptions, which we used as models of the attribution of affective states to others (please refer to the Appendix for details).

On average, subjective ratings shared the 11.4% ± 8.6% of the variance with the self-directed emotion attribution model and the 35.3% ± 16.1% with the other-directed model, indicating that the subjective emotional experience could be inferred from portrayed emotions only in part. Moreover, in line with previous studies highlighting the role of right TPJ in the attribution of mental states to others (Saxe and Kanwisher, 2003; Van Overwalle, 2009; Skerry and Saxe, 2015), the other-directed emotion attribution model significantly explained activity in this region (Supplementary Figure 8 in the Appendix). However, none of the first six components obtained from the other-directed emotion attribution model (i.e., 87% of the explained variance) retained a topographic organization in right TPJ (Supplementary Table 5 in the Appendix). In addition, we used canonical correlation analysis to transform the other-directed model into the space defined by subjective emotion ratings and tested whether starting from a third-person complex description of portrayed emotions, one can fully reconstruct the brain topography of emotion dimensions.

Noteworthy, only the first aligned component was mapped in a topographic manner within right TPJ (reconstructed polarity: $\varrho = 0.221$, p-value = 0.036; reconstructed complexity: $\varrho = 0.150$, p-value = 0.384; reconstructed intensity: $\varrho = 0.207$, p-value = 0.092). This finding suggests that the information coded in the other-directed emotion attribution model is not sufficient to fully reconstruct the topography of the subjective emotional experience.

3.3.6 Characterization of emotion dimension gradients

To detail how right TPJ gradients encode perceived affective states, we have reconstructed fMRI activity for movie segments connoted by either positive or negative polarity, as well as higher or lower complexity and intensity. The orientation of the three emotion dimension gradients was represented by the symmetry axis of our region of interest. Specifically, for polarity events connoted by positive emotions increased activity in ventrorostral territories, lying close to the superior temporal sulcus, whilst highly negative events augmented hemodynamic activity in dorsocaudal portions of right TPJ, extending to the posterior banks of Jensen sulcus (Figure 6A and 4D).

Events connoted by higher complexity (e.g., concurrent presence of happiness and sadness) were associated to signal increments in rostrolateral territories of right TPJ, whereas those rated as having lower complexity (e.g., fearful events) increased hemodynamic activity in its caudal and medial part, encompassing the ascending ramus of the superior temporal sulcus (Figure 6B and 4D). Higher levels of intensity were related to increased activity in rostrodorsal and ventrocaudal territories, reaching the ascending ramus of the lateral sulcus and posterior portions of the middle temporal gyrus, respectively. On the contrary, low-intensity events augmented hemodynamic activity in a central belt region of right TPJ, located along the superior temporal sulcus (Figure 6C and 4D). Noteworthy, the orthogonal arrangement of polarity and complexity and the fact that intensity was represented both superiorly and inferiorly to the superior temporal sulcus determined that the variety of emotional states elicited by the Forrest Gump movie (see Figure 2) could be mapped within a single patch of cortex.

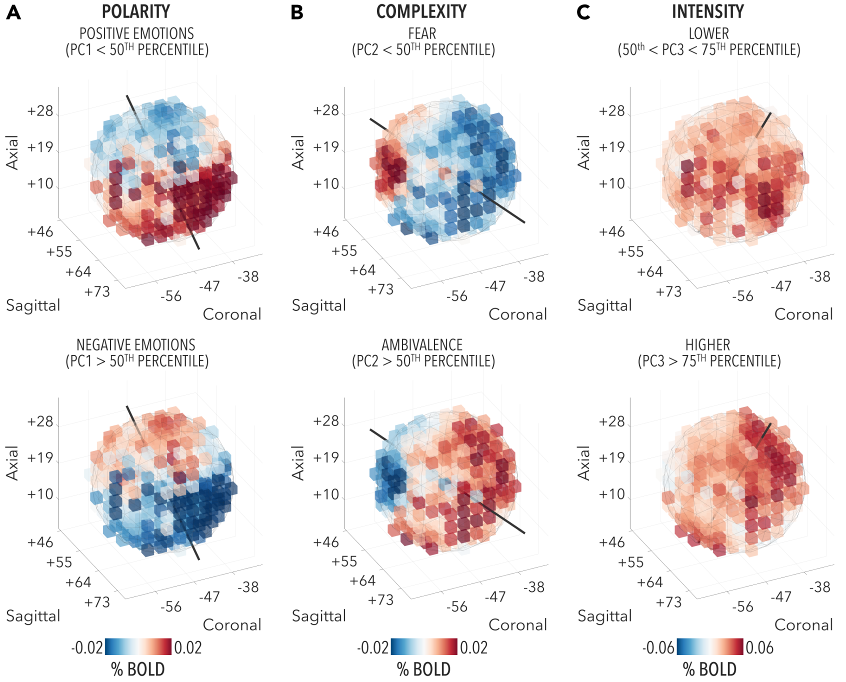


Figure 6 Characterization of Emotion Dimension Gradients in right TPJ - A. Average hemodynamic activity in right TPJ related to the scores below and above the 50th percentile for polarity. **B**. Average hemodynamic activity in right TPJ related to the scores below and above the 50th percentile for complexity. **C**. Since intensity is not bipolar as the other two components (i.e., scores ranged from ~0 to positive values only), for this dimension we mapped the average TPJ activity above the 75th percentile and within 50th and 75th percentile. PC = Principal Component.

Moreover, in sensory areas, topographies result from the maximal response of neurons to a graded stimulus feature. To parallel right TPJ emotion dimension gradients with those observed in primary sensory regions, we investigated whether distinct populations of voxels were selective for specific polarity, complexity and intensity scores. Thus, we employed the population receptive field method (Dumoulin and Wandell, 2008) to estimate the tuning curve of right TPJ voxels for each emotion dimension. The map of voxel selectivity was consistent with the topography obtained from the original gradient estimation for the three emotion dimensions (polarity: $\varrho = 0.547$, p-value = 0.001; complexity: $\varrho = 0.560$, p-value < 0.001 and intensity: $\varrho = 0.596$, p-value < 0.001). Specifically, results demonstrated the existence of four populations of voxels tuned to specific polarity values, which encoded highly and mildly positive and negative events, respectively (Figure 7A). Also, two distinct populations of voxels were tuned to maximally respond during cognitively mediated affective states (i.e., highly and mildly positive complexity values), and two other populations were selective for emotions characterized by higher and lower levels of automatic responses (i.e., highly and mildly negative complexity values; Figure 7B). Lastly, for the intensity dimension two specific populations of voxels were engaged depending on the strength of the emotional experience (Figure 7C). This further evidence favored the parallel between emotion and sensory gradients.

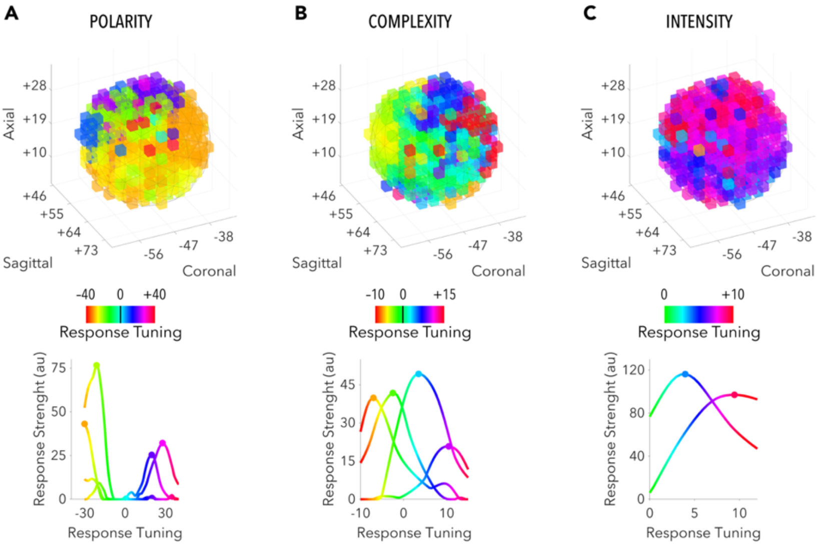


Figure 7 Population receptive field estimates in right TPJ - Response selectivity maps of right TPJ voxels for (**A**) polarity, (**B**) complexity and (**C**) intensity. Preferred responses of distinct populations of voxels were obtained using non-negative matrix factorization (Supplementary Figure 9 in the Appendix). Components explaining at least 5% of the variance were plotted as a tuning curve

(lowermost row) after averaging all the possible tuning width values for each emotion dimension score.

3.4 Discussion

Previous studies reported that activity of individual brain regions codes distinct emotion features (Vytal and Hamann, 2010) whereas others suggested that a distributed network of cortical areas conjointly interacts to represent affective states (Wager et al., 2015). Interestingly, though, the possibility that gradients may encode the emotional experience as function of either basic emotions, or emotion dimensions, has never been explored. The topological isomorphism between feature space and cortical distances has been successfully adopted to relate psychophysical characteristics of stimuli to patterns of activity in sensory regions (Sereno et al., 1995). Nonetheless, this biologically advantageous mechanism has been recently proven to lie at the basis of the cortical representation of higher-level (e.g., semantic) features as well (Harvey et al., 2013; Sha et al., 2015; Huth et al., 2016). Building upon this evidence, we tested whether different affective states could be mapped onto the cortical mantle through spatially overlapping gradients.

We demonstrated that the topography of right TPJ, a crucial cortical hub for social cognition (Saxe and Kanwisher, 2003; Kober et al., 2008; Nummenmaa et al., 2012; Morelli et al., 2012; Skerry and Saxe, 2015), is described by emotion dimensions, rather than by single basic emotions. Indeed, within this region, we discovered three orthogonal and spatially overlapping gradients encoding the polarity, complexity and intensity of the emotional experience. The peculiar arrangement of these gradients allows a gamut of emotional experiences to be represented in a single patch of cortex, including affective states perceived as pleasant, unpleasant or ambivalent, connoted by calmness or excitement and mediated by primitive reactions or mentalization processes. Therefore, TPJ organization resembles the one that can be observed in primary sensory areas, where stimulus properties are topographically arranged onto the cortical mantle, as in the case of eccentricity and polar angle in the primary visual cortex (V1), frequency in the primary auditory region (A1) and body parts in the primary somatosensory area (S1). In this regard, the evidence that emotion dimensions are encoded in a gradient-like manner supports the existence of a biologically plausible mechanism for the coding of affective states, which we named emotionotopy. Indeed, as in vision precise portions of V1 map distinct locations of the visual field, specific regions of temporo-parietal territories code unique emotional experiences. This emerged also from the analysis of response tuning, showing how within each emotional hemifield of polarity and complexity, populations of voxels code specific levels of the emotional experience.

As for the case of polar angle and eccentricity in the primary visual cortex, right TPJ emotion dimension gradients are lower-dimensional descriptions of the underlying neural activity. The retinotopic representation of azimuth and elevation in V1 overlaps with local maps of ocular dominance and orientation tuning. Therefore, multiple neural codes exist at different spatial scales and the ability to capture either global or local representations relates to the resolution of the imaging technique. Our data provide evidence for a lower-dimensional, yet biologically favorable, neural code to represent emotions in temporo-parietal regions. Considering the parallel with the organization of sensory areas, we believe that the topography of right TPJ does not prevent the existence of other neural codes, especially considering the coexistence of global and local representations and the multifaceted nature of this region.

Furthermore, the fact that subjective reports explained the activity of other cortical modules is not necessarily in contrast with the topographic organization of TPJ. In fact, as in vision a rich and complex percept relies on both the primary visual cortex to extract fundamental features and other regions to process specific stimulus properties (i.e., V5 for motion), so in emotion processing TPJ may represent an hub embedded in a distributed network of regions carrying out distinct computations.

3.4.1 Richness of the emotional experience in movie watching

In our study, we employed a naturalistic continuous stimulation paradigm since it fosters emotional contagion and empathic reactions, leading to complex subjective emotional experiences, akin to real life (Nummenmaa et al., 2012). Indeed, we found that within a 60 seconds time window, emotion transitions represented in the Forrest Gump movie are similar to those experienced in real life (Trampe et al., 2015) and are predicted by a mental model of emotion co-occurrence (Thornton and Tamir, 2017; please refer to the Appendix for details). This supports the ecological validity of our stimulus and emphasizes that movies can be successfully adopted to nurture emotional resonance (Philippot, 1993; Gross and Levenson, 1995; Schaefer et al., 2010) also in the fMRI setting.

In movie watching, actions and dialogues generally are not directed toward the observer. Thus, the subjective emotional experience results, on the one hand, from narrative choices aimed at fostering empathic responses and emotional contagion (Smith, 1995) and, on the other hand, from perspective-taking and mentalizing processes (Lombardo et al., 2010; Raz et al., 2013). The fact that character intentions and beliefs shape the subjective experience in a bystander may also explain the high between-subjects agreement in reports of experienced emotions (Philippot, 1993; Gross and Levenson, 1995; Schaefer et al., 2010). This is in line with the consistency of behavioral ratings of happiness, fear, sadness and anger present in our data. Noteworthy, surprise and disgust were not consistent across all participants and, even though this may appear as a contradiction with respect to the supposed (Mesquita and Walker, 2003) universalism of basic emotions, it should be noted that our stimulus was not specifically built to reflect the well-established definition of these six emotions. For instance, some of our subjects reported that movie scenes rated as disgusting were mainly associated to situations for which interpreting the context was necessary (e.g., the principal of the school using his power to obtain sexual favors), rather than to repulsive images. This cognitive interpretation of the basic emotion disgust was apparently not present in all the subjects, with some of them relying more on its well-established definition for their ratings. Also, the use of six distinct emotion categories offered the opportunity to compare the basic emotion model with the emotion dimension one starting from the same data. In fact, as we (and others, Smith and Ellsworth, 1985; Fontaine et al., 2007) have demonstrated, emotion dimensions can be easily derived from reports based on single emotions, whereas the opposite may not be feasible. For instance, happiness is associated to positive valence, yet the intensity of such an experience may be different depending on the context (e.g., win the lottery versus meet an old friend). Moreover, while the definition of basic emotions is common across individuals, ratings based on emotion dimensions require participants to be acquainted with the meaning of psychological constructs (e.g., dominance, Mehrabian and Russell, 1974).

Nonetheless, single basic emotions provide a coarse description of subjective experiences, since humans may perceive a complex blend of apparently conflicting emotions (Larsen et al., 2001), and affective states could emerge from psychological processes not directly reducible to single emotions (Lindquist and Barrett, 2012). Our rating model, though, does account for this possibility, since subjects were allowed to report throughout the movie the occurrence of more than one emotion, as when simultaneously experiencing happiness and sadness. Besides the divergences in literature on the precise temporal overlap in perceiving conflicting emotions (Larsen et al., 2001; Berrios et al., 2015), when subjects are free to detail their personal experience, this peculiar emotional state seems to arise.

Therefore, even though our rating model was based on six emotion categories, the possibility to report the perceived intensity of more than one emotion at a time provided a rich online description of subjects' emotional experiences. Indeed, using this procedure we identified fifteen distinct affective states elicited by the Forrest Gump movie (Figure 3), a number compatible with the one reported by previous studies aimed at investigating a wide range of emotion categories (e.g., craving, terror; Skerry and Saxe, 2015; Cowen and Keltner, 2017).

3.4.2 Polarity, Complexity and Intensity of the Emotional Experience

With respect to emotion dimensions, the components we identified were deliberately interpreted not following any known model. Nonetheless, the first dimension, polarity, mainly relates to positive against negative emotions as in valence, whereas the third one, intensity, is unipolar and mimics arousal (Russell, 1980). We considered the second component as a measure of complexity of the emotional state. Indeed, this dimension contrasts events in the movie rated as fearful, an emotion with a fast and automatic response (Adolphs, 2013), against scenes characterized by ambivalence, where cognitive processes play a significant role in generating “mixed emotions” (Russell, 2017). Even though this component does not pertain to classical emotion dimension theories, an interesting interpretation may relate complexity to the involvement of Theory of Mind (Saxe and Kanwisher, 2003) in emotion perception (Mitchell and Phillips, 2015). In addition, a recent study on the mental representation of emotions (Thornton and Tamir, 2017) described the human mind component as a cardinal dimension of the affective space. In accordance with the definition provided by the authors, this dimension maps states "[...] purely mental and human specific vs. bodily and shared with animals", which is in line with our interpretation of complexity.

3.4.3 Right Temporo-Parietal Gradients Do Not Simply Encode Portrayed Emotions

In our analyses, we used the collected behavioral ratings to explain brain activity in independent subjects. In line with previous studies (Kober et al., 2008; Kragel and LaBar, 2015), results highlighted a set of regions located mainly in the right hemisphere (Figure 4 and Supplementary Table 1 in the Appendix). Interestingly, the peak of association between emotion ratings and brain activity was located in right pSTS/TPJ. This cortical area has been consistently identified as having a central role in the attribution of mental states to others, as demonstrated by functional neuroimaging (Saxe and Kanwisher, 2003; Skerry and Saxe, 2015), noninvasive transcranial stimulation (Donaldson et al., 2015) and lesion studies (Campanella et al., 2014). In addition, this region spans across the posterior portion of the superior temporal sulcus, which is implicated in emotion perception (Kober et al., 2008; Burnett and Blakemore, 2009; Nummenmaa et al., 2012; Kragel and LaBar, 2015, 2016). In line with this, we demonstrated that activity in right TPJ is significantly explained by the process of emotion attribution to others and by the subjective emotional experience as well. Crucially, though, the information coded in the other-directed emotion attribution model is not sufficient to fully reconstruct the topography of the subjective emotional experience. In this regard, we reason that the empathic resonance and emotional contagion elicited by emotions of movie characters produced a remapping of events within a subjective framework. The final reported experience was then mapped in right temporo-parietal territories following the three cardinal axes represented by emotion dimensions. This view would reconcile previous studies demonstrating the involvement of right TPJ in the representation of subjective emotional experience (Burnett and Blakemore, 2009; Nummenmaa et al., 2012; Kragel and LaBar, 2015), in empathic processes (Morelli et al., 2012; Morelli and Lieberman, 2013), and in the attribution of beliefs and emotions to others (Saxe and Kanwisher, 2003; Van Overwalle, 2009; Skerry and Saxe, 2015).

In addition, in the current study, ratings of the emotional experience elicited by an American movie in Italian participants explained brain activity of German subjects. This suggests that the topographic representation of emotions exists regardless of linguistic or micro-cultural differences. What may depend on the cultural background of each individual is instead the mapping of distinct emotional states within these gradients. For instance, ruminative thinking, sadness and apathy characterize melancholy in the Western culture and we speculate that such an emotion would be mapped in the brain as a negative state having high complexity. However, if different levels of polarity, complexity and intensity characterize melancholy in other cultures (e.g., Eastern), this emotion would be mapped differently with respect to the three emotion dimension topographies.

Of note, the optimal description of emotion dimension gradients is obtained considering a 15mm radius sphere centered at the meta-analytic peak for the term "TPJ". This region of interest is ~42% larger in volume as compared to the reverse inference TPJ map (i.e., the likelihood that the term "TPJ" is used in a study given the presence of reported activation), but also ~32% smaller than the forward inference definition (i.e., the likelihood that a region will activate if a study uses the "TPJ" term; please refer to the Appendix for further details). Therefore, emotion dimension gradients are best represented in a patch of cortex that approximates the definition of right TPJ based on brain activation studies. Nevertheless, the fact that emotion topography generalizes across a range of spatial scales (i.e., up to 27mm radius sphere) motivated the use of "temporo-parietal territories" in describing the location and size of emotion dimension gradients.

