SISSA
Scuola
Internazionale
Superiore di
Studi Avanzati

Neuroscience Area – PhD course

in Cognitive Neuroscience

The Role of Personality in Shaping Neural Processes Associated with Theory of Mind

Candidate:
Elisabetta Pisanu

Advisor:
Prof. Raffaella Rumiati
Co-advisors:
Prof. Giorgia Silani
Prof. Alessandro Grecucci

Academic Year 2022-23

To Giorgio.

Acknowledgements

As I conclude this demanding yet stimulating journey, I would like to express my gratitude to the people who have played a significant role in my personal and professional development.

First of all, I wish to thank my supervisor, prof. Raffaella Rumiati, for pushing me towards self-confidence, and for giving me the opportunity to grow in terms of independence and scientific thinking always offering encouragement, support, and precious feedback.

I want to give a warm thank you to Dr. Marilena Aiello for introducing me to research during my SISSA internship and for her ongoing advices and support throughout these years.

I am also deeply thankful to my lab colleagues Maristella and Tania for their consistent and invaluable support during these years and to Sandra for her contributions to my projects. Within the SISSA community, I would like to acknowledge the crucial technical assistance provided by Marco and Alessio. Moreover, I am grateful to Marco for engaging in thought-provoking discussions that have greatly benefited my work.

I would also like to express my heartfelt thanks to Valeria, a fellow SISSA colleague and, above all, a cherished friend. Throughout these years, she has been a source of unwavering support, and we have shared the joys and challenges of this journey.

I also wish to thank Prof. Giorgia Silani and Prof. Alessandro Grecucci for the time spent reading and reviewing my thesis.

To my family, especially my parents, who have stood by my side in this journey as in all aspects of life, celebrating my achievements and providing solace during difficult times, I offer my heartfelt thanks. I also thank Tiziano, for having been both my demanding critic and my caring partner and for sticking by me despite all my sudden mood swings and breakdowns.

Lastly, I would like to acknowledge and thank the patients who participated in my last study. Despite the challenges they faced during the experiment, they demonstrated remarkable motivation, commitment, and patience. Their willingness to engage and help me understand their perspectives not only enhanced my professional and clinical growth but also enriched my personal development.

Index

Abstract **11**

CHAPTER 1 - General introduction **13**

1.1 Personality Neuroscience **15**

1.1.1 Neurobiology behind The Five Factor Model *15*

1.1.2 Agreeableness *20*

1.2 Theory of Mind **24**

1.2.1 An umbrella term: from lower to higher processes *24*

1.2.2 Mental state decoding *26*

1.2.3 Sex differences *28*

1.3 Autism Spectrum Disorder (ASD) **28**

1.3.1 Big Five personality traits in ASD *28*

1.3.2 The "Extreme Male Brain" theory *30*

1.3.3 The social motivational theory *32*

1.4 The aims of the thesis and outline **34**

CHAPTER 2 - Study 1: Representation of social content in dorsomedial prefrontal cortex underlies individual differences in agreeableness trait **37**

Abstract **38**

2.1 Introduction **38**

2.2 Methods **41**

2.2.1 Participants recruitment and personality assessment *41*

2.2.2 Social cognition task *42*

2.2.3 MRI acquisition *42*

2.2.4 MRI quality assessment *43*

2.2.5 MRI preprocessing *43*

2.2.6 Behavioral data analyses *44*

2.2.7 fMRI data analyses *45*

First level GLM analysis *45*

Group level GLM analysis *45*

Representational similarity analysis *46*

Univariate analysis *47*

Meta-analytic decoding *47*

2.3 Results **47**

2.3.1 Behavioral results *47*
2.3.2 Neuroimaging results *48*
Whole-brain group results *48*
Representational similarity results *49*
Univariate results *51*
Meta-analytic decoding results *52*
2.4 Discussion **52**
2.5 Conclusion **57**

CHAPTER 3 - Study 2: Personality and sex differences in social interaction perception: an ERP and source imaging study **59**
Abstract **60**
3.1 Introduction **60**
3.2 Methods **63**
3.2.1 Participants recruitment and personality assessment *63*
3.2.2 Social interaction perception task *63*
3.2.3 EEG recording and preprocessing *65*
3.2.4 Behavioral data analysis *66*
3.2.5 ERP analysis *67*
First-level GLM analysis *67*
Second level GLM analysis *67*
3.2.6 Source analysis *68*
3.3 Results **69**
3.3.1 Behavioral results *69*
3.3.2 ERP results *70*
3.3.3 Source Localization *77*
3.4 Discussion **78**
3.5 Conclusion **83**

CHAPTER 4 - Study 3: Agreeableness modulates mental state decoding: Electrophysiological evidence **85**
Abstract **86**
4.1 Introduction **86**
4.2 Methods **89**
4.2.1 Participants recruitment and personality assessment *89*
4.2.2 Biological (e)motion task *90*
4.2.3 Reading the mind in the eyes task *91*

4.2.4 EEG recording and preprocessing *93*
4.2.5 Behavioral data analysis *94*
4.2.6 ERP analysis *94*
First-level GLM analysis *95*
Second level GLM analysis *95*
4.2.7 ERSP analysis *96*
First-level GLM analysis *96*
Second level GLM analysis *97*
4.2.8 Source analysis *97*
4.3 Results **98**
4.3.1 Behavioral results *98*
4.3.2 ERP results *99*
4.3.3 ERSP results *108*
4.3.4 Source Localization *109*
4.4 Discussion **110**
4.5 Conclusion **114**

CHAPTER 5 - Study 4: Does ASD individuals' personality affect the decoding of mental states? **117**
Abstract **118**
5.1 Introduction **118**
5.2 Methods **122**
5.2.1 Participants recruitment *122*
5.2.2 Procedure and experimental tasks *123*
5.2.3 EEG recording and preprocessing *123*
5.2.4 Behavioral data analysis *123*
5.2.5 ERP analysis *125*
First-level GLM analysis *125*
Second level GLM analysis *125*
5.2.6 ERSP analysis *126*
First-level GLM analysis *126*
Second level GLM analysis *126*
5.3 Results **126**
5.3.1 Behavioral results *126*
5.3.2 ERP results *130*
5.3.3 ERSP results *136*

5.4 Discussion **137**
5.5 Conclusion **140**
CHAPTER 6 - General discussion **143**
REFERENCES **155**
APPENDIX - Supplementary materials of Chapter 2 ..... **209**

Abstract

Personality accounts for a significant amount of individual differences in people's academic and life outcomes. In my dissertation, I focused on agreeableness, a personality trait from the Five Factor Model describing individual differences in altruism, empathy, and Theory of Mind (ToM) (Allen et al., 2017). Thus, my primary aim was to investigate whether and how agreeableness might modulate social and emotional processes associated with ToM in the general population and in young adults with Autism Spectrum Disorder (ASD) without language and intellectual impairments. All these aspects are introduced in Chapter 1 of my thesis, while the original experimental studies are described in the following Chapters (2-5).

In Chapter 2, I report an fMRI study (Study 1) exploring whether neural variations during social content encoding are related to differences in agreeableness by assessing the degree of similarity between patterns of activation for social and non-social content across individuals. Results suggested that more agreeable individuals encode social and non-social information more distinctively compared with less agreeable individuals, whose encoding of the two contents is more similar in the dorso-medial prefrontal cortex (dmPFC), a brain region typically linked with ToM abilities (Saxe and Baron-Cohen, 2006). This result suggests that individual differences in agreeableness are associated with differences in processing social information.