3.4.4 Limitations

Our study presents the following limitations: first, the effect size we report for the relationship between emotion ratings and brain activity appears to be relatively small (i.e., 7% of explained variance in right TPJ). However, we would like to emphasize three aspects: (1) first, brain regions significantly encoding emotions are selected after rigorous correction for multiple comparisons; (2) second, the magnitude of the effect is in line with recent fMRI literature on the coding of emotions in the brain (Skerry and Saxe, 2015) and the evaluation of the noise ceiling suggests that our emotion dimension model explains between 30% (i.e., upper bound) and 54% (i.e., lower bound) of right TPJ activity; (3) lastly, we used a parsimonious encoding model, in which only six predictors explained 3,595 samplings of brain activity.

Second, although using a larger set of emotion categories the same polarity, complexity and intensity dimensions still emerged (see the Appendix for details), we would like to emphasize that the collected emotion ratings are specific for the present stimulus. Therefore, alternative movies depicting horror, or sexual desire (Cowen and Keltner, 2017) may produce different emotion dimensions. Indeed, independent studies are required to prove the generalizability of the topographic organization of polarity, complexity and intensity dimensions within right TPJ.

Lastly, it would be interesting to collect psychophysiological data as skin conductance and heart rate, while subjects watch the movie and provide emotion ratings. Future studies might investigate the association between these parameters and the emotional experience.

In summary, our results showed that moment-by-moment ratings of perceived emotions explain brain activity recorded in independent subjects. Most importantly, we demonstrated the existence of orthogonal and spatially overlapping right temporo-parietal gradients encoding emotion dimensions, a mechanism that we named emotionotopy.

4. Emotional experience timecourse explains brain connectivity dynamics during naturalistic stimulation

4.1 Introduction

Emotions have often been considered as an intense and brief reaction of the body and the mind in response to a precise external or internal event (Frijda, 1993; Ekman and Davidson, 1994; Lewis et al., 2010). Differently from mood, which is not related to a specific experience and can be continuously present (Parkinson et al., 1996; Lischetzke, 2014), emotions can rapidly change over time (Frijda, 2017). Specifically, affective states are the consequence of an active interplay between an individual and the environment and their duration is likely to be affected by several determinants (see Waugh et al., 2015 and Verduyn et al., 2015 for a review). Over the past years, increasing attention has been oriented to the study of the phasic nature of emotions, describing them as dynamic states that unfold over time (Scherer, 2009; Barrett, 2013; Waugh et al., 2015), and not as bounded static phenomena.

Therefore, different models have been proposed to delineate the temporal characteristics of emotions (e.g., duration, onset, resurgence) and how these are represented in the brain, with the dynamics between specific regions related to different emotional experiences (Raz et al., 2012, 2013, 2016; Waugh et al., 2015; Résibois et al., 2017). A prominent theory postulates that affective states emerge from a constructive process involving several components interacting, including physiological, cognitive, social and perceptual ones (Barrett, 2006, 2012). According to this approach, the activity and interplay of a network of brain areas would be necessary to support and give rise to the gamut of emotional experiences found in real life (Barrett, 2013; Wilson-Mendenhall et al., 2015). In this regard, previous studies have indeed identified a set of regions consistently involved in the experience of affective states (Kober et al., 2008; Lindquist et al., 2012; Lindquist and Barrett, 2012). Interestingly, it has also been described a so called "extended social-affective default mode network", related to socio-affective and self- and other- related mental processes and involving regions as the temporo-parietal junction, the dorsomedial prefrontal cortex and the middle temporal gyrus among others, already known to be implicated in mentalizing and empathic processing (Amft et al., 2014).

However, the majority of neuroimaging studies so far have been employing brief or static stimuli to investigate emotion perception, which do not take into account the dynamism of affective states and their unfolding over time (e.g., Posner et al., 2009; Baucom et al., 2012; Kim et al., 2017). Using pictures or short videos, previous researches captured the peak of an affective state, missing the following cascade of events characterizing the emotional experience (Waugh and Schirillo, 2012).

On this matter, the use of movies in the functional magnetic resonance imaging (fMRI) setting could provide valuable insights. Indeed, it has been shown that movies are ecological stimuli, able to elicit a wide variety of emotional states akin to real life (Philippot, 1993; Gross and Levenson, 1995; Schaefer et al., 2010; Lettieri et al., 2019). Furthermore, they would allow to follow for a long period of time the dynamic interaction between different brain areas over the course of emotional responses.

Of note, a recent work employing brief movie exerts showed that the functional connectivity between the salience and amygdala-based networks is associated to the perceived intensity of sadness, fear and anger (Raz et al., 2016). Nevertheless, a small set of negative emotions is far from capturing the complexity of subjective emotional experiences.

Here, we aimed to overcome this limitation and explore the dynamic interplay between different brain regions throughout naturalistic situations. To do so, we tracked changes in functional connectivity among distinct brain regions during the watching of an emotionally charged movie. Continuous behavioral ratings of the perceived intensity of emotional states in an independent sample of subjects were then related to functional connectivity dynamics.

4.2 Materials and Methods

To explore the dynamic dialogue between brain regions during an emotionally charged movie, we employed the same behavioral and fMRI data collected in the first study. Starting from these, we investigated whether network properties related to functional connectivity data encode variations in the polarity, complexity and intensity of the subjective emotional experience.

4.2.1 Behavioral processing

First, we used the moment-by-moment emotion ratings obtained during the Forrest Gump movie to derive the 3 emotion dimensions of polarity, complexity and intensity (i.e., PCA analysis, Figure 2). Following the same procedure employed in the previous study, we downsampled, lagged and temporally smoothed the emotion dimension data using a moving average procedure (50 timepoints window).
The obtained timeseries expressed changes in the polarity, complexity and intensity of the emotional experience, that we then correlated with brain functional connectivity dynamics.

4.2.2 fMRI processing and connectivity dynamics estimation

In this study, fMRI data were preprocessed exactly as described in the previous chapter. To select the regions of interest for the estimation of functional connectivity dynamics, we employed a voxel-wise encoding procedure (Naselaris et al., 2011) in which the emotion ratings obtained from the behavioral experiment were used as predictors of brain activity in independent subjects exposed to the same movie (Hanke et al., 2016). The model significantly explained activity in right inferior frontal gyrus (IFG), rostral middle frontal gyrus (rMFG), medial superior frontal gyrus (mSFG), occipitotemporal sulcus (OTS), precentral sulcus (preCS), posterior part of the superior temporal sulcus/temporoparietal junction (pSTS/TPJ), middle occipital gyrus (MOG) and posterior middle temporal gyrus (pMTG) and in the left supramarginal gyrus (SMG) and pMTG ($q < 0.01$ FDR corrected and cluster size > 10; Figure 4A, 8).

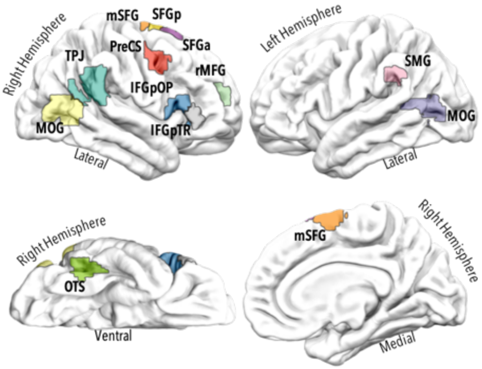


Figure 8 Voxelwise encoding – Results of the Voxelwise encoding procedure. TPJ = temporo-parietal junction, MOG = middle occipital gyrus, IFGpOP = inferior frontal gyrus pars opercularis, IFGpTR = inferior frontal gyrus pars triangularis, rMFG = rostral middle frontal gyrus, SFGa = anterior superior frontal gyrus, SFGp = posterior superior frontal gyrus, SMG = supramarginal gyrus.

We thus investigated if changes in functional connectivity dynamics of these 12 brain regions were associated to the three emotion dimensions.

Functional connectivity dynamics (FCD) is a novel metric useful to capture brain network reconfigurations through time (Hansen et al., 2015). Therefore, it reflects the time-varying dialogue between different brain regions.

First, using the same 50 timepoint-sliding windows employed for the behavioral data, we assessed the correlation (i.e., FCD) among all voxels pairings within the 12 region network. We then measured the association of polarity, complexity and intensity with FCD. We obtained 3 correlation matrices depicting how much each emotion dimension explained voxels interactions in time. Each matrix was then thresholded at $p < 0.05$ using the fitting between FCD and surrogate emotion dimension models (IAAFT) as null distribution. To assess whether above-threshold interactions occurred in specific ROIs pairings, we divided each correlation matrix in 78 sections representing the connectivity among the 12 ROIs and counted the number of significant occurrences for the whole matrix and each section (Danti et al., 2018). For each surrogate matrix, we selected the same overall number of interactions among the highest and counted their occurrence in each section. Thus, we obtained for each portion of the 3 matrices the number of interactions under the null hypothesis and tested the significance of above-threshold occurrences. Lastly, the exact p-value was estimated using a tail approximation method (Winkler et al., 2016) and corrected for multiple comparisons using Bonferroni (Figure 9).

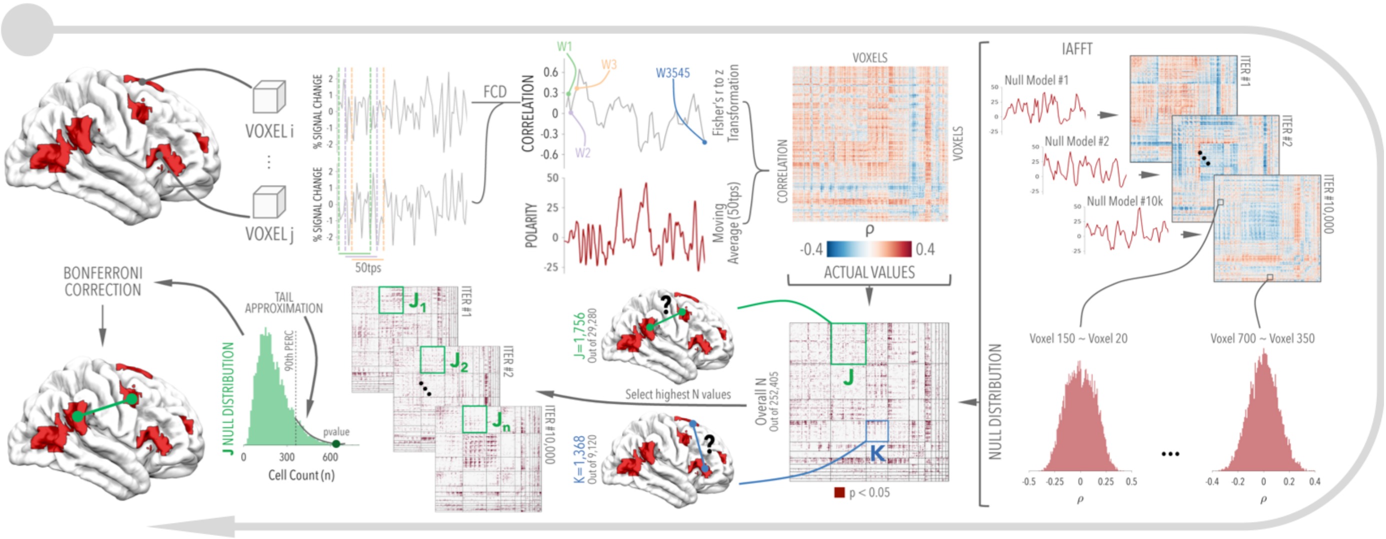


Figure 9 Connectivity dynamics estimation - Analysis pipeline to test whether the timecourse of emotion dimensions explain changes in connectivity among regions encoding emotion ratings. Details of the procedure are provided in the Methods section. FCD = functional connectivity dynamics; IAFFT = iterative amplitude adjusted Fourier transformed procedure.

4.3 Results

We found that connectivity of the right precentral sulcus (preCS) was significantly modulated by the timecourse of the polarity, complexity and intensity of the emotional experience.
Specifically, polarity modulated the coupling between the right preCS and the right occipito-temporal sulcus (OTS; Figure 10).

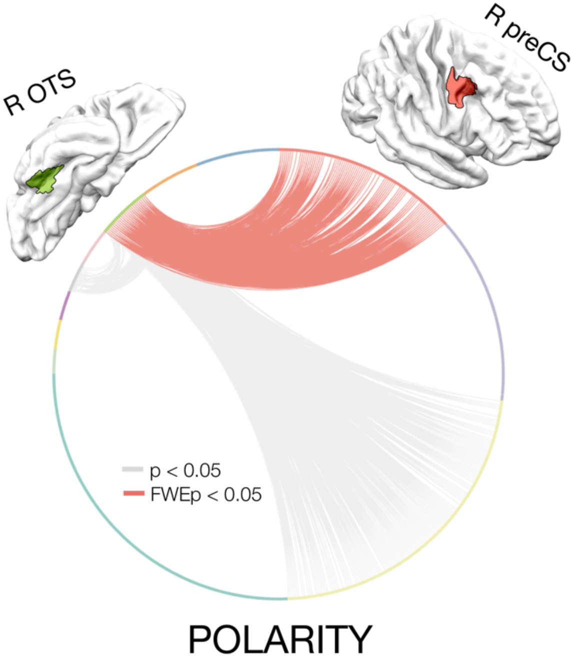


Figure 10 Functional connectivity dynamics modulated by polarity – Polarity modulated the coupling between the right preCS and the right occipito-temporal sulcus (OTS). Results corrected for multiple comparisons (FWEp < 0.05).

On the other hand, intensity modulated the interaction of the right preCS with the right medial superior frontal gyrus (mSFG; Figure 11).

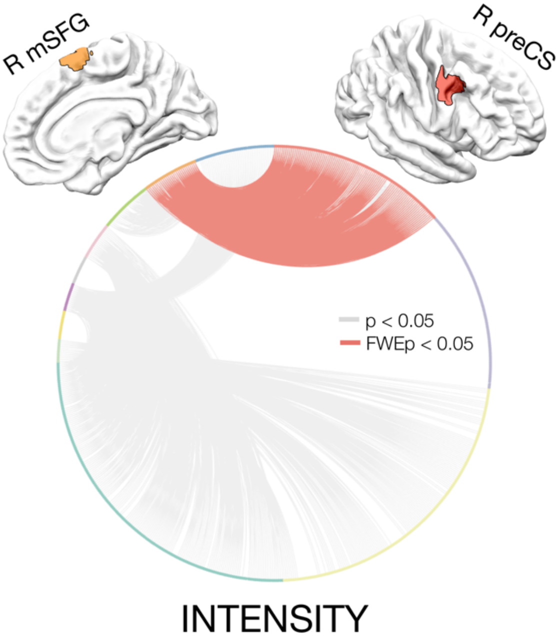


Figure 11 Functional connectivity dynamics modulated by intensity – Intensity modulated the interaction of the right preCS with the right medial superior frontal gyrus (mSFG). Results corrected for multiple comparisons (FWEp < 0.05).

Lastly, complexity regulated the intrinsic connectivity of the right preCS (Figure 12).

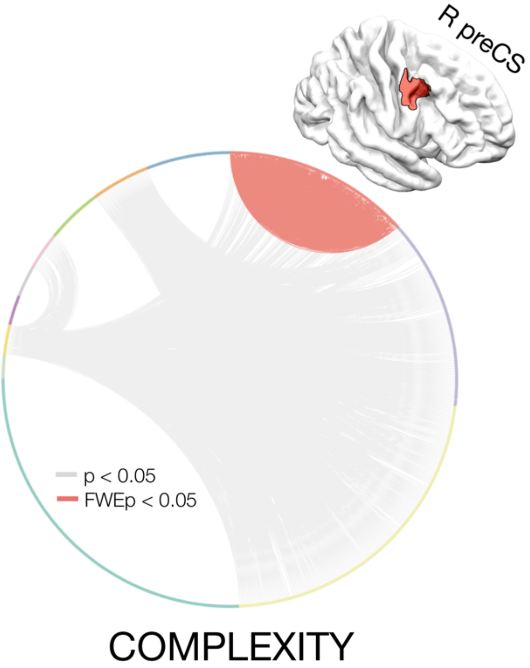


Figure 12 Functional connectivity dynamics modulated by complexity – Complexity regulated the intrinsic connectivity of the right preCS. Results corrected for multiple comparisons (FWEp < 0.05).

Furthermore, we also observed the contribution of each emotion dimension to the hemodynamic activity of brain regions. We found a peculiar organization of the right preCS, where separate yet adjoining subregions code distinct emotion dimensions (Figure 13). Indeed, as depicted in Figure 13, we can identify a posterior part of the sulcus being associated to intensity, a middle one to complexity and an anterior portion related to polarity.

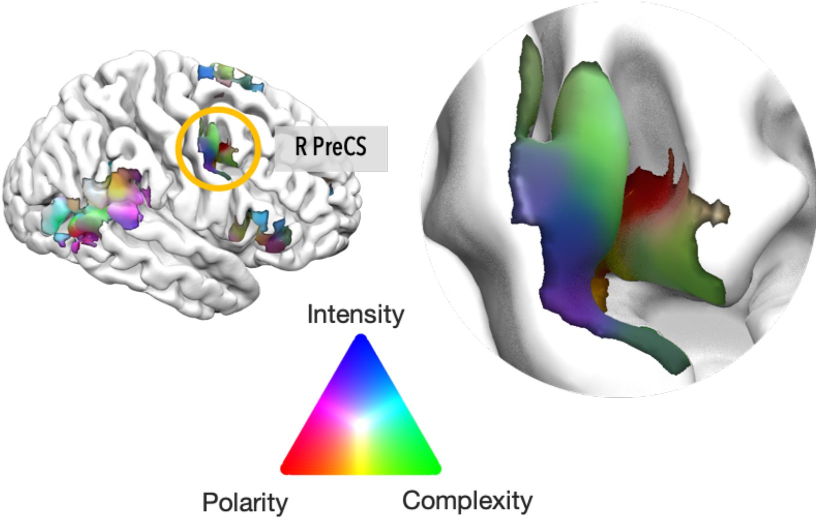


Figure 13 Emotion dimension organization in the preCS – The right preCS shows a peculiar organization in relation to each emotion dimension. The picture depicts the individual contribution of polarity, complexity and intensity to the hemodynamic activity of this region. We found a posterior part of the sulcus associated to intensity, a middle one to complexity and an anterior portion related to polarity.

Lastly, using a more liberal threshold ($p < 0.05$ uncorrected), it is interesting to observe the functional connectivity dynamics of other brain regions being regulated by the three emotion dimensions.

We can find that polarity modulated the functional connectivity dynamics of the right IFGpTR and the right OTS, together with the coupling between the right OTS and the right MOG (Figure 10). On the other hand, intensity modulated the dialogue between the left SMG and the right TPJ and between the right OTS and the right mSFG. Furthermore, it also influenced the connectivity dynamics of the right SFGp with the right IFGpTR, and of the right rMFG with the right MOG (Figure 11). Complexity instead modulated the dialogue between the right mSFG with the bilateral MOG and the left SMG, and between the mSFG with the left MOG (Figure 12).