Based on this finding, I developed three studies that aimed at characterizing the neural correlates of agreeableness and different aspects of ToM by means of electroencephalography (EEG). This approach allowed me to identify at which stage of the neural processing agreeableness kicks in, in order to distinguish the impact on early, perceptual processes from slower, inferential processing. Since ToM involves distinct processes, I first set out to investigate the social (Study 2) and emotional dimensions (Study 3) separately; then I examined the same processes within a

group of individuals within the autism spectrum (Study 4). Thus Chapter 3 contains Study 2 in which I assessed the impact of agreeableness on social interaction perception. Results revealed agreeableness allows a deeper processing of social stimuli compared to non-social ones, mainly in fronto-parietal areas. This relatively early effect in the whole sample and in females was more massive and extensive in males and translated into behavior as a tendency towards a better ability to discriminate social interactions. Chapter 4 contains Study 3 in which I investigated the modulation of agreeableness on the decoding of emotional cues by means of two tasks with different degrees of complexity spanning from the perception of basic emotional actions to the recognition of complex mental states. Results showed that agreeableness affects brain's activity only during mental state decoding in male participants, mainly involving the ventro-mPFC, while in females this modulation seems to be negligible. Study 4, in Chapter 5, investigates the personality of young adults with ASD without language and intellectual impairment, and whether agreeableness might differentially modulate the same neural processes investigated in the previous two EEG studies in this population. This investigation underscored that while young adults with this diagnosis may exhibit lower scores in all the five traits, they do not uniformly fall into the extreme ends of all personality dimensions. Results further highlighted the role of agreeableness in shaping socio-emotional processing, particularly in the context of ToM tasks. Notably, ASD individuals showcased a capacity for compensatory mechanisms, achieving behavioral performance akin to less agreeable healthy controls despite differences in neural responses.

Overall, the findings of my thesis advance our understanding of the neural determinants of agreeableness and its significance in the social domain. This has potential implications for clinical interventions by tailoring treatments to individual variations, thus enhancing effective therapeutic strategies. In the final Chapter 6, I contextualize these findings within the existing literature and I discuss their implications.

CHAPTER 1

General introduction

Imagine the challenge of summarizing your personality within a few succinct lines. In attempting to do so, a multitude of descriptive adjectives would likely come to mind. For instance, Clara may be characterized as a very friendly and extroverted individual, driven by a passion for art and adventure, yet lacking in organizational and planning skills. These characteristics exert a profound influence on several aspects of Clara's life, encompassing academic achievements, occupational pursuits, well-being, interpersonal relationships, and overall life satisfaction.

The Five-Factor Model (FFM) of personality commonly referred to as the Big Five, conceptualizes these characteristics as *traits*, namely enduring internal dispositions that generally exhibit stability across time and situations. Therefore, it has become evident that personality and individual differences play a crucial role in a wide range of research domains aimed at comprehending the complexities of human nature. However, a new awareness has been reached that relying solely on observable behavior is insufficient. As a result, the field of personality neuroscience has emerged, seeking to provide a more comprehensive understanding of the neural underpinnings associated with these traits and their influence on our brain.

Through this thesis, my objective is to contribute to this investigation that is still in its infancy but holds immense potential and importance. In the following paragraphs I will provide an overview of the emerging field of personality neuroscience with particular reference to the FFM. Specifically, I will delve into the trait of agreeableness, which serves as the central theme and *fil rouge* connecting all my studies, and then I will present the Theory of Mind (ToM), which is closely associated with agreeableness, and its distinct social and emotional processes. In particular, I will pause on the *decoding of mental states* as the primary process investigated in relation to agreeableness. In the concluding part of this general introduction, I will provide an overview of the current understanding of individual differences in the Autism Spectrum Disorder (ASD), with reference to the FFM and ToM. Specifically, I will introduce to two influential theories namely the *Extreme Male Brain Theory* by Baron-Cohen and *Social Motivation Theory* by Chevallier that have been put forward to account for the challenges and differences in

processing social and non-social stimuli by people with ASD.

1.1 Personality Neuroscience

1.1.1 Neurobiology behind The Five Factor Model

Personality neuroscience is an interdisciplinary field that investigates the neural basis of human personality traits, aiming to understand how individual differences in personality arise from and are influenced by underlying brain processes. Moreover, personality neuroscience investigates the genetic and environmental factors that interact with brain processes to shape personality traits. By integrating findings from genetics, neuroscience, and psychology, personality neuroscience aimed at offering a comprehensive approach to unraveling the complex interplay between biology and culture, advancing our understanding of human nature and individual differences (Allen and DeYoung, 2017; DeYoung, 2010). DeYoung suggests, "the relationship between personality psychology and neuroscience should be viewed as a two-way street" (2010, p. 1176) with personality psychology guiding neuroscience hypotheses and organizing findings, while neuroscience data influencing the development of trait models. Exploiting neuroscience methods to study personality has the potential to create explanatory biological models for previously descriptive trait taxonomies, contributing to a comprehensive theory of personality as a dynamic system generating behavior and experience. However, it is a prestigious yet highly complex and challenging goal to achieve, primarily due to the specific methodological constraints that have often been disregarded, resulting in difficulties in interpreting and generalizing the results. Consequently, further research in this field requires extensive additional studies, along with the inclusion of meta-analyses, to shed light on, synthesize, and integrate the emerging findings.

Central to its original program, neuroscientists interested in personality adopted the Five Factor Model (FFM) or Big Five as a theoretical framework. This is the most widespread and used model of personality which provides a comprehensive taxonomy of traits, typically labeled Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness/Intellect. The model

is the result of an accurate lexical and psychometric research, which identified these five broad dimensions of personality by applying a factor analysis to the trait-related verbal descriptions and evaluations that people use to describe themselves and others (Digman, 1997; John et al., 2008; Markon et al., 2005). Traits are defined as probabilistic descriptions of the intensity and frequency with which individuals display different behavioral, motivational, emotional, and cognitive states. Each trait has two extremes poles and each person positions herself between these two poles according to her own individuality. Individuals with high levels of one trait will experience the states associated with that trait more often and more intensely than individuals with low levels of the same trait (Costa and McCrae, 1992a, 2019; Goldberg, 1990; John et al., 2008).

As represented in Figure 1.1, the FFM provides a framework for understanding dispositional traits as pivotal, biologically grounded tendencies that manifest as characteristic adaptations within a culturally contextualized framework. Characteristic adaptations encompass several aspects such as goals, strivings, attitudes, as well as the self-concept, which incorporates self-schemas and the life story. Noteworthy, the FFM imposes a reciprocity between characteristic adaptations and external influences, mediated by objective biography, while the influence of basic tendencies on characteristic adaptations and self-concept is modelled as unidirectional. Therefore, basic tendencies, that generally respond to maturational forces, can also exhibit shifts in response to external factors, only in the case these factors affect the biological foundations of personality (Costa and McCrae, 2019).