4.4 Discussion

Emotions are complex descriptions of events that unfold over time (Scherer, 2009; Waugh et al., 2015). Therefore, to properly study how emotions are represented in the brain, it is necessary to consider both their complexity and their dynamic nature at the same time. In our study, we employed an emotionally charged movie able to elicit affective experiences akin to real life (please refer to the Appendix for further details), allowing us to investigate the complex inner cascade of events associated to it. Moreover, we tracked changes in functional connectivity among distinct brain regions during the watching of the same movie in an independent sample of subjects (Hanke et al., 2016). Therefore, the methodology we adopted allowed us to take into consideration at the same time the complexity and the dynamics of the subjective emotional experience. Using this approach, we demonstrated that the functional connectivity dynamics of the right preCS is significantly modulated by changes in three cardinal emotion dimensions.

More precisely, we found that variations in the polarity dimension modulated the coupling between preCS and a face-selective area in OTS. Indeed, the latter is a brain region part of the face-processing network in the occipital lobe (Jonas et al., 2016). Therefore, the functional connectivity between the preCS and OTS might be related to the evaluation of emotional facial expressions, likely suggesting the importance of processing them to determine positive and negative affective states (Radua et al., 2014). In this regard, as it has been described in the first chapter of this thesis, to understand other individuals, we are used to decipher a variety of social signals coming from them. Surely, one of the most useful signal representatives of affective states is manifested through facial activity. Facial expressions are an easy and almost immediate way to communicate perceived emotions and are universally shared, at least the ones associated to basic affective states (Ekman and Friesen, 1971; Ekman, 1989; 1999). The interpretation and understanding of facial expressions play a fundamental role in responding to emotional situations and contributes to generate the inner affective experience.

We also observed that changes in the intensity of the emotional states modulated the functional connectivity dynamics of the right preCS and the right mSFG, a crucial area for emotion regulation

(Billot et al., 2017). In this regard, affective states can have severe consequences for an individual when they occur at an inappropriate time or intensity level. Therefore, the ability to regulate these responses is of fundamental importance for the good functioning of an individual in daily life. Moreover, it has also been found that the mSFG is involved in the perception of fearful and painful facial and bodily expressions (Grézes et al., 2007; Budell et al., 2015). Therefore, this evidence is in line with changes in the intensity of perceived emotions to modulate the coupling between the right preCS and the right mSFG.

Lastly, we found that the intensity of the emotional experience regulated the intrinsic connectivity of the right preCS. It is interesting to also note the peculiar organization of this region, where we can observe separate yet adjoining subregions coding distinct emotion dimensions. This spatial arrangement is likely to be at the basis of our findings. Further, the right preCS has been involved in the perception of dynamic facial expressions and their affective labelling (Foley et al., 2012; Telzer et al., 2014), and also during the viewing of emotionally engaging short movies (Hutcherson et al., 2005).

Moreover, the central role of the precentral sulcus highlighted by our results, might reside in the importance of this region for the mirror neuron system (Rizzolatti and Craighero, 2004; Rizzolatti et al., 2009). Indeed, as it has been discussed in the Introduction section of this dissertation, there are relevant findings relating emotional responses to mirror mechanisms (for a review see Gallese, 2001 and Iacoboni, 2009). Several neuroimaging studies support the existence of a "mirror mechanism" as the basis of socio-affective abilities. In line with this, the experience of empathy would originate through a simulation of the mental states of other people, made possible thanks to the mirror neurons system (Rizzolatti and Caruana, 2017; Iacoboni, 2009). According to this view, the mirror system supports the representation and imitation of emotional expressions, empathy and empathic pain, among other functions. In light of this perspective, the precentral sulcus would be a core region fundamental for the generation of mirror mechanisms lying at the basis of emotional responses.

Of note, canonical fMRI studies are used to dissect the phenomenon of interest in its very basic components to actually reduce its complexity and control possible intervening effects (see for instance Simony et al., 2016). This approach is effective in understanding isolated processes but has some limitations in drawing conclusions regarding global aspects of perception and cognition, limiting de facto our understanding of brain functioning in real-life situations (e.g., Felsen and Dan, 2005). Conversely, the use of naturalistic and dynamic stimuli, such as movies (Betti et al., 2013; Lettieri et al., 2019) and video-clips of everyday activities (Zacks et al., 2011), in the fMRI experimental setting can provide a characterization of the dynamic interplay among brain regions that likely occur in ecological situations. In this regard, it seems that dynamic changes in brain functional connectivity are associated to distinct behavioral performances and track variations occurring in the surrounding environment (Sadaghiani et al., 2015). The use of naturalistic stimulation to track dynamic changes in the brain functional connectivity is a novel approach able to introduce a holistic perspective that can pave the way to new and valuable insights (Bullmore and Sporns, 2009; Baldassano et al., 2017).

In this regard, we demonstrated that it is possible to track the functional connectivity dynamics of brain regions during a naturalistic stimulation. Moreover, we employed continuous behavioral ratings expressing a wide gamut of affective states. This approach showed that, during naturalistic stimulation, the timecourse of the emotional experience can be described by polarity, complexity and intensity dimensions. Importantly, the perceived affective states tracked by changes in the three components explained connectivity dynamics of right preCS, a region processing social interactions and emotional situations (Hutcherson et al., 2005; Foley et al., 2012; Telzer et al., 2014).

5. Conclusions

We demonstrated that the topography of right TPJ, a crucial cortical hub for social cognition (Saxe and Kanwisher, 2003; Kober et al., 2008; Nummenmaa et al., 2012; Morelli et al., 2012; Skerry and Saxe, 2015), is described by emotion dimensions, rather than by single basic emotions. Indeed, within this region, we discovered three orthogonal and spatially overlapping gradients encoding the polarity, complexity and intensity of the emotional experience. The peculiar arrangement of these gradients allows a gamut of emotional experiences to be represented in a single patch of cortex, including affective states perceived as pleasant, unpleasant or ambivalent, connoted by calmness or excitement and mediated by primitive reactions or mentalization processes. Therefore, TPJ organization resembles the one that can be observed in primary sensory areas, where stimulus properties are topographically arranged onto the cortical mantle, as in the case of eccentricity and polar angle in the primary visual cortex (V1), frequency in the primary auditory region (A1) and body parts in the primary somatosensory area (S1). In this regard, the evidence that emotion dimensions are encoded in a gradient-like manner supports the existence of a biologically plausible mechanism for the coding of affective states, which we named emotionotopy. Indeed, as in vision precise portions of V1 map distinct locations of the visual field, specific regions of temporo-parietal territories code unique emotional experiences. This emerged also from the analysis of response tuning, showing how within each emotional hemifield of polarity and complexity, populations of voxels code specific levels of the emotional experience.

Furthermore, it is important to highlight the existence of cortical neurodegenerative diseases as the behavioral variant of the frontotemporal dementia, which present neuroanatomical damages to structures of the Theory of Mind and emotion processing network, with posterior regions as TPJ and pSTS significantly affected by the degeneration processes. As a consequence, these patients show several behavioral disturbances, particularly involving the social cognition domain. Indeed, they usually report a significant impairment in the recognition and attribution of emotions, in empathizing abilities and in the affective component of the Theory of Mind (Kipps et al., 2009; Poletti et al., 2012). The pSTS/TPJ region, therefore, seems to play a fundamental role in the integration of these capacities, as testified also by lesion and noninvasive stimulation studies (Van Overwalle, 2009; Campanella et al., 2014; Donaldson et al., 2015). In line with this, the results presented under the "Emotionotopy in the human right temporo-parietal cortex" chapter, fit well with findings coming from neuropsychological studies and provide a more detailed description of the role of the TPJ region in the processing of affective states.

Lastly, the fact that subjective reports explained the activity of other cortical modules is not necessarily in contrast with the topographic organization of TPJ. In fact, as in vision a rich and complex percept relies on both the primary visual cortex to extract fundamental features and other regions to process specific stimulus properties (i.e., V5 for motion), so in emotion processing TPJ may represent an hub embedded in a distributed network of regions carrying out distinct computations.

Indeed, we also found that the perceived affective states tracked by changes in the polarity, complexity and intensity of the emotional experience, explain the connectivity dynamics of the right preCS, a region processing social interactions and emotional situations (Hutcherson et al., 2005; Foley et al., 2012; Telzer et al., 2014). The right preCS acts as a distinct hub in this regard, showing also a peculiar spatial arrangement related to the three emotion dimensions.

6. Appendix

Emotionotopy in the human right temporo-parietal cortex

Generalization of emotion dimensions to a wider sample of emotion categories

To test whether our *emotion dimensions* depended on behavioral ratings of six basic emotions, we verified their existence starting from a comprehensive set of emotion categories, which included secondary affective states as well. We employed Labs and colleagues' ratings (2015) describing portrayed emotions of Forrest Gump characters considering embedded affective states (i.e., *other-directed* emotions). Participants tagged 205 randomly presented movie segments choosing among a large set of emotion categories (Ortony et al., 1990; N=22) and were allowed to watch each scene more than once. Other than *happiness*, *fear*, *sadness* and *anger*, these 22 emotion categories included secondary and social states as *admiration*, *contempt*, *gratitude*, *hate*, *love*, and *pride* among others (for a complete description please refer to Labs et al., 2015). Thus, we applied PCA to Labs data after lagging and temporally smoothing the 22 emotion timeseries, as we did for the subjective emotion ratings. The first six dimensions (~85% of explained variance) were selected to match the dimensionality of our emotion rating model and were transformed by rotating PC scores using the procrustes criterion. Results of this procedure are presented in Supplementary Figure 11, in which factor loadings of *polarity*, *complexity* and *intensity* dimensions (panel A) can be compared with the unrotated (panel B) and rotated (panel C) version of Labs and colleagues' PCs. The first three rotated components represented respectively the 20.6%, 19.8% and 16.6% of the explained variance, and were positively associated with our three *emotion dimensions* (panel E). Correlation for rotated PC_1 versus *polarity* was Spearman's $\varrho = 0.589$, for rotated PC_2 versus *complexity* was Spearman's $\varrho = 0.533$ and for rotated PC_3 versus *intensity* was Spearman's $\varrho = 0.488$.

It is important to note that, other than basic emotions (i.e., *happiness*, *fear*, *sadness* and *anger*), only four secondary/social

affective states - i.e., *love, contempt, admiration* and *gloating* - substantially contributed to the first six components derived from Labs and colleagues data, even considering the unrotated version (panel B). Indeed, the majority of emotional episodes involved the five categories of *anger*, *fear*, *happiness*, *love* and *sadness*, whereas other secondary/social categories available to subjects (e.g., *resentment, gratification, satisfaction*) were used infrequently or employed only by a subset of observers (Labs et al., 2015).
In summary, the same *polarity*, *complexity* and *intensity* dimensions emerge even when a broader set of emotion categories are used.

Does Forrest Gump reflect real life emotion dynamics?

Forrest Gump is an emotionally evocative movie that elicits a variety of affective states in a relatively short amount of time. Although movies have been successfully used to study emotions in the laboratory setting (Philippot, 1993; Gross and Levenson, 1995; Schaefer et al., 2010), we cannot exclude that the dynamics of portrayed emotions mimic those experienced in the real life. To explore this possibility, we took advantage of Thornton & Tamir 2017 experience-sampling dataset (i.e., Study 3, see also Trampe et al., 2015) comprising ~65,000 ratings obtained from ~10,000 participants, who were asked to report their own emotional state throughout the day, choosing among 18 categories (i.e., *alertness, amusement, awe, gratitude, hope, joy, love, pride, satisfaction, anger, anxiety, contempt, disgust, embarrassment, fear, guilt, offense* and *sadness*). In this study, the authors used the collected reports to build an experience-based description of emotion transitions (i.e., *real life emotion transitions*). Specifically, by considering each reported emotion and the one following in time, they tested whether the co-occurrence of emotions is predicted by a mental representation of emotion transitions (for further details please refer to Thornton and Tamir, 2017). We particularly selected the model based on study 3, as nine out of 18 emotion categories included in this dataset (i.e., *anger, sadness, fear, contempt, satisfaction, gratitude, hope, love* and *pride*) were also adopted by Labs and colleagues 2015 to label portrayed emotions in Forrest Gump. Starting from these data, we thoroughly followed the methods reported in Thornton and Tamir 2017, and converted ratings into discrete outcomes (i.e., emotion present or not) for each timepoint. We then built a transition count matrix by measuring the number of transitions between all possible emotion pairings in adjacent timepoints (i.e., between t and t+1). This matrix was further normalized by frequency-based expectations obtaining the odds of each transition. The log-transformed version of this matrix (i.e., *movie emotion transitions*) was then compared to real-life data using Spearman's ρ. To assess the statistical significance of this association, we generated surrogate timeseries for the nine emotion categories through the IAAFT procedure (N = 1,000; see *Methods* for details). For each of the 1,000 null models, a transition count matrix was then obtained, normalized and log-transformed. The obtained matrices were correlated with real-life data, generating a null distribution against which the actual association between *movie* and *real life emotion transitions* was tested.

Results showed that emotion transitions obtained from movie and real-life data were significantly associated (Spearman's ρ = 0.646; p = 0.001; Supplementary Figure 12). In addition, as this analysis explores the similarity between *movie* and *real life* data in a short time window (2s), we also evaluated whether this relationship exists at different time scales. Therefore, we built a number of *movie-based* models, each measuring the likelihood of emotion transitions between timepoint *t* and timepoint *t+n* in the future, with a maximum delay of 120 seconds (60 timepoints). These models were then correlated with real-life data and statistical significance was assessed using the procedure described above. Results are reported in panel D of Supplementary Figure 12 and show that the *real life model* predicts emotion transitions in the movie up to 58 seconds.

Of note, *happiness* is one of the emotion categories most present in Forrest Gump tagging data, yet it has not been used in reports collected for study 3. Hence, we decided to include this emotion in the *movie* model using *joy*, *awe* or *amusement* as its counterpart in the *real life* model. This allowed us to estimate the robustness of the association between movie and real-life data considering different facets of the basic emotion *happiness*.

Interestingly, using *joy, awe* or *amusement* as proxies of *happiness,* the association between *movie* and *real life emotion transitions* is significant (*joy*: Spearman's ρ = 0.702; p = 0.001; *awe*: Spearman's ρ = 0.702; p = 0.001; *amusement*: Spearman's ρ = 0.686; p = 0.001). In all these three cases, emotion transitions observed in real-life data predict those occurring in the movie up to 64 seconds in the future. Altogether, these analyses show that within a ~60 seconds time window our stimulus reflects emotion transitions similar to those experienced in real life and predicted by a mental model of emotion co-occurrence. These findings substantiate the ecological validity of our stimulus.

Searchlight analysis

We performed a data-driven searchlight analysis to test whether right TPJ was the only region significantly encoding all the three emotion dimension gradients. Thus, for each voxel significantly associated to emotion ratings (i.e., shaded and outlined regions in Supplementary Figure 5) we built a spherical region of interest (i.e., searchlight; 15mm radius) and derived the Euclidean distance of voxel coordinates and of β coefficients related to the fitting of the three *emotion dimensions*. Functional and anatomical dissimilarity matrices were then compared using Spearman's ϱ coefficient and the computation of p-value was based on surrogate data (i.e., 1,000 IAAFT-based null models). Results were corrected using the False Discovery Rate procedure and minimum cluster size > 10. The combination of the three *emotion dimension gradients* was represented within a patch of cortex centered in right pSTS/TPJ only (Supplementary Figure 5; red-colored regions; $q < 0.05$ FDR corrected; CoG: $x = 58$, $y = -53$, $z = 21$). This evidence corroborated the original findings based on the hypothesis-driven approach (i.e., NeuroSynth "TPJ").

Furthermore, we searched for individual *emotion dimension* topographies in regions encoding the emotion rating model. To do this, we ran three separate searchlight analyses measuring the topographic arrangement of *polarity*, *complexity* and *intensity*. The resulting maps were then combined into a comprehensive description of the distribution of gradients across the brain. Briefly, we employed a specific coding in the RGB color space (Handjaras et al., 2015). The red channel was assigned to *polarity*, the green to *complexity* and the blue to *intensity*. Color brightness relates to the log transformed p-value of the fitting of each component. This procedure highlighted regions predominantly involved either in *polarity*, *complexity* or *intensity*, as well as in any combination of the three (Supplementary Figure 6). Results showed that right pSTS/TPJ region is the only area of overlap of the three *emotion dimension* gradients even considering *polarity*, *complexity* and *intensity* separately.

Effect size and noise-ceiling estimation

To evaluate the effect size of the association between emotion ratings and right TPJ activity, we correlated the predicted fMRI signal obtained from the encoding procedure, with the actual BOLD activity within the same peak voxel (i.e., $R^2 = 0.07$). The association between the two timeseries was Spearman's $\varrho = 0.23$ and Kendall's $\tau = 0.15$.

To allow a direct and unbiased comparison between R^2 values obtained from the fitting of different emotion models in right TPJ, we also performed a cross-validation using a half-run split method (Supplementary Figure 8). Specifically, we randomly selected one of the two halves as the training data for the estimation of β coefficients. We then measured the goodness of fit of our model by multiplying the predictors of the remaining half with estimated β coefficients, thus reconstructing the predicted fMRI signal. The latter was then correlated with the actual fMRI activity, obtaining the final cross-validated R^2 coefficient. To avoid possible confounds introduced by selecting the first or the second part of each run as training/test dataset, we repeated the same procedure 200 times (i.e., bootstrapping), each one randomly assigning the first or second half to the training/test set. The use of this procedure resulted in an effect size of $R^2 = 0.04$ for the right TPJ peak.

Moreover, we conducted a noise-ceiling analysis for right TPJ data, similarly to what has been done by Ejaz and colleagues, 2015. For each right TPJ voxel, we calculated the average association (i.e., R^2 value) between single-subject timeseries and group-level activity. This procedure considers group-level fMRI data as the ground-truth model. However, the average signal is biased as it includes single-subject information from all the enrolled participants, ultimately producing an overestimate of the actual noise-ceiling level (i.e., the upper bound). Therefore, to obtain an estimate of the lower bound of noise-ceiling, we iteratively measured the association between each individual timeseries and the group-level average signal obtained from all the other participants (i.e., leave-one-subject-out procedure). We found that lower and upper noise ceiling bounds of the right TPJ peak voxel were $R^2 = 0.13$ (Spearman's $\varrho = 0.33$ and Kendall's $\tau = 0.22$) and $R^2 = 0.23$ (Spearman's $\varrho = 0.45$ and Kendall's $\tau = 0.31$), respectively.

Rotation of the emotion dimension model

We developed a novel approach to test the correspondence between anatomo-functional gradients and PC rotations, to reveal which stimulus features are actually encoded onto the cortical mantle (Huth et al., 2016).

First, we restricted our analysis to the three *emotion dimensions* consistent across subjects (i.e., *polarity*, *complexity* and *intensity*), which showed a gradient-like organization in right TPJ as well. Second, we performed only orthogonal rotations because of two reasons: (1) any orthogonal rotation of the original components will explain the same total variance; (2) the computation of gradient direction requires the accurate estimate of β coefficients obtained from a multiple linear regression analysis. This approach is however not robust if predictors are collinear, which may be the case when oblique rotations are applied. Therefore, we first estimated all the possible elemental rotations along the axes defined by the three *emotion dimensions* (i.e., x: *polarity*, y: *complexity* and z: *intensity*). We explored rotations between ±45° with 1° step, as this range ensured univocal solutions that would not produce the shifting of PC labels. As a matter of fact, considering a convenient bi-dimensional example, we can assert that 60° orthogonal rotations for PC_1 and PC_2 would produce solutions in which PC_1 approximates the unrotated version of PC_2 and PC_2 resembles the 180°-rotated (i.e., flipped) version of PC_1. Such a solution, though, would be identical to a 30° rotation, except for the PC sign. In line with this, rotations of ±90° would simply shift PC labels (e.g., rotated *complexity* would become now unrotated *intensity*), whereas ±180° rotations would result in sign flipping. The latter case leads to brain activity estimates (i.e., β values) being the topographically mirrored version of those obtained using the unrotated dimensions and, thus, to ρ values of the same magnitude for the association between anatomical and functional distance. As all the possible rotations between ±45°

produce ~750k solutions - which is already computationally intense -, we uniformly sampled 70k rotations from the original space. Further, the intuitive mapping of gradient magnitude (i.e., Spearman's ρ between anatomical and functional distance) in the manifold defined by the rotated solutions is non trivial and the method we propose is illustrated in Supplementary Figure 4A.