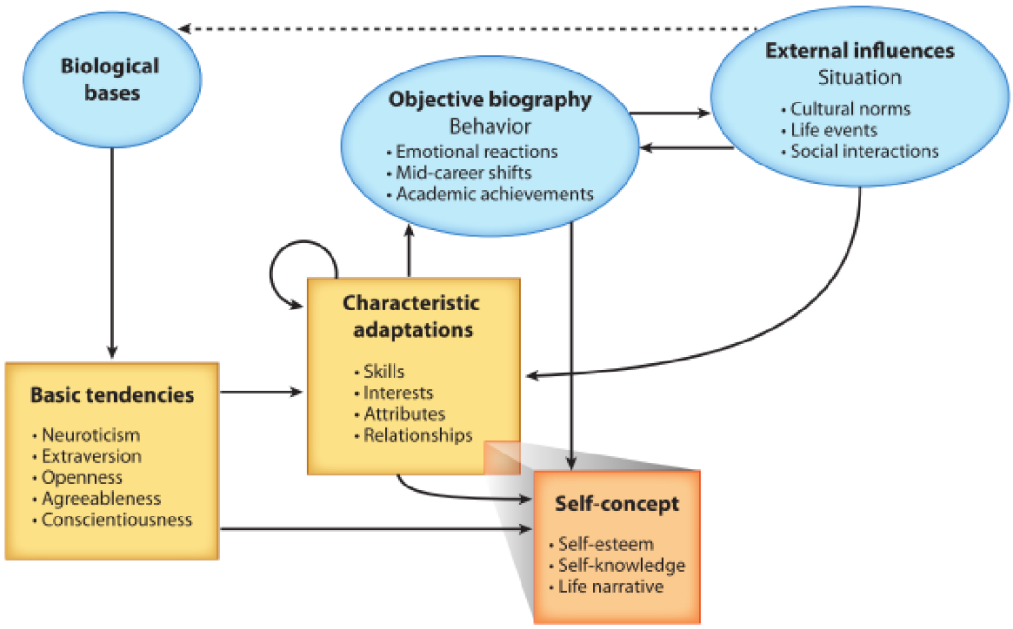


Figure 1.1. The figure represents the system of the FFM. Main components are represented within the rectangles, while in ovals are depicted the factors interacting with them. Dashed arrow represents pathways outside the system. Adapted from Costa and McCrae (2019).

DeYoung and Blain (2020) conducted an extensive literature review encompassing studies aimed at unraveling the complex neural underpinnings of the five personality traits. Presented herewith is a concise summary of the primary findings (Figure 1.2).

Extraversion is associated with increased activation and connectivity in reward-related brain regions, such as the ventral striatum and prefrontal cortex (PFC). High extraverts individuals tend to show greater responsiveness to positive stimuli and exhibit a propensity for approach-related behaviors (Allen and DeYoung, 2017; DeYoung, 2010). Accordingly, dopamine plays a pivotal role in sensitivity to rewards and drives behavior aimed at approaching potential rewards (Depue and Collins, 1999). Functional MRI (fMRI) studies have shown that extraversion correlates positively with brain activity during rest or in response to positive/rewarding stimuli in regions such as the medial orbitofrontal cortex (OFC), nucleus accumbens, amygdala, and striatum (Canli et al., 2001, 2002; Cohen et al., 2005; Mobbs et al., 2005). Additionally, an association between extraversion and larger volume in the medial OFC has been found by several structural MRI studies (DeYoung et al., 2010; Omura et al., 2005; Rauch et al., 2005) although other studies

were unable to replicate these results (Liu et al., 2013; Riccelli et al., 2017).

Neuroticism refers to the inclination to experience negative emotions and thoughts associated with threat and punishment, such as anxiety, depression, anger, and vulnerability. Neuroimaging studies have revealed that Neuroticism correlates with brain activity in regions like the amygdala, insula, and anterior cingulate cortex (ACC) during rest or in response to aversive or novel stimuli (Cools et al., 2005; Deckersbach et al., 2006; Eisenberger et al., 2005; Etkin et al., 2004; Haas et al., 2007a; Keightley et al., 2003; Reuter et al., 2004). Additionally, neuroticism has been linked to neural activity in the medial PFC (Haas et al., 2007a; Williams et al., 2006), indicating difficulties in emotion regulation, as well as reduced volume in that region (DeYoung et al., 2010). Neuroticism has been associated with higher baseline cortisol levels, indicating chronic stress, but reduced cortisol responses to specific stressors (Netter, 2004), suggesting difficulties in mobilizing resources to cope with stressful situations. Lower serotonergic function (Cools et al., 2005; Hennig, 2004; Schinka et al., 2004; Sen et al., 2004; Tauscher et al., 2001) and higher levels of norepinephrine (Hennig, 2004; White and Depue, 1999; Zuckerman, 2005) have been also linked to neuroticism.

Conscientiousness is a personality trait that reflects individual differences in self-discipline and organizational skills required for top-down control. Highly conscientious individuals are more likely to achieve academic and occupational success and engage in behaviors that promote health and longevity (Ozer and Benet-Martinez, 2006; Roberts et al., 2017). The PFC, responsible for complex rule-following and planning abilities (Bunge and Zelazo, 2006), has been associated with conscientiousness (DeYoung et al., 2010).

The two components of *Openness/Intellect* can be described as Intellect and Openness to Experience, reflecting engagement with intellectual and sensory information (artistic and aesthetic interests), respectively. These traits are equally fundamental and are associated with attentional processes and the ability to handle complex information (Allen and DeYoung, 2017). Intellect has a stronger correlation with intelligence and working memory compared to openness (DeYoung,

2020; DeYoung et al., 2009, 2014). People with high levels of Openness/Intellect are creative, curious, innovative, imaginative. A neuroimaging study revealed that only intellect was associated with brain activity in the left frontal pole and the posterior medial FC during a challenging working memory task (DeYoung et al., 2009). The frontal pole is involved in abstract integration and drawing analogies (Gilbert et al., 2006; Green et al., 2006; Ramnani and Owen, 2004), while the posterior medial FC is responsible for monitoring goal-directed performance and detecting errors in cognitive tasks (Brown and Braver, 2005; Ridderinkhof et al., 2004).

In the following section I will dedicate more attention to *Agreeableness* as this trait frequently interacts with social processes.

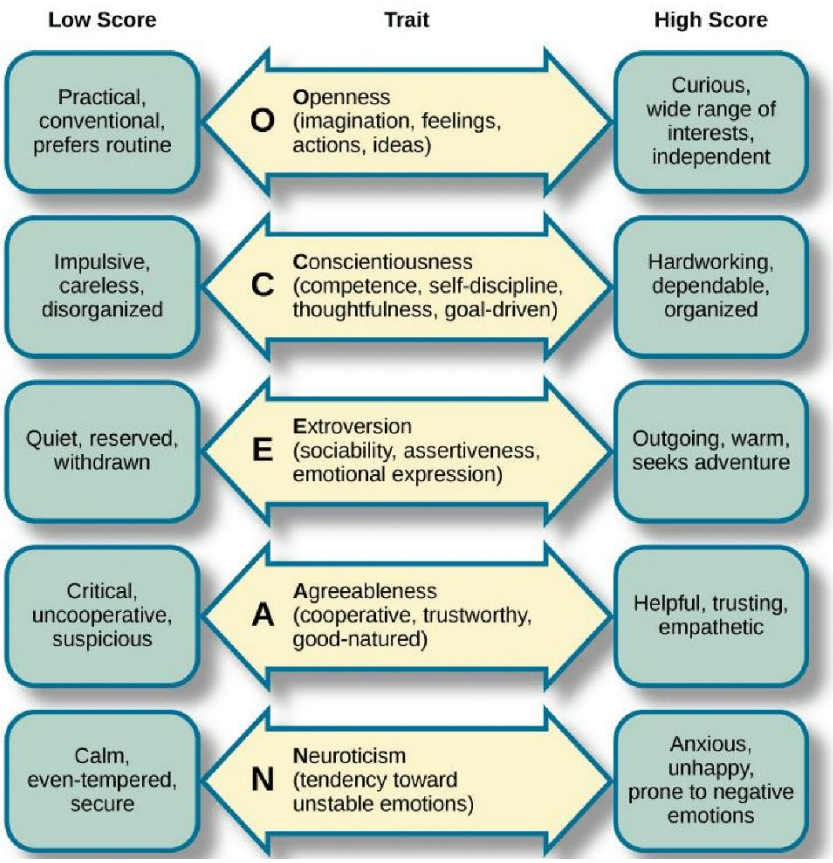


Figure 1.2. Schematic representation of the five personality traits.