In brief, we represented gradient intensity of the unrotated *emotion dimensions* as the central point of a 3D manifold described by all the ±45° explored rotations. We also mapped gradient intensity of all the rotated solutions as points in this space, color-coding the magnitude of the association between anatomical and functional distance. Rotations are expressed according to three cardinal trajectories originating from the central point (i.e., the unrotated *emotion dimensions*), each one determining the orthogonal rotation of two components while maintaining fixed the other one. Therefore, points lying on the red trajectory depict solutions in which the original unrotated version of *polarity* is present and *complexity* and *intensity* are actually rotated. The same applies also to the green and blue trajectories in which *complexity* and *intensity* respectively maintain their original unrotated form. All the other mapped solutions describe orthogonal rotations concurrently applied to the three *emotion dimensions*. The larger the geodesic distance in the solution space between axes origin and a specific point, the larger is the applied rotation to the original *emotion dimensions*. Lastly, the position of each solution with respect to the central point also defines the direction of the rotation (i.e., positive or negative).

Results show that the original unrotated version of the *polarity*, *complexity* and *intensity* dimensions is the optimal solution to explain the gradient-like organization of right TPJ. Indeed, within the space defined by PC rotations, no solutions retained ρ coefficients (i.e., gradient magnitude) larger than those associated with the unrotated components for all the three *emotion dimensions*.

In addition, rotations in which the gradient magnitude is similar across the three *emotion dimensions* are arranged close to the unrotated solution (i.e., white areas in Supplementary Figure 4B), whereas moving away from axes origin at least one of the three dimensions is not represented as a gradient in right TPJ (i.e., yellow and cyan areas in Supplementary Figure 4B). Of note, considering all the explored solutions, very few rotations produce gradients encoding combined *polarity* and *intensity*, but not *complexity* (i.e., lack of magenta areas in Supplementary Figure 4B).

As the original unrotated solution was the best among ~70k explored rotations, we assessed the probability of occurrence of such behavior using a Monte Carlo simulation. Therefore, we created 1,000 PC models by selecting 100 consecutive timepoints from the *emotion dimension* timeseries to predict randomly sampled right TPJ activity ($N = 100$ consecutive timepoints). For each iteration, we then mapped the results of the multiple linear regression analysis (i.e., β coefficients) on a 3-D grid of 25 voxels and computed the correspondence between the anatomical and functional distance obtained using the unrotated and rotated (±45° with 5° step; ~7k explored solutions) predictors. Lastly, we counted the number of iterations in which the gradient magnitude of the rotated predictors was higher with respect to the original unrotated solution.

Results of the Monte Carlo simulation confirm the peculiarity of real data. Indeed, while the unrotated version of *emotion dimensions* represents the optimal solution in explaining right TPJ topography, rotated components produce stronger gradients in the vast majority of simulated cases (96.2%; $p < 0.05$). Of note, we tested the reliability of the results obtained from the Monte Carlo simulation by also varying the length of the timeseries (50, 100 and 200 timepoints), the number of voxels ($N = 25, 100$) and by generating synthetic PC models and fMRI signal using Gaussian noise. Results for all these procedures were consistent with the original simulation (data not shown).

Comparison between emotion gradients and meta-analytic definition of right TPJ

The existence of *emotion dimension* gradients generalizes across several definition of the ROI size, yet the optimal solution is represented by a 15mm radius sphere (11,556 mm^3 volume). In fact, although *emotion dimension* gradients are significantly represented also considering a 27mm ROI (i.e., Supplementary Table 2), the effect size decreases for radii larger than 15mm. To clarify the extent of our *emotion dimension* gradients, we performed a quantitative comparison of the size of our ROI with the definition of right TPJ based on the neuroimaging literature.

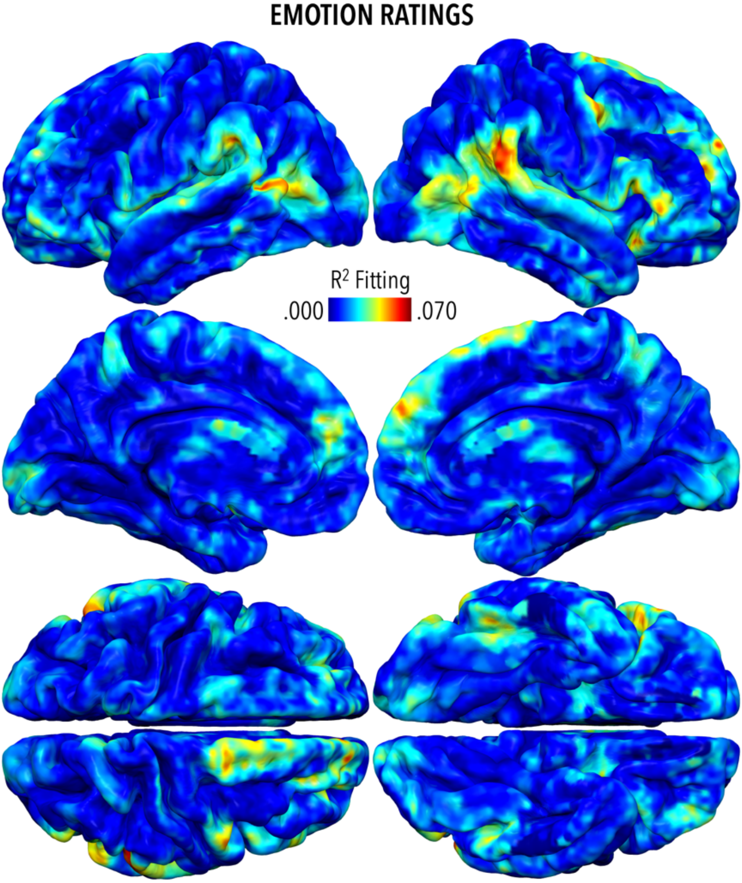
To do so, we considered the right TPJ region obtained from the Neurosynth database (http://old.neurosynth.org/analyses/terms/tpj/). This meta-analytic definition relies on brain activations elicited by classic Theory of Mind and affective processing tasks, such as false-belief (Aichhorn et al., 2009; Döhnel et al., 2012), emotion perception (Garrett et al., 2006) or reappraisal tasks (Silvers et al., 2014), providing a reliable estimate of the right TPJ size. Considering the Neurosynth *TPJ reverse inference map* - i.e., $p(F \mid A)$ -, the volume of the largest cluster was 8,127 mm^3 (coordinates: $x = +58$, $y = -50$, $z = +16$), whereas the volume of the spherical ROI that better represents emotion topography in our study (i.e., 15mm radius) was 11,556 mm^3. Yet, considering the *TPJ forward inference map* - i.e., $p(A \mid F)$ -, the volume of the largest cluster was 16,929 mm^3 (coordinates: $x = +58$, $y = -50$, $z = +16$). Altogether, these results indicate that the optimal description of *emotion dimension* gradients is represented in a patch of cortex that approximates the definition of right TPJ based on brain activation studies (i.e., ~42% larger in volume as compared to the *reverse inference* map, but also ~32% smaller than the *forward inference* definition). Supplementary Figure 15 depicts a comparison between the Neurosynth map and our spherical ROI.

Do right TPJ emotion dimension gradients depend on low-level acoustic and visual features?

To further ensure that the right TPJ *emotion dimension* gradients do not depend on low-level sensory information confounds, we built more complex descriptions of visual and acoustic features of Forrest Gump, based on well-established models. Specifically, we selected spectral power density as a model of low-level acoustic information (de Heer et al., 2017), and GIST descriptors for visual features (Oliva and Torralba, 2001; Rice et al., 2014). We derived the power spectrum for each 2 s segment of the audio track and calculated the power in dB units. The procedure we used is identical to the one described in de Heer and colleagues (de Heer et al., 2017): Welch method, Gaussian window with SD of 5 ms, length 30 ms, 1 ms spacing between windows. The resulting model comprised 449 columns describing the power spectrum of the acoustic signal ranging from 0 Hz to 15 kHz in steps of 33.5 Hz. For the visual model, we segmented each movie frame into a 4x4 grid and sampled the responses to Gabor filters having four different sizes and four possible orientations. This procedure generated a vector of 256 elements, which described each video frame in terms of spatial frequencies, Gabor filter orientations and positions in the visual field. All the GIST descriptors were averaged within a 2 s time window. Timeseries of 449 acoustic and 256 visual features were lagged by 2s and temporally smoothed using a 10s window, similarly to the emotion ratings model. As all our procedures rely on multiple linear regression, which advocate for the use of orthogonal predictors, we performed a PCA on the acoustic and visual models separately and selected the first 21 PCs, which explained more than 90% of the total variance. We then regressed out low-level visual and acoustic stimulus features from brain activity and tested the existence of right TPJ emotion dimension gradients. Importantly, right TPJ *emotion dimension* gradients were not affected by regressing out low-level properties from BOLD signal: *polarity* ($\varrho = 0.258$, p-value = 0.031, 95% CI: 0.252 to 0.264), *complexity* ($\varrho = 0.261$, p-value = 0.013, 95% CI: 0.254 to 0.267) and *intensity* ($\varrho = 0.270$, p-value = 0.016, 95% CI: 0.264 to 0.277). Overall, this evidence indicates that the topographic organization of affective states in right TPJ is not explained by low-level sensory information confounds (Supplementary Figure 17).

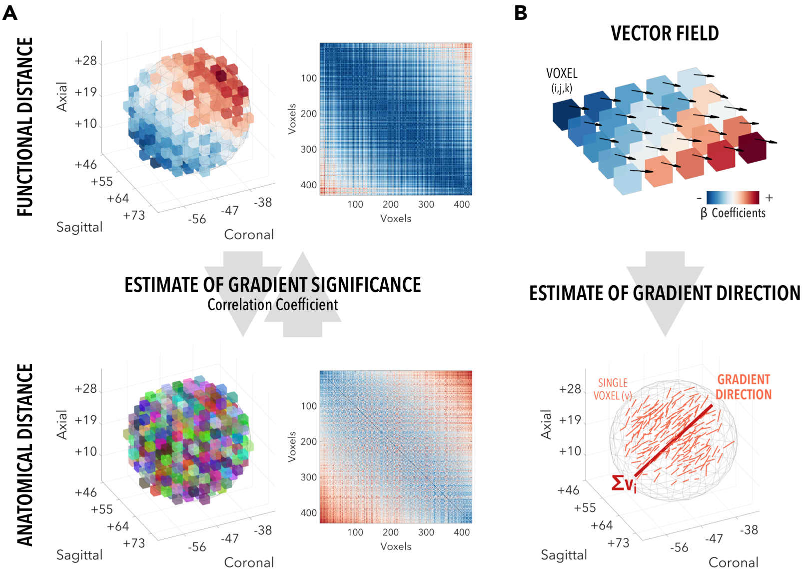
Supplementary Figures

Supplementary Figure 1. Brain regions encoding emotion ratings



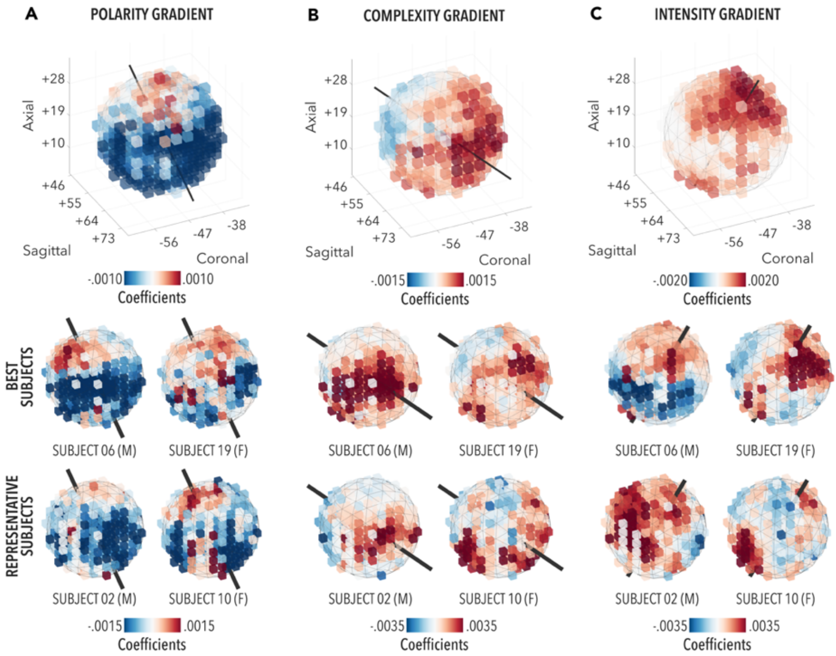
Map of the R^2 fitting of emotion ratings.

Supplementary Figure 2. Characterization of emotion gradients



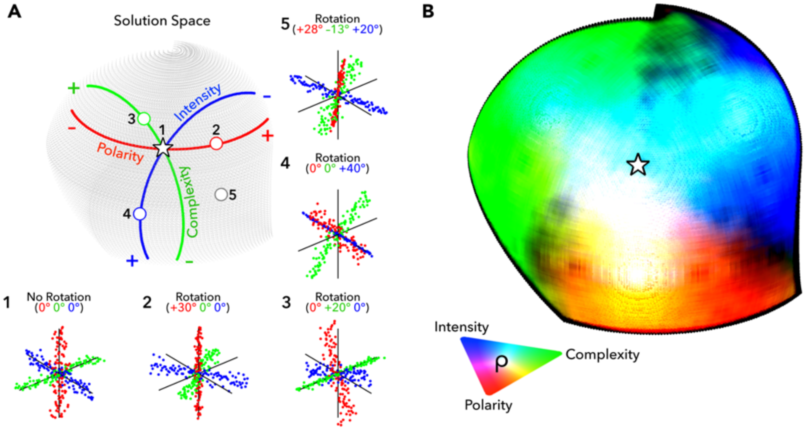
The presence of a gradient-like organization is verified by testing the similarity between functional and anatomical information. Starting from a specific patch of cortex, two dissimilarity matrices are computed (**A**): one using the Euclidean distance of voxel coordinates (i.e., anatomical distance), and the other one using the Euclidean distance of β coefficients related to the fitting of a specific model (i.e., functional distance). Spearman's ϱ is used to measure the strength and assess the significance of the relationship (panel **A**). To derive the main direction of a (linear) gradient, the vector field determined by β coefficients is then estimated and summed across voxels (panel **B**).

Supplementary Figure 3. Single-subject emotion gradients in right TPJ



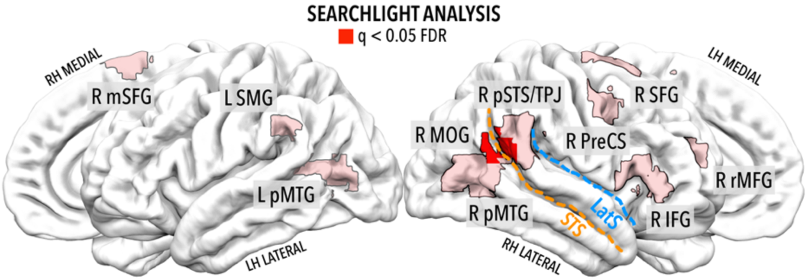
Upmost part of the figure depicts group-level results of gradient mapping (as in Figure 4). Below, results obtained from two of the best subjects (first row; one male and one female) and for two representative subjects (second row; one male and one female). Subjects coding is the one adopted in the *studyforrest* project. Column **A, B** and **C** report β coefficients of the *polarity, complexity* and *intensity* gradients, respectively.

Supplementary Figure 4. Fitting rotated emotion dimensions in right TPJ



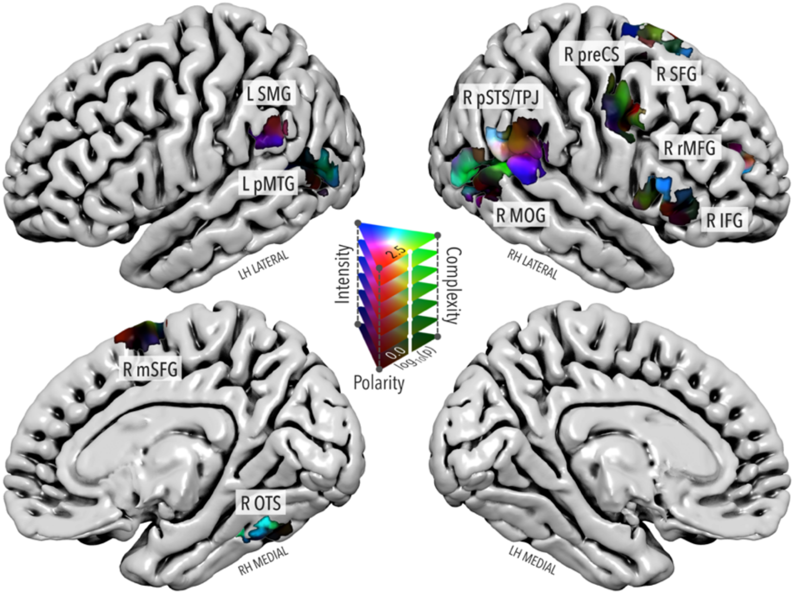
To test whether the rotated version of *emotion dimensions* explains the topographic organization of right TPJ, we systematically applied orthogonal rotations to *polarity*, *complexity* and *intensity* components, fitted each solution in brain activity and then estimated the magnitude of the obtained gradients. Panel **A** depicts the solution space: the pentagram (1) represents the solution determined by the unrotated principal components, whereas the red (2), green (3) and blue (4) round markers express the orthogonal rotation of two components while keeping fixed the other one. The grey round marker (5) maps a solution obtained by applying orthogonal rotations to the three axes. Scatter plots illustrate the transformations applied to the data for each of the points represented in the solution space. Panel **B** shows the effect size (i.e., Spearman's ρ) of the estimate of gradient for all the explored solutions. Magnitude of rotated *polarity*, *complexity* and *intensity* gradients is expressed by hue and brightness.

Supplementary Figure 5. Brain areas representing the combination of emotion dimension gradients as revealed by searchlight analysis



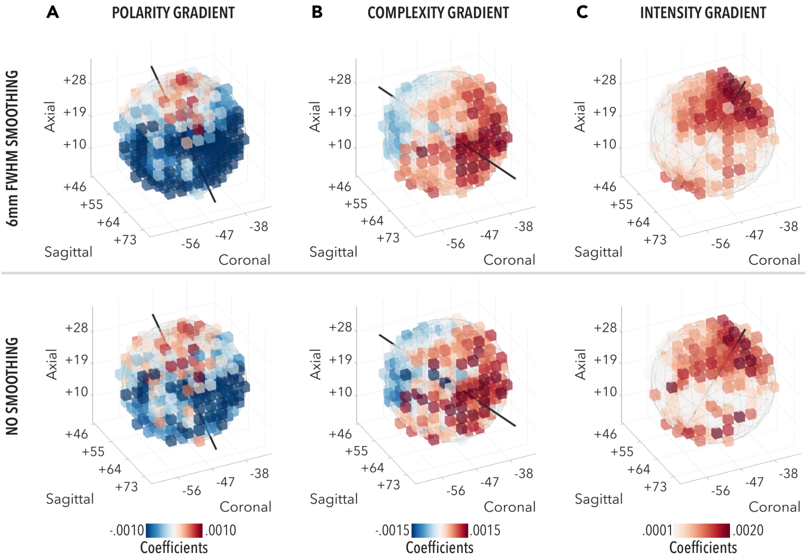
Shaded and outlined regions indicate voxels significantly encoding emotion ratings. In red results corrected using the False Discovery Rate procedure. Datasets for these results are available in the public repository. pSTS/TPJ = posterior part of the superior temporal sulcus/temporoparietal junction, pMTG = posterior middle temporal gyrus, preCS = precentral sulcus, IFG = inferior frontal gyrus, mSFG = medial superior frontal gyrus, SMG = supramarginal gyrus, rMFG = rostral middle frontal gyrus, MOG = middle occipital gyrus.