1.1.2 Agreeableness

The crucial functions underlying agreeableness are cooperation and altruism, which involve aligning our own goals with those of others (Graziano et al., 2007). This implies that agreeableness is associated with the capacity to comprehend others' perspectives and adapt the behavior accordingly (Nettle and Liddle, 2008). All the features encompassed by agreeableness seem to reflect a tendency towards altruism and cooperation rather than engaging in antisocial or exploitative behavior. Agreeable individuals demonstrate an interest in and consideration for the needs, desires, and feelings of others, avoiding aggression and the imposition of their own will. Such behavior necessitates the capacity to comprehend the emotions, intentions, and mental states of others (Allen and DeYoung, 2017; DeYoung 2010; Tackett et al., 2019). Accordingly, previous research has shown that agreeableness predicts performance on tests of empathy, ToM, and other social information processing tasks (Graziano et al., 2007; Nettle and Liddle, 2008; Wilkowski et al., 2006).

Agreeableness encompasses two aspects: *compassion*, reflecting emotional empathy and sympathy, and *politeness*, which involves conforming to social norms and refraining from aggression or exploitation (DeYoung et al., 2007, 2013). Consistent with these associations, agreeableness has been found to negatively correlate with externalizing type disorders and psychopathic traits (Decuyper et al., 2009; Miller and Lynman, 2001), and positively with the quality of social relationship with peers, family, and romantic partners, prosocial behavior, and reduced prejudice (Caprara et al., 2010; Habashi et al., 2016; Malouff et al., 2010; Sibley and Duckitt, 2008; Tackett et al., 2019). Finally, there are some evidences that agreeableness is associated with indicators of academic competence (McAbee and Oswald, 2013; O'Connor and Paunonen, 2007; Vedel, 2014) although findings are mixed and more research is needed (Tackett et al., 2019).

Studies about endogenous substances, such as hormones, suggest that serotonin might play a role in facilitating agreeableness and regulating aggression (Montoya et al., 2012; Pardini et al.,

2014); furthermore they report a negative correlation between testosterone levels and agreeableness, and conversely, a positive correlation with oxytocin, involved in social bonding and attachment (DeYoung et al., 2013; Perry et al., 2015; Turan et al., 2014).

Regarding sex differences, while it has been observed that agreeableness tends to slightly favor girls (Furnham and Cheng, 2015; Weisberg et al., 2011), these differences remain relatively small yet consistent from childhood through adulthood (Soto and Tackett, 2015). Biological theories, such as evolutionary psychology, posit that sex differences within personality traits arise from inherent temperamental variances between males and females, which have evolved through natural selection depending on the distinct adaptive challenges that the sexes faced throughout evolutionary history. For instance, females' greater biological investment in child-rearing, influenced by factors such as pregnancy, childbirth, and lactation, may contribute to their prioritization of relationships with offspring compared to men. Therefore, traits such as agreeableness and empathy in women could have conferred survival advantages to their progeny, offering an evolutionary edge (Costa et al., 2001).

Although neuroscientific research on this specific personality trait is still lacking (DeYoung and Blain, 2020), several studies have investigated the relationships between individual differences in agreeableness, and structural and functional characteristics of the brain by means of different measurement techniques.

Structural MRI studies. Evidence from structural MRI studies is mixed, with some studies finding no robust associations between agreeableness and brain structural indices (Avinun et al., 2020; Bjørnebekk et al., 2013; Chen and Canli, 2022; Liu et al., 2013; Riccelli et al., 2017), and others leading to consistent associations (DeYoung et al., 2010; Hu et al., 2011; Kapogiannis et al., 2013) between agreeableness and the volume of brain regions associated with social information processing, including the superior temporal sulcus (STS), posterior cingulate cortex (PCC), and fusiform gyrus (FG; DeYoung et al., 2010).

In the attempt to unfold these inconsistencies, Owens et al. (2019) designed a study to

replicate and extend the results of Riccelli et al. (2017) in an independent sample of equal size (N = 507), drawn from the same population using the same methods and then aggregate the results of their original sample and the replication sample (N = 1104). This study represents the largest investigation of the structural morphometry of personality to date. In the replication sample, agreeableness was found to positively correlate with local gyrification index in the left dorsolateral PFC (DLPFC), left precentral/postcentral gyrus and left dorsomedial PFC (DMPFC), as well as with the right temporoparietal junction (TPJ), inferior parietal lobule and superior parietal lobule, postcentral gyrus and lateral occipital cortex. Additionally, agreeableness was found to negatively correlate with local gyrification index in the left lingual gyrus and right parahippocampal gyrus. Within the full sample, clusters of association with agreeableness were found for thickness in the left DLPFC (negative) and the right lingual gyrus (positive). For the correlation of agreeableness and volume, a cluster of positive association was found in the right DLPFC.

fMRI studies. While agreeableness has not been extensively investigated in neuroscience, several fMRI investigations have utilized self-report measures of empathy, which is a facet of agreeableness. These studies have established connections between individual differences in empathy and neural activity during the observation and imitation of others' actions (Gazzola et al., 2006; Kaplan and Iacoboni, 2006), as well as the perception of others' emotional expressions (Chakrabarti et al., 2006; Schulte-Rüther et al., 2007). Empathy positively correlates with activity in the MPFC, STS, and the so called ‘mirror neuron system’, encompassing regions in the frontal and parietal cortex that exhibit similar activation during the execution or observation of the same action. Another study (Tankersley et al., 2007) demonstrated a positive association between a self-report measure of altruism (another facet of agreeableness) and activity in the posterior STS when observing task performance, as opposed to performing the task oneself. Moreover, a targeted fMRI study focusing specifically on agreeableness identified an association with activity in the left DLPFC, a region implicated in emotion regulation (Haas et al., 2007b). Additionally, in line with agreeable individuals' tendency to suppress aggressive impulses and socially disruptive emotions

(Meier et al., 2006), another fMRI investigation revealed that agreeableness predicted increased activation in the right LPFC when processing fearful faces, potentially reflecting automatic engagement of emotion regulation in response to potential threat or conflict (Haas et al., 2007b). Furthermore, empirical evidences have shown that more agreeable people are faster in emotional attribution decision and report a greater right TPJ (key node within the "social brain"; Frith and Frith, 2010) activity compared with less agreeable ones (Haas et al., 2015). On the other hand, low agreeableness in association with borderline personality traits is marked by hypoactivity in the right TPJ during emotional perspective-taking task (Haas and Miller, 2015).

Resting-state studies have revealed that increased levels of agreeableness are linked to heightened functional connectivity between ACC and precuneus, as well as regions implicated in empathy and social information processing (Adelstein et al., 2011). Furthermore, a recent meta-analysis has demonstrated a shared disrupted cognitive control network among individuals with psychiatric diagnoses characterized by aggressiveness (Wong et al., 2019). These findings suggest that networks involved in cognitive control may have associations with agreeableness.

Furthermore, other resting-state fMRI studies have reported a positive relationship between agreeableness and the connectivity among major hubs of the default network (DMN; Adelstein et al., 2011; Sampaio et al., 2014). Taking a step further, two recent studies investigated the neural correlates of mentalizing ability and exploring their relationship with agreeableness (Allen et al., 2017; Udochi et al., 2022). Although both of them found the DMN to be affected by individual variations in mentalizing, Udochi et al. (2022), when employing a task-based approach, found that the brain activity of the dorsomedial arears within prefrontal regions predicted both social cognitive ability and agreeableness during social animation processing.