Supplementary Figure 6. Brain areas representing individual emotion dimension gradients as revealed by searchlight analysis



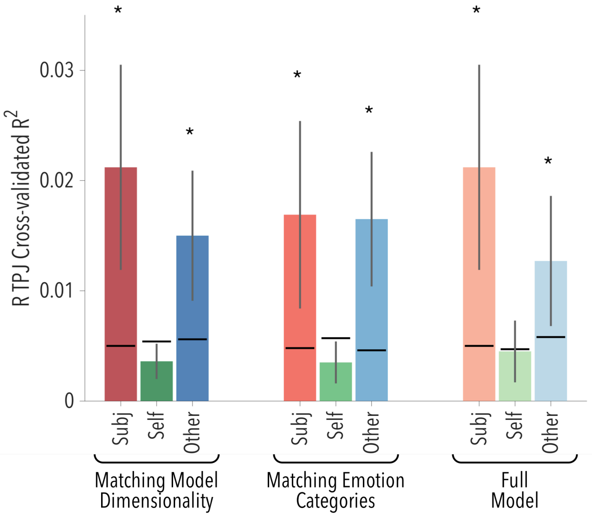
Results for the three separate searchlight analyses measuring the topographic arrangement of *polarity* (red channel), *complexity* (green channel) and *intensity* (blue channel). Color brightness relates to the log transformed p-value of the fitting of each component. Datasets for these results are available in the public repository. pSTS/TPJ = posterior part of the superior temporal sulcus/temporoparietal junction, pMTG = posterior middle temporal gyrus, preCS = precentral sulcus, IFG = inferior frontal gyrus, mSFG = medial superior frontal gyrus, SMG = supramarginal gyrus, rMFG = rostral middle frontal gyrus, MOG = middle occipital gyrus; OTS = occipitotemporal sulcus.

Supplementary Figure 7. Emotion dimension gradients using unsmoothed fMRI data



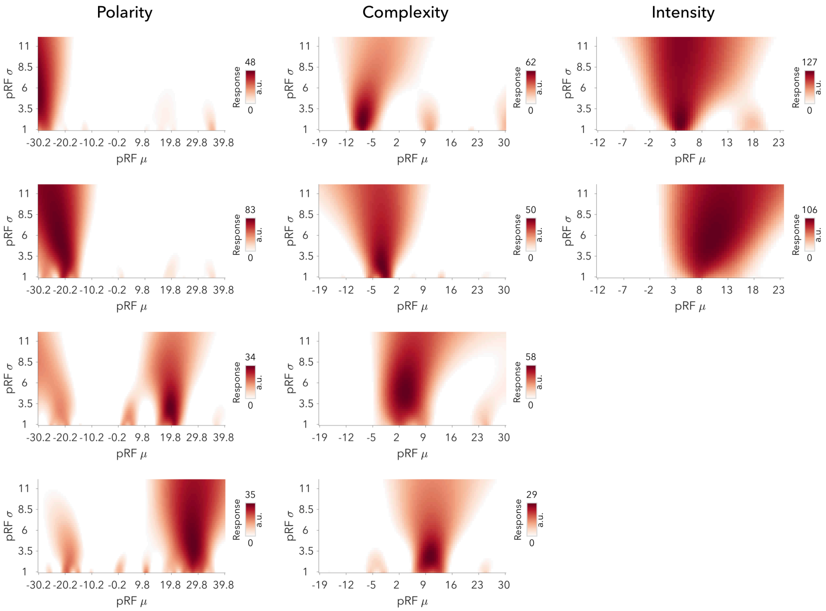
Uppermost row depicts β coefficients of *emotion dimensions* obtained when applying 6mm FWHM smoothing (3dBlurToFWHM). Lowermost row depicts β coefficients of *emotion dimensions* obtained without applying any spatial filtering. Panel **A**, **B** and **C** represent *polarity*, *complexity* and *intensity* gradients in right TPJ, respectively.

Supplementary Figure 8. Subjective emotion rating model versus third person emotion attribution models



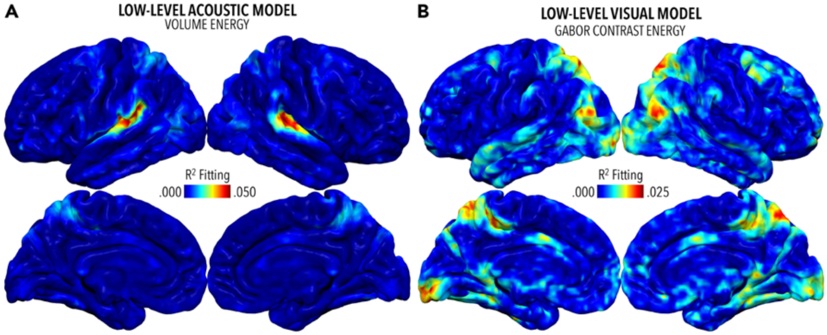
We measured the extent to which the two *third-person emotion attribution models* explained brain activity in right TPJ. We assessed the significance of fitting using three different procedures: (**A**) matching the dimensionality across models by selecting the first six principal components only; (**B**) matching the emotion categories in ratings, by performing PCA on the four basic emotions shared across models (i.e., *happiness*, *fear*, *sadness* and *anger*); (**C**) using the full model regardless of the dimensionality (i.e., six for our *subjective emotion rating* and 22 for the *emotion attribution models*). Results showed that only the *subjective emotion rating model* and the *other-directed emotion attribution* one significantly explained activity of right TPJ (**A**: *subj* $R^2 = 0.021$, $p < 0.002$; *other* $R^2 = 0.015$, $p < 0.002$; *self* $R^2 = 0.004$, $p = 0.269$. **B**: *subj* $R^2 = 0.017$, $p < 0.002$; *other* $R^2 = 0.016$, $p < 0.002$; *self* $R^2 = 0.003$, $p = 0.335$. **C**: *subj* $R^2 = 0.021$, $p < 0.002$; *other* $R^2 = 0.013$, $p < 0.002$; *self* $R^2 = 0.004$, $p = 0.078$). The *subjective emotion rating* and the *other-directed emotion attribution model* did not significantly differ in explaining activity of right TPJ ($p > 0.05$). The lower and upper noise ceiling bounds averaged across all the right TPJ voxels were $R^2 = 0.11$ and $R^2 = 0.20$. * denotes $p < 0.05$; error bar indicates standard error; bold horizontal line is the 95th percentile of the null distribution.

Supplementary Figure 9. Preferred responses of distinct populations of voxels using non-negative matrix factorization



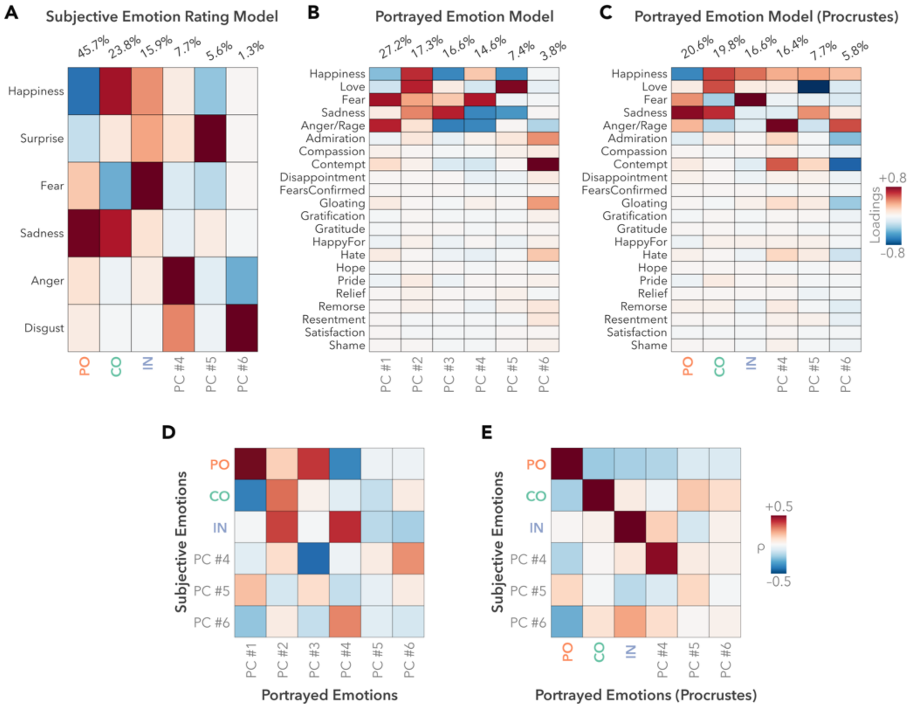
Prototypical responses of populations of voxels as function of affective states. We decomposed the pRF data (i.e., voxels t-values for all the explored μ and σ in the grid-search procedure) using non-negative matrix factorization. The figure depicts resulting components retaining at least 5% of the variance for *polarity* (i.e., first column), *complexity* (i.e., second column) and *intensity* (i.e., third column). Results highlight the existence of four distinct populations of voxels tuned to specific scores of *polarity* and *complexity*. Two populations represented distinct *intensity* values.

Supplementary Figure 10. Fitting of low-level features



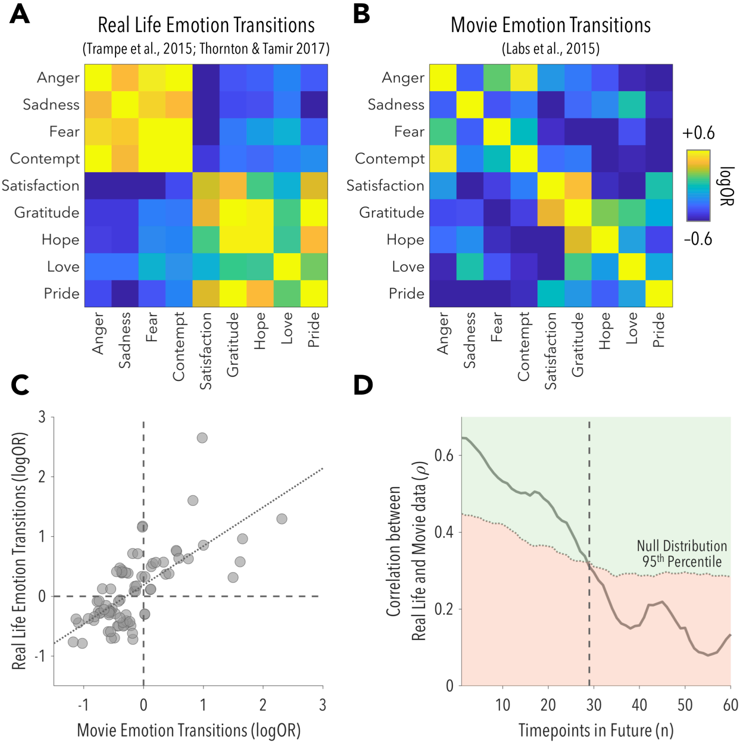
The fitting of low-level stimulus properties was estimated to verify the adequacy of adopted models. **A**. Peak of fitting for the volume energy model (i.e., RMS of the audio track) is located in primary auditory cortex. **B**. Gabor contrast energy (i.e., low and high spatial frequencies) of movie frames explained activity in primary visual cortex as well as in other associative areas (e.g., retrosplenial, parahippocampal and superior parietal cortex).

Supplementary Figure 11. Reconstruction of the emotion dimensions from portrayed emotions



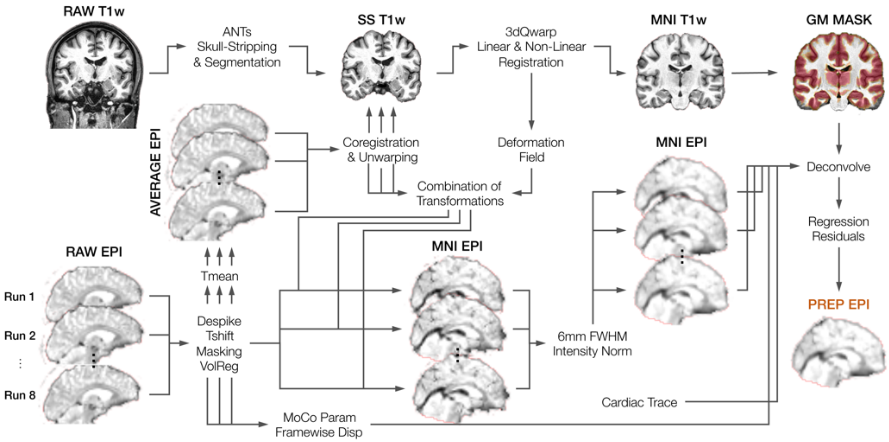
Panel **A** shows the six PCs obtained from subjective emotion ratings. Panel **B** depicts the output of PCA for Labs and colleagues[41] data. The first six dimensions represented ~85% of the explained variance. Panel **C** demonstrates that our *polarity*, *complexity* and *intensity* dimensions emerge from the portrayed emotion model after rotating PC scores using the procrustes criterion. Lowermost panels report the correlation between our original PCs and the unrotated (**D**) and rotated (**E**) version of components derived from portrayed emotions.

Supplementary Figure 12. Real life versus Forrest Gump emotion transitions



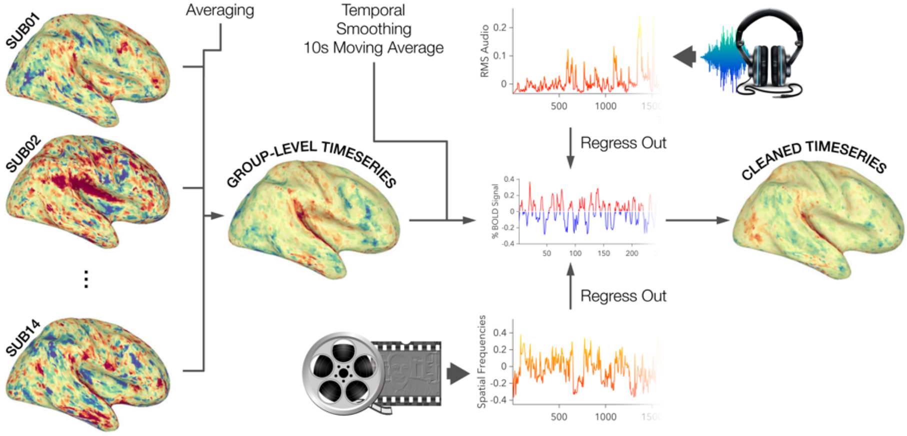
We analyzed data relative to study 3 of Thornton & Tamir 2017 and selected the emotion categories in common with Labs and coauthors 2015. (**A**) Matrix depicting *real life emotion transitions*: each cell represents the log odds of a particular emotion transition. We built this matrix from an experience-sampling dataset of subjects reporting their affective state throughout the day. (**B**) Matrix showing *movie emotion transitions*: each cell represents the log odds of a particular emotion transition during Forrest Gump. We built this matrix from the reports of portrayed emotions. (**C**) *Real life emotion transitions* are significantly associated to the *movie-based emotion transitions* (Spearman's $\rho = 0.646$; $p = 0.001$). (**D**) We built a number of *movie-based* models, each measuring the likelihood of emotion transition between timepoint *t* and timepoint *t+n* in the future, with a maximum delay of 120 seconds (60 timepoints). These models were then correlated with (**A**) and results show that the *real life* model predicts emotion transitions in the movie up to 58 seconds.

Supplementary Figure 13. fMRI data single-subject preprocessing



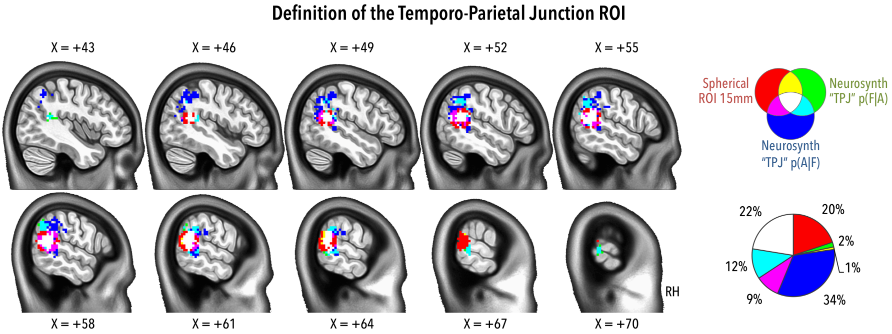
Preprocessing pipeline for structural and functional MRI data.

Supplementary Figure 14. fMRI data group-level preprocessing



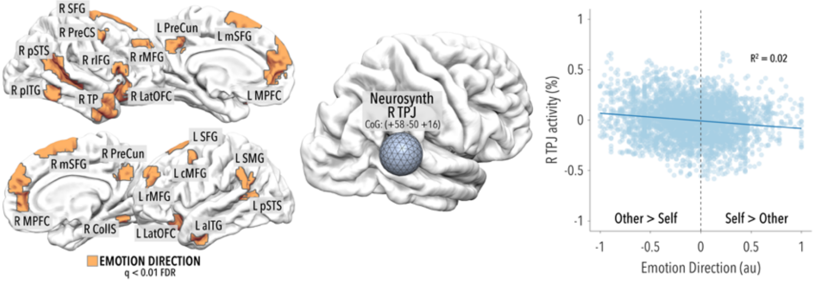
Single-subject preprocessed fMRI data were averaged to obtain group-level hemodynamic activity. For each voxel a windowing procedure was employed to temporally smooth data (moving average: 10s window). From the obtained aggregated and smoothed timeseries, the timecourse of low-level acoustic (i.e., volume energy - RMS of the signal) and visual (i.e., Gabor contrast energy for 0.5 and 8 cyc/deg spatial frequencies for each frame) movie features were regressed out, so as to mitigate the possible collinearity between emotion ratings and low-level psychophysical properties of the stimulus.

Supplementary Figure 15. Comparison between emotion gradients and meta-analytic definition of right TPJ



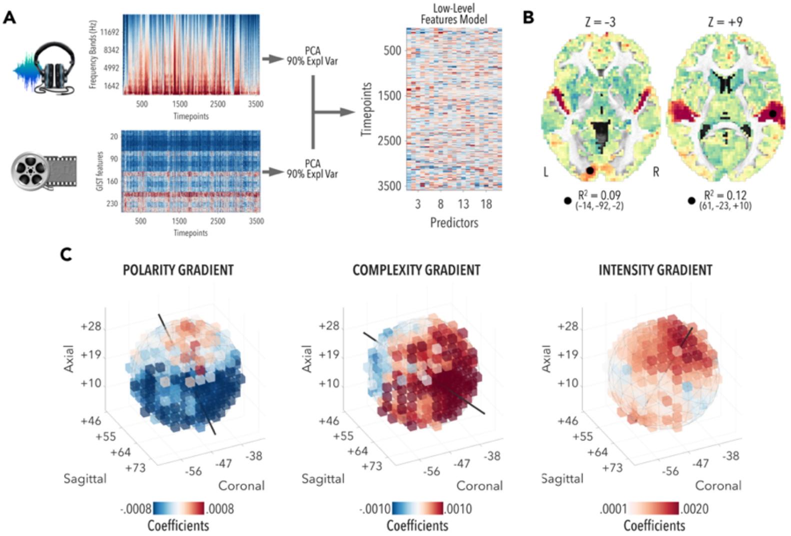
We obtained from the Neurosynth database (http://old.neurosynth.org/analyses/terms/tpj/) two meta-analytic maps representing a reliable estimate of the right TPJ size, against which we compared the volume of our spherical ROI. Neurosynth *TPJ reverse inference map* - p(F | A) - is represented in green, whereas the *TPJ forward inference map* - p(A | F) - is in blue and our spherical ROI (i.e., 15mm radius) is in red. Pie chart represents the percentage of volume related to our ROI, the two meta-analytic maps and their overlap.

Supplementary Figure 16. Regions associated to the direction of portrayed emotions



We performed a voxel-wise encoding of the *direction* of portrayed emotions on group-averaged BOLD signal. The higher the BOLD of right TPJ, the more raters labeled emotions as *other-directed* (right TPJ peak R^2: 0.04; right TPJ average R^2: 0.02). Significant associations ($p < 0.01$ FDR corrected) between *emotion direction* and BOLD signal were also found in other brain regions of the Theory of Mind, empathy and emotion processing networks, closely resembling the pattern found by Hanke and colleagues, 2016. Datasets for these results are available in the public repository. CoG = center of gravity; pSTS = posterior part of the superior temporal sulcus, preCS = precentral sulcus, IFG = inferior frontal gyrus, mSFG = medial superior frontal gyrus, SMG = supramarginal gyrus, rMFG = rostral middle frontal gyrus, TP = temporal pole, CollS = collateral sulcus, aITG = anterior inferior temporal gyrus, cMFG = caudal middle frontal gyrus, SFG = superior frontal gyrus, LatOFC = lateral orbitofrontal cortex, MPFC = middle prefrontal cortex, PreCun = precuneus, pITG = posterior inferior temporal gyrus.

Supplementary Figure 17. Right TPJ emotion dimension gradients do not depend on low-level acoustic and visual features



Panel A represents the 21 PCs used to model acoustic and visual low-level properties of the Forrest Gump movie derived from power spectral features and GIST descriptors. Using this model, we have more than doubled the explained variance in sensory cortical areas (12% in Heschl's gyrus and 9% in pericalcarine cortex; **B**), as compared to RMS and contrast energy models (Supplementary Figure 10). Of note, upper and lower noise ceiling bounds for the highest R^2 voxels were 0.268-0.172 in primary auditory cortex and 0.412-0.330 in early visual cortex. These numbers suggest that our 21 PCs model explains up to 70% and 27% of brain activity within these regions. Most importantly, we found that right TPJ emotion dimensions gradient were not affected by regressing out low-level properties from BOLD signal (**C**): *polarity* (ϱ = 0.258, p-value = 0.031, 95% CI: 0.252 to 0.264), *complexity* (ϱ = 0.261, p-value = 0.013, 95% CI: 0.254 to 0.267) and *intensity* (ϱ = 0.270, p-value = 0.016, 95% CI: 0.264 to 0.277).

Supplementary Tables

Supplementary Table 1. Brain regions encoding emotion ratings

	Cluster	Peak			CoG		
	Size	x	y	z	x	y	z
R pSTS/TPJ	343	61.5	-40.5	19.5	55.9	-54.1	16
L pMTG	96	-49.5	-52.5	10.5	-51	-63.8	8.4
R preCS	95	49.5	1.5	55.5	48.7	3.1	44.6
R IFG	42	55.5	22.5	10.5	50	22.1	5.6
R mSFG	30	1.5	-4.5	70.5	6	-1.3	68.8
R OTS	28	49.5	-46.5	-19.5	46.7	-44.6	-20.7
L SMG	20	-55.5	-40.5	31.5	-60.8	-41	32.6
R IFG	15	55.5	31.5	-1.5	53.2	33.1	1.3
R mSFG	15	13.5	16.5	61.5	14	14.4	61.1
R mSFG	15	16.5	1.5	67.5	14.8	3	69
R rMFG	12	22.5	52.5	19.5	23.7	53	17.5

Table showing regions significantly associated to emotion ratings ($q < 0.01$; minimum cluster size > 10 voxels). Voxel size = 3 mm isotropic; CoG = center of gravity; pSTS/TPJ = posterior part of the superior temporal sulcus/temporoparietal junction, pMTG = posterior middle temporal gyrus, preCS = precentral sulcus, IFG = inferior frontal gyrus, mSFG = medial superior frontal gyrus, OTS = occipitotemporal sulcus, SMG = supramarginal gyrus, rMFG = rostral middle frontal gyrus.

Supplementary Table 2. Emotion gradients in TPJ

	Radius	Emotion Dimensions		Basic Emotions	
	(mm)	**ϱ**	**p-value**	**ϱ**	**p-value**
Right TPJ	9	0.316	0.037	0.286	0.076
	12	0.375	0.001	0.326	0.024
	15	0.399	<0.001	0.352	0.004
	18	0.387	<0.001	0.332	0.006
	21	0.372	<0.001	0.319	0.007
	24	0.342	0.001	0.299	0.012
	27	0.292	0.003	0.266	0.021
Left TPJ	15	0.251	0.144	0.208	0.356

To identify the patch of cortex with the highest significant association between anatomical and functional distance, we started from the reverse inference peak for the term "*TPJ*" in the NeuroSynth database. We then created a set of spherical ROIs having as center of gravity this peak and with radius ranging from 9 to 27 mm. For each ROI, we tested the relationship between anatomical and functional distance using the procedure detailed above and depicted in Supplementary Figure 2. The procedure was performed using either the three *emotion dimensions* or the four basic emotions stable across all subjects. Results demonstrated that within a 15 mm radius ROI, relative spatial arrangement and functional features of right TPJ were significantly and maximally correlated either considering the basic emotion model or the *emotion dimensions* one. Moreover, we included a 15 mm ROI centered at the left TPJ as control region (Neurosynth definition). TPJ = Temporoparietal junction.

Supplementary Table 3. Emotion gradients in TPJ relative to emotion dimensions or basic emotions

	Radius	Polarity		Complexity		Intensity		PC #4		PC #5		PC #6	
	(mm)	**ϱ**	**p-value**	**ϱ**	**p-value**	**ϱ**	**p-value**	ϱ	p-value	ϱ	p-value	ϱ	p-value
Right TPJ	15	0.241	0.041	0.271	0.013	0.229	0.049	0.044	0.975	0.239	0.052	0.114	0.598
Left TPJ	15	0.132	0.354	0.157	0.222	0.149	0.257	0.088	0.643	0.049	0.889	0.171	0.169

	Radius	Happiness		Surprise		Fear		Sadness		Anger		Disgust	
	(mm)	**ϱ**	**p-value**	ϱ	p-value	**ϱ**	**p-value**	**ϱ**	**p-value**	**ϱ**	**p-value**	ϱ	p-value
Right TPJ	15	0.275	0.013	0.202	0.112	0.197	0.091	0.182	0.160	0.141	0.379	0.097	0.724
Left TPJ	15	0.158	0.216	0.028	0.964	0.142	0.293	0.156	0.213	0.073	0.733	0.163	0.179

For each individual *emotion dimension* and basic emotion, we tested the existence of a gradient-like organization in a spherical ROI (15 mm radius) located within the TPJ region (Neurosynth definition). Results for *emotion dimensions* and basic emotions consistent across all subjects are reported in black (see the *Agreement across subjects of the six basic emotions* and *Agreement across subjects of the emotion dimensions* sections). Significant results are marked with bold. TPJ = Temporoparietal junction.

Supplementary Table 4. Single-subject emotion dimension gradients in right TPJ

Sub ID	Gender	Polarity		Complexity		Intensity	
		ϱ	**p-value**	**ϱ**	**p-value**	**ϱ**	**p-value**
02	M	0.464	0.019	0.446	0.002	0.411	0.071
03	F	0.349	0.013	0.362	0.029	0.521	<0.001
04	F	0.607	0.002	0.465	<0.001	0.423	0.007
05	M	0.074	0.244	0.199	0.229	-0.004	0.508
06	M	0.660	0.003	0.639	0.001	0.480	<0.001
09	M	0.354	0.009	0.313	0.017	0.245	0.056
10	F	0.423	0.009	0.347	0.038	0.263	0.045
14	F	0.306	0.098	0.495	<0.001	0.231	0.117
15	M	0.232	0.044	0.183	0.098	0.151	0.182
16	M	0.380	0.006	0.256	0.103	0.561	<0.001
17	M	0.291	0.015	0.188	0.081	0.407	0.041
18	M	0.154	0.210	0.448	0.008	0.369	0.017
19	F	0.619	0.001	0.485	0.015	0.559	<0.001
20	F	0.594	<0.001	0.425	0.001	0.580	<0.001

We tested the consistency of *emotion dimension* gradients in right TPJ using single-subject data. Firstly, preprocessed fMRI single-subject timeseries were smoothed in time (10s moving average window) and cleaned from low-level visual and acoustic features of the movie, as in the group-level analysis pipeline. Subsequently, we performed an encoding analysis using the behavioral ratings and obtained β values for *polarity*, *complexity* and *intensity*. Afterwards, we measured the relationship between single-subject maps and those obtained from group-level analysis using Spearman's ϱ coefficient. To measure the statistical significance of these associations, we employed a surrogate-based approach by generating 1,000 *emotion dimension* encoding models using the IAAFT procedure, as described in the *Methods* section. Bold values represent significant associations between single-subject and group-level gradients.

Supplementary Table 5. Topographic organization of portrayed emotions in right TPJ

PC	Other-directed model		Other-directed CCA	
#	ϱ	p-value	ϱ	p-value
1	0.136	0.404	0.221	0.036
2	0.111	0.594	0.150	0.384
3	0.119	0.554	0.207	0.092
4	0.081	0.844	0.062	0.897
5	0.157	0.276	0.101	0.684
6	0.118	0.533	0.107	0.664
7	0.147	0.363		
8	0.105	0.676		
9	0.178	0.234		
10	0.190	0.154		
11	0.124	0.546		
12	0.177	0.200		
13	0.158	0.270		
14	0.124	0.511		
15	0.154	0.340		
16	0.115	0.587		
17	0.094	0.735		
18	0.290	0.004		
19	0.101	0.688		
20	0.105	0.669		
21	0.130	0.485		
22	0.124	0.519		

We tested right TPJ topography for the *other-directed emotion attribution* PCs. None of the first six components retained a topographical organization in this region. Only the 18th PC, explaining the 0.3% of the variance, appeared to be encoded in a gradient-like manner. However, the pattern associated to this component was also collinear with activity evoked by *polarity* (ϱ = 0.494) and *intensity* (ϱ = 0.475) dimensions. Moreover, using CCA (canonical correlation analysis) we transformed the 22-dimensional space defined by the *other-directed model* to match our subjective reports. Noteworthy, when fitting the aligned components into right TPJ activity, only the first PC (i.e., reconstructed *polarity*) was represented through a gradient. PC = principal component.

Supplementary Table 6. Topographies in right TPJ considering spatial smoothing and cortical folding

	Radius	Polarity		Complexity		Intensity		PC #4		**PC #5**		PC #6	
	(mm)	ϱ	p-value	ϱ	p-value	ϱ	p-value	ϱ	p-value	ϱ	p-value	ϱ	p-value
Right TPJ	15	0.167	0.033	0.186	0.010	0.184	0.010	0.014	0.996	0.194	0.018	0.076	0.584

	Radius	Polarity		Complexity		Intensity		PC #4		PC #5		PC #6	
	(mm)	ϱ	p-value	ϱ	p-value	ϱ	p-value	ϱ	p-value	ϱ	p-value	ϱ	p-value
Right TPJ	15	0.248	0.026	0.314	0.001	0.249	0.013	0.012	0.961	0.130	0.323	0.083	0.577

Table showing the robustness of *emotion dimension* gradients in right TPJ using unsmoothed data. The first row regards the evaluation of gradients in the unfiltered volumetric space. The second row refers to the results of the same analysis conducted with unfiltered data into surface space, using the Dijkstra algorithm. PC = principal component; TPJ = Temporoparietal junction.

7. References

Abu-Akel, A., & Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia, 49*(11), 2971-2984.

Adolphs, R. (2008). Fear, faces, and the human amygdala. *Current opinion in neurobiology*, *18*(2), 166-172.

Adolphs, R. (2013). The biology of fear. *Current Biology, 23*(2), R79-R93.

Adolphs, R. (2017). How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences. *Social Cognitive Affective Neuroscience*, 12(1), 24-31.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., ... & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia, 37*(10), 1111-1117.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*(7021), 68.

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., & Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. *Journal of Cognitive Neuroscience, 21*(6), 1179-1192.

Amft, M., Bzdok, D., Laird, A. R., Fox, P. T., Schilbach, L., & Eickhoff, S. B. (2015). Definition and characterization of an extended social-affective default network. *Brain Structure and Function*, *220*(2), 1031-1049.

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., ... & Sobel, N. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature neuroscience, 6*(2), 196.

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience, 16*(10), 1773-1784.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron, 95*(3), 709-721.

Baron-Cohen, S. (1997). Mindblindness: An essay on autism and theory of mind. MIT press.

Baron-Cohen, S. (1998). Does the study of autism justify minimalist innate modularity?. *Learning and Individual Differences*, *10*(3), 179-191.

Baron-Cohen, S. (2005). The empathizing system. *Origins of the social mind: Evolutionary psychology and child development*, 468-492.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163-175.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37-46.

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. *Journal of Child Psychology and Psychiatry*, *42*(2), 241-251.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, *30*(2), 285-297.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813-822.

Barrett, L. F. (2006). Are emotions natural kinds?. *Perspectives on Psychological Science*, *1*(1), 28-58.

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition & Emotion*, *23*(7), 1284-1306.

Barrett, L. F. (2012). Emotions are real. *Emotion*, *12*(3), 413.

Barrett, L. F. (2013). Psychological construction: The Darwinian approach to the science of emotion. *Emotion Review*, *5*(4), 379-389.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience*, *12*(1), 1-23.

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*, *8*(1), 10-14.

Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Current opinion in neurobiology*, *23*(3), 361-372.

Barrett, L. F., & Wager, T. D. (2006). The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science*, *15*(2), 79–83.

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, *58*, 373-403.

Batson, C. D. (2009). These things called empathy: eight related but distinct phenomena.

Batson, C. D., & Shaw, L. L. (1991). Evidence for altruism: Toward a pluralism of prosocial motives. *Psychological inquiry*, *2*(2), 107-122.

Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L., & Ortiz, B. G. (2007). An additional antecedent of empathic concern: valuing the welfare of the person in need. *Journal of Personality and Social Psychology*, *93*(1), 65.

Baucom, L. B., Wedell, D. H., Wang, J., Blitzer, D. N., & Shinkareva, S. V. (2012). Decoding the neural representation of affective states. *Neuroimage*, *59*(1), 718-727.

Bechara, A. (2004). Disturbances of emotion regulation after focal brain lesions. Int Rev Neurobiol, 62(159), 93.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*, *10*(3), 295-307.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531-539.

Benjamini Y, Hochberg Y. (1995). Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289-300.

Benuzzi, F., Lui, F., Duzzi, D., Nichelli, P. F., & Porro, C. A. (2008). Does it look painful or disgusting? Ask your parietal and cingulate cortex. *Journal of Neuroscience*, *28*(4), 923-931.

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual review of neuroscience*, *35*, 1-23.

Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony.

Berrios, R., Totterdell, P., & Kellett, S. (2015). Eliciting mixed emotions: a meta-analysis comparing models, types, and measures. *Frontiers in Psychology*, *6*, 428.

Betti, V., Della Penna, S., de Pasquale, F., Mantini, D., Marzetti, L., Romani, G. L., & Corbetta, M. (2013). Natural scenes viewing alters the dynamics of functional connectivity in the human brain. *Neuron, 79*(4), 782-797.

Bilenko, N. Y., & Gallant, J. L. (2016). Pyrcca: regularized kernel canonical correlation analysis in python and its applications to neuroimaging. *Frontiers in neuroinformatics*, *10*, 49.

Billot, P. E., Andrieu, P., Biondi, A., Vieillard, S., Moulin, T., & Millot, J. L. (2017). Cerebral bases of emotion regulation toward odours: A first approach. *Behavioural brain research*, *317*, 37-45.

Bodden, M. E., Dodel, R., & Kalbe, E. (2010). Theory of mind in Parkinson's disease and related basal ganglia disorders: a systematic review. *Movement Disorders*, *25*(1), 13-27.

Bodden, M. E., Kübler, D., Knake, S., Menzler, K., Heverhagen, J. T., Sommer, J., ... & Dodel, R. (2013). Comparing the neural correlates of affective and cognitive theory of mind using fMRI: Involvement of the basal ganglia in affective theory of mind. *Advances in cognitive psychology*, *9*(1), 32.

Bohart, A. C., Elliott, R., Greenberg, L. S., & Watson, J. C. (2002). Empathy.

Borod, J. C. (Ed.). (2000). *The neuropsychology of emotion*. Oxford University Press.

Borod, J. C., Cicero, B. A., Obler, L. K., Welkowitz, J., Erhan, H. M., Santschi, C., ... & Whalen, J. R. (1998). Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology, 12*(3), 446.

Bowers, D., Blonder, L. X., & Heilman, K. M. (1998). *Florida affect battery*. Florida, USA: Center for Neuropsychological Studies, Department of Neurology.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry, 25*(1), 49-59.

Bradley, M. M., & Lang, P. J. (1999b). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Vol. 30, No. 1, pp. 25-36). Technical report C-1, the center for research in psychophysiology, University of Florida.

Bradley, M., & Lang, P. J. (1999a). *The International affective digitized sounds (IADS)[: stimuli, instruction manual and affective ratings*. NIMH Center for the Study of Emotion and Attention.

Brosch, T., Scherer, K. R., Grandjean, D. M., & Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision-making. *Swiss medical weekly, 143*, w13786.

Brothers, L., & Ring, B. (1992). A neuroethological framework for the representation of minds. *Journal of cognitive neuroscience,* 4(2), 107-118.

Brown, S., Gao, X., Tisdelle, L., Eickhoff, S. B., & Liotti, M. (2011). Naturalizing aesthetics: brain areas for aesthetic appraisal across sensory modalities. *Neuroimage, 58*(1), 250-258.

Brüne, M. (2005). "Theory of mind" in schizophrenia: a review of the literature. *Schizophrenia bulletin, 31*(1), 21-42.

Budell, L., Kunz, M., Jackson, P. L., & Rainville, P. (2015). Mirroring pain in the brain: Emotional expression versus motor imitation. *PloS one, 10*(2), e0107526.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience, 10*(3), 186-198.

Burnett, S., & Blakemore, S. J. (2009). Functional connectivity during a social emotion task in adolescents and in adults. *European Journal of Neuroscience, 29*(6), 1294-1301.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*(2), 337-342.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. *Handbook of emotions, 2*, 173-191.

Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature neuroscience, 3*(11), 1077.