EEG studies. EEG researches on agreeableness are very scares. Studies aimed at decoding personality traits from EEG recordings elicited during emotion processing found different results. For instance, Zhao et al. (2017) observed that parietal theta frequency band, frontal beta frequency band, and frontal theta and alpha asymmetry were associated with agreeableness trait. Klados et

al. (2020) reported an association between agreeableness and brain activity in the frontal and the occipital lobes with a dominant theta frequency band instead. In a resting state task-independent study, agreeableness was reported to negatively correlate with a posterior alpha power, and positively correlate with a parietal beta power (Jach et al., 2020). In a study using event-related potentials (ERPs), agreeableness was found to modulate risky decision-making within groups. Specifically, more agreeable individuals showed increased risk-taking behavior and a reduced P300 during group decision-making compared to individual decision-making (Wang et al., 2019). This result was not observed for less agreeable participants.

Unfortunately, compared with other Big Five traits, the underlying mechanisms of agreeableness have received less attention, with limited research exploring its neurocognitive correlates. Furthermore, the majority of the studies carried out to date has been developed by adopting resting state and/or structural measurements. However, task-based approaches within both EEG and fMRI studies are needed to understand how agreeableness might affect specific cognitive processes and their neural correlates, aspects that the former approaches cannot capture. Since agreeableness has been associated with many desirable life outcomes, I argue that further investigation is warranted.

1.2 Theory of Mind

1.2.1 An umbrella term: from lower to higher processes

The ability to effectively navigate social relationships and interactions requires adaptability in response to changing environmental demands. This involves understanding both the behaviors of others and the underlying drives behind those behaviors, encompassing their affective and cognitive states. The conceptual framework that recognizes the influence of mental states on observable behavior is commonly referred to as ToM (Wellman, 1990), that has emerged as a central process in facilitating social interactions (Schurz et al., 2021). The term “theory” is employed due to the nature of our inability to directly observe the mental states of others. Instead,

we formulate hypotheses about their existence and make probabilistic assessments regarding their specific contents, drawing upon a variety of pertinent evidence (Sabbagh and Bowman, 2018).

ToM is an umbrella term that can be fractioned into multiple component processes, including mental state decoding and mental state integration, each relying on distinct types of information (Sabbagh, 2004). Typically, in the real life, mental state decoding skills work in tandem with mental state reasoning skills to provide reliable inferences about others' intentions and actions; this the reason why it is challenging to isolate the aspects of neural activity specifically associated with decoding others' mental states. This distinction parallels the distinction between social- perceptual (or social-affective) and social-cognitive components of ToM (Meinhardt-Injac et al., 2018). The former involves inferring an individual's emotional state based on external cues such as facial expressions and non-verbal behaviors, while the latter involves inferring mental states based on subtle contextual cues that may not be physically apparent. Although measures of both the two components show a moderate relationship (Fossati et al., 2017), there is evidence to suggest that they rely on distinct neural circuits. The social-perceptual component is argued to be associated with circuits involving the OFC, medial temporal regions, and amygdala (Baron-Cohen et al, 1999; Schurz et al., 2021). In adults, this component has typically been measured using the "Reading the Mind in the Eyes test" (RMET; Baron-Cohen et al., 2001), in which a participant infers an actor's mental state from their eye gaze. In contrast, the social-cognitive component is mediated by the MPFC (specifically dorsally and in the left hemisphere), the TPJ, and the precuneus (Allen et al., 2017; Haas et al., 2015; Sabbagh, 2004; Saxe and Kanwisher, 2003; Spunt and Lieberman, 2012; Young et al., 2010). Importantly, the latter circuit overlaps significantly with the DMN, which is activated when subjects passively generate mental content associated to memories, imagination, and social cognition (Allen et al., 2017; Spreng and Andrews-Hanna, 2015). The social-cognitive component is typically measured by asking the participant to identify an actor's belief or intention after listening to a complex social story or scenario.

Impaired ToM has been linked with several disorders such as schizotypy (Blain et al., 2017;

Bora, 2020), schizophrenia (Abram et al., 2016; Pedersen et al., 2012), autism (Baron-Cohen et al., 1985; Brüne and Brüne -Cohrs, 2006), Williams syndrome (Tager-Flusberg and Sullivan, 2000), and also with autistic traits (Best et al., 2008; Blain et al., 2017). Noteworthy, ToM impairments in such populations are often associated with only one of the two components depending on the pathology considered.

The developmental trajectory of ToM understanding originates from early social perception. As early as three months of age, infants are capable of tracking another person's eye gaze (D'Entremont et al., 1997; Moore, 1999) and engaging in joint visual attention in social contexts (Corkum and Moore, 1998; Deák et al., 2000). These abilities demonstrate sensitivity to others' attention and the perceptual information available to them, which may play a crucial role in establishing an understanding of the relationship between the observable world and internal mental states. Developmentally, major changes in ToM reasoning occur during the preschool years (between 3 and 5 years of age; Wellman et al., 2001), and this developmental timeline remains consistent across cultures with diverse worldviews and languages (Sabbagh et al., 2006). These findings support the hypothesis that ToM is a fundamental human conceptual understanding shared worldwide.

1.2.2 Mental state decoding

Baron-Cohen et al. (2001) defined the mental state decoding as the first stage of attribution of ToM. The process of attributing the relevant mental state, such as compassion, represents an important aspect ToM albeit not encompassing its entirety. Noteworthy, this process specifically refers to the identification of the type of mental state starting from visible cues and does not encompass the subsequent stage of inferring the content associated with that mental state, such as compassion for a specific event like the loss of one's dear friend. However, even within this category, brain processes of varying complexity can be encompassed depending on the task and the cues employed. For instance, in addition to the previously mentioned and well-established

RMET, other tasks can be employed such as identifying others' focus of attention from their gaze or the recognition of basic emotions from point-light displays (PLDs). PLDs have been demonstrated to carry information about motion allowing a quick perception and categorization of emotions such as happiness, sadness, and anger (Atkinson et al., 2004; Dittrich et al., 1996). The decoding of actions and emotions from PLDs has been associated with reduced alpha-band (8–12 Hz) EEG oscillations over the sensorimotor cortex (Siqi-liu et al., 2018). This so called "μ-alpha" suppression is believed to be indicative of mental simulation of action (Niedermeyer and da Silva, 2005; Ulloa and Pineda, 2007), which in turn has been argued to improve internal representation of others' emotional states (Kilner et al., 2006; Pineda and Hecht, 2009), therefore, supporting mental state decoding (Sabbagh and Bowman, 2018).

Previous evidence suggests the involvement of the right hemisphere in mental state decoding, as indicated by studies on ERPs and neuroimaging in tasks requiring mental state inference (Baron-Cohen et al., 1999; Sabbagh et al., 2004). Furthermore, individuals with right-hemisphere damage exhibit difficulties similar to those observed in individuals with ASD in tasks requiring ToM reasoning (Brownell et al., 2000; Sabbagh, 1999). Additionally, specific ERP components that are suggested to be modulated during mental state decoding tasks have been identified. For example, the N2 component over frontal regions, the P3 over parietal regions (~300-600 ms), and the late slow potential over frontal regions (positive or negative, ~500-1000 ms) (Baron-Cohen et al., 2013). Sabbagh et al. (2004) also found the involvement of the frontal N270-400 during the RMET located within the orbitofrontal and medial temporal cortex according to the source estimation procedure. Noteworthy, the N270 is involved in processing emotions (Sato et al., 2001). Using the same task, mental state decoding skills turned out to be stronger among participants with right-lateralized frontal activation (Sabbagh and Flynn, 2006). Unfortunately few EEG studies have specifically targeted the decoding of mental states. Accordingly, further studies are needed to confirm these results and obtain a reliable and complete picture of the neural dynamics of this process.