Campanella, F., Shallice, T., Ius, T., Fabbro, F., & Skrap, M. (2014). Impact of brain tumour location on emotion and personality: A voxel-based lesion–symptom mapping study on mentalization processes. *Brain, 137*(9), 2532-2545.

Cappella, J. N., & Planalp, S. (1981). Talk and silence sequences in informal conversations III: Interspeaker influence. *Human Communication Research*, *7*(2), 117-132.

Chakrabarti, B., & Baron-Cohen, S. (2006). Empathizing: neurocognitive developmental mechanisms and individual differences. *Progress in brain research*, *156*, 403-417.

Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2017). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*, *27*(3), 1944-1948.

Colibazzi, T., Posner, J., Wang, Z., Gorman, D., Gerber, A., Yu, S., ... & Peterson, B. S. (2010). Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion*, *10*(3), 377.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience, 3*(3), 201.

Corradi-Dell'Acqua, C., Hofstetter, C., & Vuilleumier, P. (2011). Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. *Journal of Neuroscience*, *31*(49), 17996-18006.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *PNAS, 114*(38), E7900-E7909.

Couto, B., Sedeno, L., Sposato, L. A., Sigman, M., Riccio, P. M., Salles, A., ... & Ibanez, A. (2013). Insular networks for emotional processing and social cognition: comparison of two case reports with either cortical or subcortical involvement. *Cortex, 49*(5), 1420-1434.

Craig, K. D. (2004). Social communication of pain enhances protective functions: a comment on Deyo, Prkachin and Mercer.

Dal Monte, O., Schintu, S., Pardini, M., Berti, A., Wassermann, E. M., Grafman, J., & Krueger, F. (2014). The left inferior frontal gyrus is crucial for reading the mind in the eyes: brain lesion evidence. *Cortex*, 58, 9-17.

Damasio, A., Damasio, H., & Tranel, D. (2012). Persistence of feelings and sentience after bilateral damage of the insula. *Cerebral Cortex*, *23*(4), 833-846.

Danti, S., Handjaras, G., Cecchetti, L., Beuzeron-Mangina, H., Pietrini, P., & Ricciardi, E. (2018). Different levels of visual perceptual skills are associated with specific modifications in functional connectivity and global efficiency. *International Journal of Psychophysiology, 123*, 127-135.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology, 44*(1), 113.

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual review of neuroscience, 15*(1), 353-375.

Davis, M. H. (2018). *Empathy: A social psychological approach*. Routledge.

de Greck, M., Shi, Z., Wang, G., Zuo, X., Yang, X., Wang, X., ... & Han, S. (2012). Culture modulates brain activity during empathy with anger. *Neuroimage, 59*(3), 2871-2882.

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *Journal of Neuroscience, 37*(27), 6539-6557.

De Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why?. *Trends in cognitive sciences, 10*(10), 435-441.

de Waal, F. B. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology., 59*, 279-300.

de Waal, F. B., & Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nature Reviews Neuroscience, 18*(8), 498.

Decety, J. (2011). Dissecting the neural mechanisms mediating empathy. *Emotion review, 3*(1), 92-108.

Decety, J. E., & Ickes, W. E. (2009). *The social neuroscience of empathy*. MIT Press.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews, 3*(2), 71-100.

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology, 19*(6), 643-647.

Döhnel, K., Schuwerk, T., Meinhardt, J., Sodian, B., Hajak, G., & Sommer, M. (2012). Functional activity of the right temporo-parietal junction and of the medial prefrontal cortex associated with true and false belief reasoning. *Neuroimage, 60*(3), 1652-1661.

Donaldson, P. H., Rinehart, N. J., & Enticott, P. G. (2015). Noninvasive stimulation of the temporoparietal junction: A systematic review. *Neuroscience & Biobehavioral Reviews, 55*, 547-572.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage, 39*(2), 647-660.

Dvash, J., & Shamay-Tsoory, S. G. (2014). Theory of mind and empathy as multidimensional constructs: Neurological foundations. *Topics in Language Disorders, 34*(4), 282-295.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., ... & Convit, A. (2006). Introducing MASC: a movie for the assessment of social cognition. *Journal of autism and developmental disorders, 36*(5), 623-636.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual review of psychology, 51*(1), 665-697.

Eisenberg, N., & Fabes, R. A. (1990). Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion, 14*(2), 131-149.

Ejaz, N., Hamada, M., & Diedrichsen, J. (2015). Hand use predicts the structure of representations in sensorimotor cortex. *Nature Neuroscience, 18*(7), 1034.

Ekman, P. (1989). The argument and evidence about universals in facial expressions. *Handbook of social psychophysiology*, 143-164.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169-200.

Ekman, P. (1999). Facial expressions. Handbook of cognition and emotion, 16, 301–20.

Ekman, P. E., & Davidson, R. J. (1994). *The nature of emotion: Fundamental questions*. Oxford University Press.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion review*, *3*(4), 364-370.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124.

Ekman, P., & Keltner, D. (1997). Universal facial expressions of emotion. *Segerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture*, 27-46.

Elamin, M., Pender, N., Hardiman, O., & Abrahams, S. (2012). Social cognition in neurodegenerative disorders: a systematic review. *Journal of Neurology, Neurosurgery, and Psychiatry*, *83*(11), 1071-1079.

Eslinger, P. J. (1998). Neurological and neuropsychological bases of empathy. *European neurology*, *39*(4), 193-199.

Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 903-911.

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of experimental psychology: General*, *113*(3), 464.

Feinstein, J. S. (2013). Lesion studies of human emotion and feeling. *Current opinion in neurobiology*, *23*(3), 304-309.

Feinstein, J. S., Adolphs, R., Damasio, A., & Tranel, D. (2011). The human amygdala and the induction and experience of fear. *Current biology, 21*(1), 34-38.

Feinstein, J. S., Buzza, C., Hurlemann, R., Follmer, R. L., Dahdaleh, N. S., Coryell, W. H., ... & Wemmie, J. A. (2013). Fear and panic in humans with bilateral amygdala damage. *Nature neuroscience, 16*(3), 270.

Feinstein, J. S., Duff, M. C., & Tranel, D. (2010). Sustained experience of emotion after loss of memory in patients with amnesia. *Proceedings of the National Academy of Sciences*, *107*(17), 7674-7679.

Feldman, L. A. (1995). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of personality and social psychology*, *69*(1), 153.

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature neuroscience, 8*(12), 1643-1646.

Fenske, M. J., & Eastwood, J. D. (2003). Modulation of focused attention by faces expressing emotion: evidence from flanker tasks. *Emotion*, *3*(4), 327.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage, 9*(2), 195-207.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. Mcgraw-Hill Book Company.

Foley, E., Rippon, G., Thai, N. J., Longe, O., & Senior, C. (2012). Dynamic facial expressions evoke distinct activation in the face perception network: a connectivity analysis study. *Journal of cognitive neuroscience, 24*(2), 507-520.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage, 47*, S102.

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18*(12), 1050-1057.

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews neuroscience, 8*(9), 700.

Fox, P. T., Parson, L. M., and Lancaster, J. L. (1998). Beyond the single study: Function/location metanalysis in cognitive neuroimaging. *Current Opinion in Neurobiology, 8*:178–187.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Frijda, N. H. (1993). Moods, emotion episodes, and emotions.

Frijda, N. H. (2017). *The laws of emotion*. Psychology Press.

Frijda, N. H., Ortony, A., Sonnemans, J., & Clore, G. L. (1992). The complexity of intensity: Issues concerning the structure of emotion intensity.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531-534.

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology, 17*(16), R724-R732.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual review of psychology, 63*, 287-313.

Frith, U. (2003). Autism: Explaining the enigma. Blackwell Publishing.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., ... & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current biology*, *19*(7), 573-576.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., ... & Perez, J. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of psychiatry & neuroscience*.

Gainotti, G. (1972). Emotional behavior and hemispheric side of the lesion. *Cortex*, *8*(1), 41-55.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in cognitive sciences, 7*(2), 77-83.

Gallese, V. (2001). The'shared manifold'hypothesis. From mirror neurons to empathy. *Journal of consciousness studies*, *8*(5-6), 33-50.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences, 2*(12), 493-501.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in cognitive sciences*, *8*(9), 396-403.

Garrett, A. S., & Maddock, R. J. (2006). Separating subjective emotion from the perception of emotion-inducing stimuli: an fMRI study. *Neuroimage, 33*(1), 263-274.

Gentili, C., Messerotti Benvenuti, S., Lettieri, G., Costa, C., & Cecchetti, L. (2019). ROI and phobias: The effect of ROI approach on an ALE meta-analysis of specific phobias. *Human brain mapping, 40*(6), 1814-1828.

Gerber, A. J., Posner, J., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., ... & Peterson, B. S. (2008). An affective circumplex model of neural systems subserving valence, arousal, and cognitive overlay during the appraisal of emotional faces. *Neuropsychologia*, *46*(8), 2129-2139.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... & Smith, S. M. (2016). A multi-modal parcellation of human cerebral cortex. *Nature, 536*(7615), 171.

Goldman, A. I. (2006). Simulating minds: The philosophy, psychology, and neuroscience of mindreading. Oxford University Press.

Goleman, D. (2006). *Emotional intelligence*. Bantam.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. Mapping the mind: Domain specificity in cognition and culture, 257.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language, 1*(2), 158-171.

Gray, H. (1966). Anatomy of the Human Body, CM Goss, Ed. *Philadelphia. Lea and Febiger*, *28*, 735-790.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389-400.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *PNAS*, *100*(1), 253-258.

Grezes, J., Pichon, S., & De Gelder, B. (2007). Perceiving fear in dynamic body expressions. *Neuroimage*, *35*(2), 959-967.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & emotion*, *9*(1), 87-108.

Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Do ratings become more extreme with age?. *Behavior Research Methods*, *40*(2), 512-521.

Haak, K. V., Marquand, A. F., & Beckmann, C. F. (2017). Connectopic mapping with resting-state fMRI. *Neuroimage, 170*, 83-94.

Haidt, J. (2003). The moral emotions. *Handbook of affective sciences*, *11*(2003), 852-870.

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in cognitive sciences*, *16*(9), 458-466.

Handjaras, G., Bernardi, G., Benuzzi, F., Nichelli, P. F., Pietrini, P., & Ricciardi, E. (2015). A topographical organization for action representation in the human brain. *Human Brain Mapping, 36*(10), 3832-3844.

Handjaras, G., Leo, A., Cecchetti, L., Papale, P., Lenci, A., Marotta, G., ... & Ricciardi, E. (2017). Modality-independent encoding of individual concepts in the left parietal cortex. *Neuropsychologia, 105*, 39-49.

Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F. R., ... & Stadler, J. (2016). A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Scientific data, 3*, 160092.

Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic representation of numerosity in the human parietal cortex. *Science*, *341*(6150), 1123-1126.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1992). Primitive emotional contagion. *Review of personality and social psychology*, *14*, 151-177.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). Emotional contagion. New York: Cambridge University Press.

Hatfield, E., Rapson, R. L., & Le, Y. C. L. (2011). Emotional contagion and empathy. *The social neuroscience of empathy.*, 19.

Heilman, R. M., Crişan, L. G., Houser, D., Miclea, M., & Miu, A. C. (2010). Emotion regulation and decision making under risk and uncertainty. *Emotion*, *10*(2), 257.

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, *68*(1), 149-160.

Hervé, P. Y., Razafimandimby, A., Jobard, G., & Tzourio-Mazoyer, N. (2013). A shared neural substrate for mentalizing and the affective component of sentence comprehension. *PloS one, 8*(1), e54400.

Hervé, P. Y., Razafimandimby, A., Vigneau, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2012). Disentangling the brain networks supporting affective speech comprehension. *Neuroimage, 61*(4), 1255-1267.

Hillis, A. E. (2014). Inability to empathize: brain lesions that disrupt sharing and understanding another's emotions. Brain, 137(4), 981-997.

Hoffman, M. L. (2008). Empathy and prosocial behavior. *Handbook of emotions, 3*, 440-455.

Hughes, C., & Leekam, S. (2004). What are the links between theory of mind and social relations? Review, reflections and new directions for studies of typical and atypical development. *Social development, 13*(4), 590-619.

Huntenburg, J. M., Bazin, P. L., & Margulies, D. S. (2018). Large-scale gradients in human cortical organization. *Trends in Cogn Sciences, 1*, 21-31.

Hutcherson, C. A., Goldin, P. R., Ochsner, K. N., Gabrieli, J. D., Barrett, L. F., & Gross, J. J. (2005). Attention and emotion: does rating emotion alter neural responses to amusing and sad films?. *Neuroimage, 27*(3), 656-668.

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

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual review of psychology, 60*, 653-670.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS biology, 3*(3), e79.

Izard, C. E. (1984). Emotion-cognition relationships and human development. *Emotions, cognition, and behavior*, 17-37.

Izard, C. E. (1993). Organizational and motivational functions of discrete emotions.

Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research.

Izard, C. E. (2013). *Human emotions*. Springer Science & Business Media.

Jabbi, M., Bastiaansen, J., & Keysers, C. (2008). A common anterior insula representation of disgust observation, experience and imagination shows divergent functional connectivity pathways. *PloS one, 3*(8), e2939.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia, 44*(5), 752-761.

Jacobs, D. H., Shuren, J., Bowers, D., & Heilman, K. M. (1995). Emotional facial imagery, perception, and expression in Parkinson's disease. *Neurology, 45*(9), 1696-1702.

James, W. (1884). What is an emotion? *Mind*, 16, 188-205.

Jonas, J., Jacques, C., Liu-Shuang, J., Brissart, H., Colnat-Coulbois, S., Maillard, L., & Rossion, B. (2016). A face-selective ventral occipitotemporal map of the human brain with intracerebral potentials. *PNAS, 113*(28), E4088-E4097.

Kalbe, E., Schlegel, M., Sack, A. T., Nowak, D. A., Dafotakis, M., Bangard, C., ... & Kessler, J. (2010). Dissociating cognitive from affective theory of mind: a TMS study. *Cortex, 46*(6), 769-780.

Kanske, P., Böckler, A., Trautwein, F. M., & Singer, T. (2015). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy and Theory of Mind. *Neuroimage, 122*, 6-19.

Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G., & Just, M. (2013). A. Identifying emotions on the basis of neural activation. *PloS One, 8*(6), e66032.

Keltner, D., & Kring, A. M. (1998). Emotion, social function, and psychopathology. *Review of General Psychology*, *2*(3), 320-342.

Keltner, D., Ekman, P. (2000). Facial expressions of emotion. In: Haviland-Jones, J.M.,

Kensinger, E. A., & Schacter, D. L. (2006). Processing emotional pictures and words: effects of valence and arousal. *Cognitive Affective & Behavioral Neuroscience,* 6(2), 110-126.

Kim, J., Shinkareva, S. V., & Wedell, D. H. (2017). Representations of modality-general valence for videos and music derived from fMRI data. *NeuroImage*, *148*, 42-54.

Kim, J., Wang, J., Wedell, D. H., & Shinkareva, S. V. (2016). Identifying core affect in individuals from fMRI responses to dynamic naturalistic audiovisual stimuli. *PloS one*, *11*(9), e0161589.

Kipps, C. M., Nestor, P. J., Acosta-Cabronero, J., Arnold, R., & Hodges, J. R. (2009). Understanding social dysfunction in the behavioural variant of frontotemporal dementia: the role of emotion and sarcasm processing. *Brain*, *132*(3), 592-603.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 1.

Klimecki, O. M., Leiberg, S., Ricard, M., & Singer, T. (2013). Differential pattern of functional brain plasticity after compassion and empathy training. *Social cognitive and affective neuroscience, 9*(6), 873-879.

Klimecki, O., & Singer, T. (2012). Empathic distress fatigue rather than compassion fatigue? Integrating findings from empathy research in psychology and social neuroscience. *Pathological altruism*, 368-383.

Klumpers, F., Morgan, B., Terburg, D., Stein, D. J., & van Honk, J. (2014). Impaired acquisition of classically conditioned fear-potentiated startle reflexes in humans with focal bilateral basolateral amygdala damage. *Social Cognitive and Affective Neuroscience*, *10*(9), 1161-1168.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage, 42*(2), 998-1031.

Kohler, C. G., Anselmo-Gallagher, G., Bilker, W., Karlawish, J., Gur, R. E., & Clark, C. M. (2005). Emotion-discrimination deficits in mild Alzheimer disease. *The American journal of geriatric psychiatry, 13*(11), 926-933.

Kohler, C. G., Turner, T. H., Gur, R. E., & Gur, R. C. (2004). Recognition of facial emotions in neuropsychiatric disorders. *CNS spectrums, 9*(4), 267-274.

Krach, S., Cohrs, J. C., de Echeverría Loebell, N. C., Kircher, T., Sommer, J., Jansen, A., & Paulus, F. M. (2011). Your flaws are my pain: Linking empathy to vicarious embarrassment. *PLoS One, 6*(4), e18675.

Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social cognitive and affective neuroscience, 10*(11), 1437-1448.

Kragel, P. A., & LaBar, K. S. (2016). Decoding the nature of emotion in the brain. *Trends in cognitive sciences, 20*(6), 444-455.

Kreibig, S. D. Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394-421 (2010).

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience, 12*(5), 535.

Kucharska-Pietura, K., Phillips, M. L., Gernand, W., & David, A. S. (2003). Perception of emotions from faces and voices following unilateral brain damage. *Neuropsychologia, 41*(8), 1082-1090.

Kunda, Z. (1999). *Social cognition: Making sense of people*. MIT press.

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*(1), 54.

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron, 20*(5), 937-945.

Labs, A., Reich, T., Schulenburg, H., Boennen, M., Mareike, G., Golz, M., ... & Peukmann, A. K. (2015). Portrayed emotions in the movie" Forrest Gump". *F1000Research, 4*.

Laird, J. D., & Bresler, C. (1992). The process of emotional experience: A self-perception theory.

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of cognitive neuroscience, 19*(1), 42-58.

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, *54*(3), 2492-2502.

Lamm, C., Rütgen, M., & Wagner, I.C. (2017). Imaging empathy and prosocial emotions. *Neuroscience letters.*

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological review, 97*(3), 377.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention, 1*, 39-58.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*(3), 261-273.

Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2011). Can people feel happy and sad at the same time?. *Journal of Personality and Social Psychology, 81*(4), 684.

Laukka, P., Elfenbein, H. A., Chui, W., Thingujam, N. S., Iraki, F. K., Rockstuhl, T., & Althoff, J. (2010). Presenting the VENEC corpus: Development of a cross-cultural corpus of vocal emotion expressions and a novel method of annotating emotion appraisals. In *Proceedings of the LREC 2010 Workshop on Corpora for Research on Emotion and Affect* (pp. 53-57). Paris, France: European Language Resources Association.

LeDoux, J. E. (2000). Cognitive-emotional interactions. *Cognitive neuroscience of emotion*, 129-155.

LeDoux, J. E. (2012). Evolution of human emotion: a view through fear. In *Progress in brain research* (Vol. 195, pp. 431-442). Elsevier.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature, 401*(6755), 788.

Lee, T. W., Dolan, R. J., & Critchley, H. D. (2007). Controlling emotional expression: behavioral and neural correlates of nonimitative emotional responses. *Cerebral Cortex, 18*(1), 104-113.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual review of psychology, 66*, 799-823.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in cognitive sciences, 8*(12), 528-533.

Lettieri, G., Handjaras, G., Ricciardi, E., Leo, A., Papale, P., Betta, M., ... & Cecchetti, L. (2019). Emotionotopy in the human right temporo-parietal cortex. *Nature Communications*, in press.

Levenson, R. W. (2011). Basic emotion questions. *Emotion review, 3*(4), 379-386.

Lewis, M., Haviland-Jones, J. M., & Barrett, L. F. (Eds.). (2010). *Handbook of emotions*. Guilford Press.

Lind, S. E., & Bowler, D. M. (2009). Recognition memory, self-other source memory, and theory-of-mind in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 39*(9), 1231.

Lindquist, K. A., & Barrett, L. F. (2012). A functional architecture of the human brain: emerging insights from the science of emotion. *Trends in cognitive sciences, 16*(11), 533-540.

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2015). The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cerebral Cortex, 26*(5), 1910-1922.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and brain sciences*, *35*(3), 121-143.

Lischetzke, T. (2014). Mood. In A. C. Michalos (Ed.), Encyclopedia of quality of life and well-being research (pp. 4115-4120). Dordrecht, Netherlands: Springer.

Liu, Y., Sourina, O., & Nguyen, M. K. (2010, October). Real-time EEG-based human emotion recognition and visualization. In *2010 international conference on cyberworlds* (pp. 262-269). IEEE.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., ... & Baron-Cohen, S. (2010). Shared neural circuits for mentalizing about the self and others. *Journal of cognitive neuroscience, 22*(7), 1623-1635.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 91*, 630.

Lundqvist, L. O. (1995). Facial EMG reactions to facial expressions: A case of facial emotional contagion?. *Scandinavian journal of psychology*, *36*(2), 130-141.

Lundqvist, L. O., & Dimberg, U. (1995). Facial expressions are contagious. *Journal of Psychophysiology*.

Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research,* 9(Nov), 2579-2605.

MacLean, P. D. (1967). The brain in relation to empathy and medical education. *Journal of Nervous and Mental Disease*.

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human brain mapping, 18*(1), 30-41.

Mano, Y., Harada, T., Sugiura, M., Saito, D. N., & Sadato, N. (2009). Perspective-taking as part of narrative comprehension: a functional MRI study. *Neuropsychologia, 47*(3), 813-824.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., ... & Jefferies, E. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *PNAS, 113*(44), 12574-12579.

Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of empathy for ‘social pain’and subsequent prosocial behavior. *Neuroimage, 55*(1), 381-388.

Mather, M., Mitchell, K. J., Raye, C. L., Novak, D. L., Greene, E. J., & Johnson, M. K. (2006). Emotional arousal can impair feature binding in working memory. *Journal of Cognitive Neuroscience, 18*(4), 614-625.

Matsumoto, D. (1992). American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of cross-cultural psychology*, *23*(1), 72-84.

McDonald, S., Flanagan, S., Martin, I., & Saunders, C. (2004). The ecological validity of TASIT: A test of social perception. *Neuropsychological Rehabilitation*, *14*(3), 285-302.

McDonald, S., Flanagan, S., Rollins, J., & Kinch, J. (2003). TASIT: A new clinical tool for assessing social perception after traumatic brain injury. *The Journal of head trauma rehabilitation*, *18*(3), 219-238.

Mehrabian, A. (2000). Manual for the balanced emotional empathy scale. *Unpublished manuscript, Available from Albert Mehrabian, 1130*.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy 1. *Journal of personality*, *40*(4), 525-543.

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. the MIT Press.

Mesquita, B., & Walker, R. (2003). Cultural differences in emotions: A context for interpreting emotional experiences. *Behaviour Research and Therapy, 41*(7), 777-793.

Mitchell, R. L., & Phillips, L. H. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia, 70*, 1-10.

Morelli, S. A., & Lieberman, M. D. (2013). The role of automaticity and attention in neural processes underlying empathy for happiness, sadness, and anxiety. *Frontiers in Human Neuroscience, 7*, 160.

Morelli, S. A., Rameson, L. T., & Lieberman, M. D. (2012). The neural components of empathy: predicting daily prosocial behavior. *Social Cognitive and Affective Neuroscience, 9*(1), 39-47.

Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *PNAS, 96*(4), 1680-1685.

Mourao-Miranda, J., Volchan, E., Moll, J., de Oliveira-Souza, R., Oliveira, L., Bramati, I., ... & Pessoa, L. (2003). Contributions of stimulus valence and arousal to visual activation during emotional perception. *Neuroimage, 20*(4), 1955-1963.

Murphy, F. C., Nimmo-Smith, I. A. N., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience, 3*(3), 207-233.

Namburi, P., Al-Hasani, R., Calhoon, G. G., Bruchas, M. R., & Tye, K. M. (2016). Architectural representation of valence in the limbic system. *Neuropsychopharmacology, 41*(7), 1697.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage, 56*(2), 400-410.

Nummenmaa, L., & Saarimäki, H. (2017). Emotions as discrete patterns of systemic activity. *Neuroscience Letters*.

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. (2014). Bodily maps of emotions. *PNAS, 111*(2), 646-651.

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., & Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *PNAS, 109*(24), 9599-9604.

Nummenmaa, L., Hirvonen, J., Parkkola, R., & Hietanen, J. K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *Neuroimage, 43*(3), 571-580.

Nussbaum, M. C. (2003). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145-175.

Oosterwijk, S., Lindquist, K.A., Anderson, E.C., Dautoff, R., Moriguchi, Y., Barrett, L.F. (2012). States of mind: emotions, body feelings, and thoughts share distributed neural networks. *Neuroimage, 62*, 2110–28.

Ortony, A., Clore, G. L., & Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.

Panksepp, J. (1982). Toward a general psychobiological theory of emotions. *Behavioral and Brain sciences*, *5*(3), 407-422.

Parkinson, B., Totterdell, P., Briner, R. B., & Reynolds, S. A. (1996). *Changing moods: The psychology of mood and mood regulation*. Longman.

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience, 30*(30), 10127-10134.

Pelphrey, K. A., Morris, J. P., & Mccarthy, G. (2004). Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of cognitive neuroscience, 16*(10), 1706-1716.

Pessoa, L. (2013). *The cognitive-emotional brain: From interactions to integration*. MIT press.

Peters, E., Västfjäll, D., Gärling, T., & Slovic, P. (2006). Affect and decision making: A "hot" topic. *Journal of Behavioral Decision Making, 19*(2), 79-85.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage, 16*(2), 331-348.

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological science, 17*(4), 292-299.

Phelps, E. A., O'Connor, K. J., Gatenby, J. C., Gore, J. C., Grillon, C., & Davis, M. (2001). Activation of the left amygdala to a cognitive representation of fear. *Nature neuroscience, 4*(4), 437.

Philippot, P. (1993). Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and emotion, 7*(2), 171-193.

Plutchik, R. (1987). Evolutionary bases of empathy. *Empathy and its development, 1*, 38-46.

Plutchik, R., & Kellerman, H. (Eds.). (2013). *Theories of emotion* (Vol. 1). Academic Press.

Poletti, M., Enrici, I., Adenzato, M. (2012). Cognitive and affective Theory of Mind in neurodegenerative diseases: Neuropsychological, neuroanatomical and neurochemical levels. *Neuroscience and Biobehavioral Reviews*, *36*, 2147-2164.

Posner, J., Russell, J. A., Gerber, A., Gorman, D., Colibazzi, T., Yu, S., ... & Peterson, B. S. (2009). The neurophysiological bases of emotion: An fMRI study of the affective circumplex using emotion-denoting words. *Human brain mapping*, *30*(3), 883-895.

Posner, J., Russell, J., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17*(3), 715–734.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(4), 515-526.

Preston, S. D., & De Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and brain sciences*, *25*(1), 1-20.

Preti, A., Vellante, M., Baron-Cohen, S., Zucca, G., Petretto, D. R., & Masala, C. (2011). The Empathy Quotient: A cross-cultural comparison of the Italian version. *Cognitive neuropsychiatry*, *16*(1), 50-70.

Price, D. D. (2000). Psychological and neural mechanisms of the affective dimension of pain. *Science, 288*(5472), 1769-1772.

Radua, J., Sarró, S., Vigo, T., Alonso-Lana, S., Bonnín, C. M., Ortiz-Gil, J., ... & Salvador, R. (2014). Common and specific brain responses to scenic emotional stimuli. *Brain Structure and Function, 219*(4), 1463-1472.

Raz, G., Jacob, Y., Gonen, T., Winetraub, Y., Flash, T., Soreq, E., & Hendler, T. (2013). Cry for her or cry with her: context-dependent dissociation of two modes of cinematic empathy reflected in network cohesion dynamics. *Social cognitive and affective neuroscience*, *9*(1), 30-38.

Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., ... & Maron-Katz, A. (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(4), 709-723.

Raz, G., Winetraub, Y., Jacob, Y., Kinreich, S., Maron-Katz, A., Shaham, G., ... & Hendler, T. (2012). Portraying emotions at their unfolding: a multilayered approach for probing dynamics of neural networks. *Neuroimage*, *60*(2), 1448-1461.

Résibois, M., Verduyn, P., Delaveau, P., Rotgé, J. Y., Kuppens, P., Van Mechelen, I., & Fossati, P. (2017). The neural basis of emotions varies over time: different regions go with onset-and offset-bound processes underlying emotion intensity. *Social cognitive and affective neuroscience, 12*(8), 1261-1271.

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage,* 61(4), 1402-1418.

Rice, G. E., Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. *Journal of Neuroscience, 34*(26), 8837-8844.

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature communications, 9*(1), 1027.

Richell, R. A., Mitchell, D. G. V., Newman, C., Leonard, A., Baron-Cohen, S., & Blair, R. J. R. (2003). Theory of mind and psychopathy: can psychopathic individuals read the 'language of the eyes'?. *Neuropsychologia, 41*(5), 523-526.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review Neuroscience, 27*, 169-192.

Rizzolatti, G., & Caruana, F. (2017). Some considerations on de Waal and Preston review. *Nature Reviews Neuroscience*, *18*(12), 769.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature reviews neuroscience*, *2*(9), 661.

Rizzolatti, G., Fabbri-Destro, M., & Cattaneo, L. (2009). Mirror neurons and their clinical relevance. *Nature Reviews Neurology, 5*(1), 24.

Rosenberg, E. L., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion, 19*(2), 111-138.

Rousseeuw, P. J., & Kaufman, L. Finding groups in data. *Hoboken: Wiley Online Library* (1990).

Russell, J. (2017). A. Mixed emotions viewed from the psychological constructionist perspective. *Emotion Review 9*(2), 111-117.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review, 110*(1), 145.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of autism and developmental disorders, 32*(3), 189-194.

Saarimäki, H., Ejtehadian, L. F., Glerean, E., Jääskeläinen, I. P., Vuilleumier, P., Sams, M., & Nummenmaa, L. (2018). Distributed affective space represents multiple emotion categories across the human brain. *Social cognitive and affective neuroscience, 13*(5), 471-482.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., ... & Nummenmaa, L. (2015). Discrete neural signatures of basic emotions. *Cerebral cortex*, *26*(6), 2563-2573.

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., ... & Jeffries, J. (2011). Emotional perception: meta-analyses of face and natural scene processing. *Neuroimage*, *54*(3), 2524-2533.

Sachs, G., Steger-Wuchse, D., Kryspin-Exner, I., Gur, R. C., & Katschnig, H. (2004). Facial recognition deficits and cognition in schizophrenia. *Schizophrenia research*, *68*(1), 27-35.

Sadaghiani, S., Poline, J. B., Kleinschmidt, A., & D'Esposito, M. (2015). Ongoing dynamics in large-scale functional connectivity predict perception. *PNAS, 112*(27), 8463-8468.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences*, *14*(4), 303-316.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". *Neuroimage, 19*(4), 1835-1842.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia, 43*(10), 1391-1399.

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in cognitive sciences*, *19*(2), 65-72.

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

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and emotion*, *23*(7), 1307-1351.

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.

Schneider, I. K., Veenstra, L., van Harreveld, F., Schwarz, N., & Koole, S. L. (2016). Let's not be indifferent about neutrality: Neutral ratings in the International Affective Picture System (IAPS) mask mixed affective responses. *Emotion*, *16*(4), 426.

Schreiber, T., & Schmitz, A. (1996). Improved Surrogate Data for Nonlinearity Tests. *Physical Review Letters,* 77(4), 635-638.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews, 42*, 9-34.

Sebastian, C. L., Fontaine, N. M., Bird, G., Blakemore, S. J., De Brito, S. A., McCrory, E. J., & Viding, E. (2011). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social cognitive and affective neuroscience*, *7*(1), 53-63.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., ... & Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889-893.

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The animacy continuum in the human ventral vision pathway. *Journal of cognitive neuroscience*, *27*(4), 665-678.

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*, *45*(13), 3054-3067.

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*(3), 617-627.

Shamay-Tsoory, S. G., Tibi-Elhanany, Y., & Aharon-Peretz, J. (2006). The ventromedial prefrontal cortex is involved in understanding affective but not cognitive theory of mind stories. *Social neuroscience*, *1*(3-4), 149-166.

Shaw, R. J., Dong, M., Lim, K. O., Faustman, W. O., Pouget, E. R., & Alpert, M. (1999). The relationship between affect expression and affect recognition in schizophrenia. *Schizophrenia Research*, *37*(3), 245-250.

Silberman, E. K., & Weingartner, H. (1986). Hemispheric lateralization of functions related to emotion. *Brain and cognition*, *5*(3), 322-353.

Silvers, J. A., Weber, J., Wager, T. D., & Ochsner, K. N. (2014). Bad and worse: neural systems underlying reappraisal of high-and low-intensity negative emotions. *Social Cognitive and Affective Neuroscience*, *10*(2), 172-179.

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature communications, 7*, 12141.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81-96.

Singer, T., Seymour, B., O'doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*(7075), 466.

Singer, T., Seymour, B., O'doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157-1162.

Skerry, A. E., & Saxe, R. (2015). Neural representations of emotion are organized around abstract event features. *Current Biology 25*(15), 1945-1954.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology 48*(4), 813.

Smith, M. (1995). *Engaging characters: Fiction, emotion, and the cinema*. Oxford: Clarendon Press.

Solomon, R. C. (2007). *Not passion's slave: emotions and choice*. Oxford University Press.

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *Neuroimage, 35*(3), 1378-1384.

Spunt, R. P., & Adolphs, R. (2017). The neuroscience of understanding the emotions of others. *Neuroscience letters*.

Srinivasan, R., Golomb, J. D., & Martinez, A. M. (2016). A neural basis of facial action recognition in humans. *Journal of Neuroscience, 36*(16), 4434-4442.

Stephens, C. L., Christie, I. C., & Friedman, B. H. (2010). Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis. *Biological Psychology 84*(3), 463-473.

Stevenson, R. A., Mikels, J. A., & James, T. W. (2007). Characterization of the affective norms for English words by discrete emotional categories. *Behavior research methods, 39*(4), 1020-1024.

Straube, T., Weisbrod, A., Schmidt, S., Raschdorf, C., Preul, C., Mentzel, H. J., & Miltner, W. H. (2010). No impairment of recognition and experience of disgust in a patient with a right-hemispheric lesion of the insula and basal ganglia. *Neuropsychologia, 48*(6), 1735-1741.

Telzer, E. H., Fuligni, A. J., Lieberman, M. D., Miernicki, M. E., & Galván, A. (2014). The quality of adolescents' peer relationships modulates neural sensitivity to risk taking. S*ocial cognitive and affective neuroscience, 10*(3), 389-398.

Thornton, M. A., & Tamir, D. I. (2017) Mental models accurately predict emotion transitions. *PNAS 114*(23), 5982-5987.

Touroutoglou, A., Lindquist, K. A., Dickerson, B. C., & Barrett, L. F. (2015). Intrinsic connectivity in the human brain does not reveal networks for 'basic'emotions. *Social Cognitive and Affective Neuroscience, 10*(9), 1257-1265.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: a review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3*(4), 397-405.

Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in everyday life. *PloS one 10*(12), e0145450.

Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J., & Coalson, T. (2011). Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral Cortex 22*(10), 2241-2262.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping 30*(3), 829-858.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage, 48*(3), 564-584.

Vaughan, K. B., & Lanzetta, J. T. (1980). Vicarious instigation and conditioning of facial expressive and autonomic responses to a model's expressive display of pain. *Journal of personality and social psychology, 38*(6), 909.

Verduyn, P., Delaveau, P., Rotgé, J. Y., Fossati, P., & Van Mechelen, I. (2015). Determinants of emotion duration and underlying psychological and neural mechanisms. *Emotion Review, 7*(4), 330-335.

Völlm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., ... & Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *Neuroimage, 29*(1), 90-98.

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of cognitive neuroscience, 22*(12), 2864-2885.

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience, 23*(12), 4057-4066.

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS computational biology, 11*(4), e1004066.

Wager, T. D., Phan, K. L., Liberzon, I., & Taylor, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *Neuroimage, 19*(3), 513-531.

Walter, H. (2012). Social cognitive neuroscience of empathy: concepts, circuits, and genes. *Emotion Review*, *4*(1), 9-17.

Wang, Y. G., Wang, Y. Q., Chen, S. L., Zhu, C. Y., & Wang, K. (2008). Theory of mind disability in major depression with or without psychotic symptoms: a componential view. *Psychiatry research, 161*(2), 153-161.

Waugh, C. E., & Schirillo, J. A. (2012). Timing: a missing key ingredient in typical fMRI studies of emotion. *Behavioral and Brain Sciences, 35*(3), 170-171.

Waugh, C. E., Shing, E. Z., & Avery, B. M. (2015). Temporal dynamics of emotional processing in the brain. *Emotion Review, 7*(4), 323-329.

Weiss, E. M., Kohler, C. G., Brensinger, C. M., Bilker, W. B., Loughead, J., Delazer, M., & Nolan, K. A. (2007). Gender differences in facial emotion recognition in persons with chronic schizophrenia. *European psychiatry, 22*(2), 116-122.

Weiss, E. M., Kohler, C. G., Vonbank, J., Stadelmann, E., Kemmler, G., Hinterhuber, H., & Marksteiner, J. (2008). Impairment in emotion recognition abilities in patients with mild cognitive impairment, early and moderate Alzheimer disease compared with healthy comparison subjects. *The American journal of geriatric psychiatry, 16*(12), 974-980.

Wellman, H. M. (1992). The child's theory of mind. The MIT Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child development, 72*(3), 655-684.

Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., Wright, C. I., & Rauch, S. L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion, 1*(1), 70.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655-664.

Wild, B., Erb, M., Eyb, M., Bartels, M., & Grodd, W. (2003). Why are smiles contagious? An fMRI study of the interaction between perception of facial affect and facial movements. *Psychiatry Research: Neuroimaging, 123*(1), 17-36.

Williams, J. H., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Reviews, 25*(4), 287-295.

Williams, L. M., Phillips, M. L., Brammer, M. J., Skerrett, D., Lagopoulos, J., Rennie, C., ... & Gordon, E. (2001). Arousal dissociates amygdala and hippocampal fear responses: evidence from simultaneous fMRI and skin conductance recording. *Neuroimage, 14*(5), 1070-1079.

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2014). Variety in emotional life: within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks. *Social cognitive and affective neuroscience, 10*(1), 62-71.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods 8*(8), 665.

Yarrow, S., Razak, K. A., Seitz, A. R., & Seriès, P. (2014) Detecting and quantifying topography in neural maps. *PloS one 9*(2), e87178.

Yim, J., Babbage, D. R., Zupan, B., Neumann, D., & Willer, B. (2013). The relationship between facial affect recognition and cognitive functioning after traumatic brain injury. *Brain Injury*, *27*(10), 1155-1161.