1.2.3 Sex differences

There are consistent empirical evidences showing that females exhibit greater empathy and accuracy in interpreting others' emotional states (Baron-Cohen et al., 2001; Stiller and Dunbar, 2007; Udochi et al., 2022). However these differences decrease in social-perceptual tasks (Kirkland et al., 2013) and are not always evident, especially at the behavioral level in which results are aften contradictory. Accordingly, one possible explanation that has been suggested rely on sex-related brain dimorphism that might prevent behavioral differences if they are not adaptive, and foster proper behavioral response (Pavlova et al., 2010, 2015). However, being ToM an umbrella term, as previously discussed, we think is always warranted to take into account possible sex differences depending on the specific process analysed since differences among females and males in processing social and emotional stimuli have been always stressed by the literature. Specifically, hemispheric lateralization for face processing, facial expression encoding, interest in social scenes and human faces, understanding body language, intentions, and gestures are some of the main established discoveries in the research about sex differences (for a review, Proverbio 2021). These differences are explained as due to both genetic and hormonal influences that affect anatomical and functional brain differences between the two sexes, together with the evolutionary pressure primarily concerning the mother-infant bonding (see Cahill, 2006) as mentioned above.

1.3 Autism Spectrum Disorder (ASD)

1.3.1 Big Five personality traits in ASD

The role of personality and individual differences in autism is an emerging research topic that has received limited attention so far, but there is growing interest for this research area. In a recent meta-analysis, Lodi-Smith and colleagues (2019) conducted a comprehensive examination of personality research in ASD, acknowledging the substantial heterogeneity found in the literature. The results revealed that individuals with ASD tend to exhibit lower levels of openness, conscientiousness, extraversion, agreeableness, and emotional stability. However, it is important

to note that the correlations observed are of small to moderate magnitude, with stronger associations observed specifically for extraversion and neuroticism. Contrary to assumptions, results by Schwartzman et al. (2016) suggest that not all individuals with ASD display uniformly low Big Five trait profiles. These findings indicate that low Big Five traits and ASD are not equivalent. Therefore, personality trait measurement should be viewed as a potential complement rather than a substitute for traditional diagnostic tools, emphasizing the need for continued evaluation of personality and ASD within dimensional models.

The inclusion of personality trait assessments as part of a comprehensive clinical assessment might offer valuable insights into the life outcomes of individuals with ASD. Noteworthy, traits were found to be equally or even more variable in individuals with ASD compared to neurotypical samples (De Pauw et al., 2011; Hesselmark et al., 2015; Rodgers et al., 2018; Schriber et al., 2014). As for individuals without diagnosis or individuals with other diagnoses, this variability in personality traits might contribute to our understanding of the achievements and difficulties experienced by individuals with ASD. For instance, in children with ASD, emotional stability, extraversion, and agreeableness have been related to internalizing and externalizing problems (De Pauw et al., 2011; Schriber et al., 2014). Similarly, maladaptive personality characteristics among adults with ASD appear to negatively affect the quality of life (Ratner and Burrow, 2017; Rodgers et al., 2018; Schwartzman et al., 2016). The evidence from this meta-analysis suggests that the overall trait profile in ASD poses challenges for individuals throughout their lives. However, the variability in personality within individuals with ASD also offers the potential for more positive trajectories. Indeed, children with higher levels of all Big Five traits were less likely to meet diagnostic criteria for ASD after previously receiving an ASD diagnosis (Suh et al., 2016), while adults with this trait profile reported higher levels of education, greater job and overall life satisfaction (Schwartzman et al., 2016). It is possible that individuals with ASD exhibiting more typical personality traits are better equipped to adapt to societal demands. Moreover, specific aspects of the personality profile of individuals with ASD might

serve as strengths in certain contexts, particularly within academia and other workplaces (APA, 2013; Austin, 2005; Baron-Cohen et al., 2001; Strunz et al., 2015).

Noteworthy, the results of the considered studies rely only on self-report or other-report, therefore, future studies are needed that consider the association between personality traits, and more objective measures such as behavioral performance tested by means of computerized tasks, in association with neuroimaging and electrophysiological measurements. Furthermore, conflating results from samples that differ in age (i.e., children and adults) and diagnosis (i.e., high and low functioning) might complicate the interpretation of the results.

To conclude this section, the literature on personality and ASD is still emerging with relatively few studies. Personality has wide-ranging implications in both clinical and adaptive domains across an individual's lifespan. Investigating personality within the context of disorders can enhance our understanding of factors that pose risks, physiological mechanisms, phenomenology, and intervention response. Furthermore, personality shows potential as a viable focus for intervention strategies.

1.3.2 The "Extreme Male Brain" theory

The Extreme Male Brain (EMB) theory developed by Baron-Cohen (2011) represents an extension of the Empathizing-Systemizing theory (E-S; Baron-Cohen, 2009; Baron-Cohen et al., 2005) which posits that individuals can be categorized based on two dimensions: empathy, involving the ability to recognize and respond to another person's mental state, and systemizing, involving the inclination to analyze and construct rule-based systems. These dimensions follow a normal distribution in the general population, with biological factors, such as prenatal testosterone, contributing to variance. The E-S theory makes several predictions regarding sex differences: females scoring higher on empathy, males scoring higher on systemizing, a small inverse correlation between empathy and systemizing, more females having a brain type characterized by higher empathy, and more males having a brain type characterized by higher systemizing (Baron-

Cohen et al., 2014; Wheelwright et al., 2006). The theory suggests that evolutionary selection has fostered the development of specialized brains, exhibiting a greater inclination towards one domain over another depending on sex, likely due to the adaptive nature of empathy and systemizing in distinct environments, such as social and technical contexts (Greenberg et al., 2018). Building on these assumptions, the EMB proposes that individuals with autism exhibit a more "masculine" brain type, with difficulties in empathy and at least average aptitude in systemizing, which may explain the higher prevalence of autism in males. The EMB theory predicts that more autistic individuals have a brain type characterized by extreme systemizing; autistic traits are better predicted by brain type than sex; males have a higher number of autistic traits on average; individuals in STEM fields exhibit a higher number of autistic traits compared to non-STEM occupations.

These predictions were tested in a study by Greenberg et al. (2018) involving a very large sample of participants (first study, N = 671,606, including 36,648 individuals with ASD, and validation N = 14,354, including 226 individuals with ASD). At the origin of these brain specializations, and sex differences in empathy and systemizing scores, an interplay between biological and cultural factors has been proposed taking into account the possibility of exposure to fetal testosterone (Auyeung et al., 2006; Chapman et al., 2006) and genetic common variance (Strang et al., 2014; Warrier et al., 2018, 2019) together with the significance of social experiences. The study provides empirical evidences that both males and females with ASD exhibit a shift towards a more masculinized brain type, with a higher likelihood of having a systemizing brain. Brain imaging studies align with the EMB theory, revealing masculinized brain structure (Beacher et al., 2012; Lai et al., 2013) and function in autistic females (Beacher et al., 2012; Floris et al., 2018; Ypma et al., 2016). The EMB theory has also prompted investigations into sex-linked prenatal etiological factors, including confirmed elevated prenatal sex steroids (Baron-Cohen et al., 2015), circulating sex steroids (Schwarz et al., 2011), and increased rates of steroidopathy in ASD females, including polycystic ovary syndrome (Cherskov et al., 2018; Pohl et al., 2014). To

avoid misunderstandings, it is crucial to emphasize that the EMB theory does not stereotype ASD individuals as having extreme male characteristics, such as aggression. Instead, it predicts a masculinized pattern of empathy (below average) and systemizing (average or above average) which has been supported by Greenberg et al. (2018) study. Furthermore, the theory does not imply a lack of empathy in autism; rather, lower scores on empathy measures primarily reflect difficulties with cognitive empathy or ToM, while affective empathy seems to remain intact (Dziobek et al., 2008; Rueda et al., 2015). Finally, the E-S theory does not support "neurosexism" (Fine, 2011; Joel, 2015), as it does not necessarily predict that an individual's psychological profile is based on their biological sex. The scientific evidence coming from sex differences studies only permits inferences about average differences between males and females as groups (Greenberg et al., 2018).

1.3.3 The social motivational theory

The social motivational theory of autism, proposed by Chevallier and colleagues (2012), offers a comprehensive framework for understanding the core features of ASD based on motivational and reward processing differences. According to this theory, individuals with ASD exhibit unique patterns of motivation that contribute to their characteristic behaviors and difficulties in social interaction. The theory suggests that individuals with ASD have reduced social motivation, leading to a decreased interest and engagement in social interactions compared to typically developing individuals. This reduced social motivation stems from a decreased reward value associated with social stimuli and interactions. Consequently, individuals with ASD may find social interactions less inherently rewarding and, as a result, may be less motivated to seek social engagement (Chevallier 2012; 2013).

Accordingly, fMRI studies have revealed atypical activation in brain regions associated with reward processing, such as the orbitofrontal-striatum-amygdala circuit, in response to social stimuli in individuals with ASD (Bachevalier and Loveland, 2006; Masten et al., 2011; Schultz et

al., 2000; Scott Van Zeeland et al., 2010). In addition, current research on neuropeptide signaling and ASD indicates that impaired regulation of oxytocin may be significant in the social reward dysfunctions observed (Modi et al., 2011) inhering the accurate association between social stimuli and motivational values (Bartz et al., 2011). Furthermore, in the visual domain, eye-tracking studies have shown that individuals with autism attend more to non-social features than to social ones of naturalistic visual scenes. When viewing static social photographs, children with autism tend to look significantly more at the background rather than the faces or bodies of the characters (Riby and Hancock, 2008). Furthermore, adolescents and young adults with autism, when freely viewing movie clips, have been found to fixate less on people, faces, and eyes compared to other regions of interest (Klin et al., 2002; Nakano et al., 2010). These findings suggest that social information is not given the same priority in autism as it is in typical development, indicating an overall decrease in the attentional weight assigned to the social world. Accordingly fMRI studies have reported altered activation patterns in brain regions involved in attention, such as the prefrontal and parietal cortex, in individuals with ASD (Chevallier et al., 2012).

Within the auditory domain, children with autism do not exhibit a preference for salient social sounds, such as the human voice, over non-social control noise (Klin, 1991; Kuhl et al., 2005). Similarly, deficits in attention have been observed in relation to speech sounds, while attention to non-speech information appears to be increased in individuals with autism (Ceponiene et al., 2003; Gervais et al., 2004).

Altered brain processing of social and emotional stimuli probably relied on attentional and motivational impairments have been reported in people with ASD by ERPs studies showing reduced P1 and N170 components and differences in LPP (Fan et al., 2014; Hileman et al., 2011; Kang et al., 2018; Lerner et al., 2013). Furthermore, several studies found that a reduced ERP amplitude associated with the processing of social stimuli went with a higher brain response for non-social stimuli compared with typically developing individuals (Benning et al., 2016; Stavropoulos and Carver, 2014).

Therefore, according to the social motivation theory, social cognition deficits in ASD, like those observed in ToM and empathy tasks, are a consequence of a reduced sensitivity to social stimuli associated with a diminished social motivation and attention (Chevallier et al., 2012; Dawson et al., 1998; Schultz, 2005).

1.4 The aims of the thesis and outline

The primary aim of my thesis is to investigate to which extent agreeableness modulates social and emotional processes associated with ToM. I expect to find differences both at the behavioral and neural level depending on the way in which social and emotional stimuli are processed in relation to participants' agreeableness level. As a secondary aim, I will compare social and emotional processes in the general population of young and neurologically intact adults with those of individuals belonging to ASD without language and intellectual impairments.

Compared to the other personality traits, agreeableness has received less attention in previous research. Furthermore, existing studies exploring its neurobiological underpinnings have reported divergent findings and, in some cases, null results. These conflictual findings may depend on the multifaceted nature of behaviors associated with agreeableness and its complex connections with several cognitive processes. Hence, I believe that a more rigorous investigation into this personality trait and its association with distinct neural processes can yield valuable insights. Moreover, agreeableness has a direct association with a multitude of positive life outcomes, encompassing domains like interpersonal relationships, individual contentment, overall life satisfaction, and even mental health (Allen and DeYoung, 2017; DeYoung and Blain, 2020; Tackett et al., 2019). The thesis is organized in four studies. In Study 1 we examined whether individual differences in the neural representations of social information were related to differences in agreeableness (Arbula, Pisanu, Rumiati, 2021). Exploiting the positive fMRI finding, we designed two EEG studies to investigate the temporal dynamics of different neural mechanisms underlying the interaction between agreeableness and ToM. Given the complexity of

ToM, these studies were designed to analyse separately different social and emotional processes involved in perceiving and interpreting social and emotional cues. In the Study 2, using EEG we assessed the impact of agreeableness on social interaction perception while taking into account sex differences that might moderate this relation (Pisanu, Arbula, Rumiati, *under review*). In Study 3, using EEG, we investigated two emotional processes with a different degree of complexity by means of two different tasks. The two tasks are the following: 1) The Biological (e)motion task requires emotional perception processes to discriminate emotional actions represented in the movements of point-light body stimuli. 2) The Reading the Mind in the Eyes (RME) task requires higher processes of mental state recognition to discriminate mental states only through the expression of the eyes (Pisanu, Arbula, Rumiati, *under review*). In Study 4, we investigated the personality profile of young adults with ASD without language and intellectual impairment, and, by using the EEG, whether agreeableness might differentially modulate social skills in this population and account for their social difficulties.

CHAPTER 2

Study 1: Representation of social content in dorsomedial prefrontal cortex underlies individual differences in agreeableness trait

---

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

Abstract

Personality traits reflect key aspects of individual variability in different psychological domains. Understanding the mechanisms that give rise to these differences requires an exhaustive investigation of the behaviors associated with such traits, and their underlying neural sources. Here we investigated the mechanisms underlying agreeableness, one of the five major dimensions of personality, which has been linked mainly to socio-cognitive functions. In particular, we examined whether individual differences in the neural representations of social information are related to differences in agreeableness of individuals. To this end, we adopted a multivariate representational similarity approach that captured within single individuals the activation pattern similarity of social and non- social content, and tested its relation to the agreeableness trait in a hypothesis-driven manner. The main result confirmed our prediction: processing social and non-social content led to similar patterns of activation in individuals with low agreeableness, while in more agreeable individuals these patterns were more dissimilar. Critically, this association between agreeableness and encoding similarity of social and random content was significant only in the dorsomedial prefrontal cortex, a brain region consistently involved during attributions of mental states. The present finding reveals the link between neural mechanisms underlying social information processing and agreeableness, a personality trait highly related to socio-cognitive abilities, thereby providing a step forward in characterizing its neural determinants. Furthermore, it emphasizes the advantage of multivariate pattern analysis approaches in capturing and understanding the neural sources of individual variations.

2.1 Introduction

Every human being is unique, and one important part of this uniqueness is determined by personality. Understanding the neural sources of inter-individual variability in personality is the major goal of personality neuroscience (Yarkoni, 2015). Differences across personality traits are reflected in different motivational, emotional, cognitive and behavioral responses to particular

stimuli. According to many theories, these responses are relatively stable, but manifest only in certain contexts (Corr et al., 2013; DeYoung, 2010; Gray, 1982; Tellegen and Waller, 1981). Yet, numerous studies identifying the neural correlates of different personality traits have focused on spontaneous, resting state brain activity (e.g., Cai et al., 2020; Dubois et al., 2018; Kuper et al., 2019; Mulders et al., 2018; Nostro et al., 2018), or structural brain features (e.g., Avinun et al., 2020; Lewis et al., 2018; Omura et al., 2005; Owens et al., 2019; Riccelli et al., 2017; Taki et al., 2013), in which indices of personality traits were rarely clearly evident, and often diverged between studies. Another important limitation is that these types of associations are relatively uninformative about the mechanisms that contribute to distinct trait-specific behaviors.

An alternative approach holds that differences in personality affect specific cognitive mechanisms and, in turn, modulate the neural correlates observed during task-based neuroimaging. Such an association has been hypothesized between the trait agreeableness - one of the five broad dimensions of personality within the five-factor model (Costa and McCrae, 1992b; John et al., 2008) - and mentalization processes (Allen et al., 2017; Nettle and Liddle, 2008). Agreeable individuals have more empathic, altruistic and cooperative tendencies, which require the ability to understand others' mental states, emotions and intentions (Allen and DeYoung, 2017). This hypothesis has been tested both indirectly, by associating agreeableness with prosocial behavior and empathy (Graziano et al., 2007; Habashi et al., 2016; Penner et al., 1995), and directly, by showing its impact on the performance in a mentalizing task (Nettle and Liddle, 2008). More recently, two studies have taken a step further by investigating the neural correlates of ability and testing whether those correlates are related to agreeableness (Allen et al., 2017; Udochi et al., 2020). Both studies considered the default network as the neural substrate of mentalization, and found it to be modulated by individual variations in mentalizing ability, but only Udochi et al. (2020) found that it also predicted individual differences in agreeableness. One key distinction between the two studies is that Allen et al. (2017) defined the default network from resting state functional connectivity, while Udochi et al. (2020) adopted a task-based approach. In particular, in

the latter work the authors found that the default network activity during social animation processing predicted both social cognitive ability and agreeableness. While this finding provides new insights on the common neural substrates of agreeableness and mentalization, it still remains unclear how these neural variations underlie behavioral differences. Do they reflect differences in processing social information among individuals with different degrees of agreeableness?

To test this hypothesis we used representational similarity analysis (RSA; Kriegeskorte et al., 2008) that allowed us to examine whether individual differences in the representation of social content are related to differences in agreeableness of individuals. In particular, we compared the patterns of neural activity during social and random (i.e., non-social) animations viewing within each participant in order to measure on an individual basis how similarly the two types of contents are encoded. We predicted that the levels of agreeableness would correlate positively with the dissimilarity in neural representations of social and random content, and that this association will be mainly evidenced in brain regions involved in mentalization processes (i.e., mentalizing network) (Molenberghs et al., 2016; Schurz et al., 2014). Moreover, we hypothesized a possible correlation also with task accuracy, showing greater social-random pattern dissimilarity for higher accuracies. On the other hand, we did not expect any other trait to be related to this pattern dissimilarity.

With respect to univariate approaches that relate individual differences to differences in activation of single voxels, or activation averaged across voxels in restricted brain regions, the representational similarity approach is more sensitive to individual variations since it takes into account distributed patterns of activity, which hold much more information that can be compared within and between individuals (Etzel et al., 2020). On the other hand, it suffers certain limitations in inferring what type of information is encoded in the representational space (Naselaris and Kay, 2015). One way to overcome this issue is to predict neural dissimilarities between experimental conditions according to a psychological model that makes clear inferences about the observed behavior (Carlson et al., 2018; Ritchie et al., 2019). Here we adopted this strategy to investigate

the neural sources of agreeableness trait, by basing our predictions of social content encoding on well-established behavioral facets of agreeableness and its link with mentalization ability.

2.2 Methods

2.2.1 Participants recruitment and personality assessment

Participants' recruitment involved a prescreening personality assessment that was administered through an online questionnaire, advertised through different recruitment services (Facebook, Sona). Five personality traits (extraversion, agreeableness, conscientiousness, neuroticism and openness to experience) were assessed using the Italian adaptation of the Big Five Inventory (Ubbiali et al., 2013), which comprises 44 items rated on a five-point Likert scale. Fifty-five participants were selected from a larger sample (N = 143) based on their personality scores in order to cover uniformly all five traits derived from the five-factor model (Costa and McCrae, 1992b; John et al., 2008). In particular, within each trait the score distribution was normal (Kolmogorov-Smirnov ps' > 0.15) and at least 5 subjects had scores above ±1 SD from the Italian population mean (Ubbiali et al., 2013). Specifically as regards agreeableness, 11 subjects scored above and 11 subjects below 1 SD. As personality traits are continuous variables, in order to ensure that our sample accurately captures measurements at either tail of the distribution, deliberate efforts were made to select individuals with diverse scores along the z-score continuum, encompassing both negative and positive values. Six participants were excluded due to excessive motion during scanning, yielding a final sample of 49 participants (19 males) with mean age 23.14 (SD = 4.24, range = 18– 34). None of them reported any history of neurologic or psychiatric disorders. All participants gave written informed consent prior to their participation in the study and received 30 euros for their participation. The study was approved by the Regional Ethics Committee of Friuli Venezia Giulia and was conducted according to the guidelines of the Declaration of Helsinki. Behavioral and demographic information are reported in the participants.tsv file on OpenNeuro.

2.2.2 Social cognition task

Participants were presented with short animations representing different shapes that moved randomly or interacted in a socially meaningful way. The animations were originally developed by Castelli et al. (2000) and Wheatley et al. (2007) and shortened to 20 s for the battery of fMRI tasks used in the Human Connectome Project (Barch et al., 2013). During each of the 4 runs 5 different animations were presented interleaved with 15 s long fixation blocks. After each animation, a 3 s long instruction screen appeared prompting the participants to respond by pressing one of the three keys on a response pad positioned under their right hand. They were told to press their ring finger if the shapes were moving randomly, their index finger if the shapes interacted in a socially meaningful way (as if they were taking into consideration each other feelings and thoughts), and the middle finger if they were not sure about the type of interaction. Each run consisted of 2 or 3 videos of each condition. One social and one random video were presented prior to scanning for practice and were not reused during the testing phase.

2.2.3 MRI acquisition

MRI data were collected on a 3 Tesla whole-body scanner (Achieva Philips) equipped with an 8-channel head coil at the "S. Maria della Misericordia Hospital" in Udine. For each of the four runs of the social cognition task, 101 functional image volumes with 37 contiguous axial slices were collected with a T2*-weighted echo-planar sequence (TR: 2 s, TE: 30 ms, FA: 82°, voxel size: 3 × 3 × 3 mm, acquisition matrix: 80 × 80). A high-resolution T1-weighted anatomical image was acquired at the beginning of the session (170 sagittal slices, TR/TE: 8.1/3.7 ms, FA: 12°; voxel size: 1 × 1 × 1 mm, acquisition matrix: 240 × 240). Additionally, to correct for spatial distortion of functional images, a pair of spin echo images with opposite phase encoding directions and same orientation as the functional scans, were acquired at the beginning and in the middle of the whole scanning session. Stimuli were presented using E-Prime 2 (Psychology Software Tools, Inc; Schneider et al., 2012) and delivered through MRI-compatible goggles mounted on the head coil.

During the scanning session, participants performed three other tasks in a counterbalanced order, which are not reported in the present study.

2.2.4 MRI quality assessment

All MRI data were converted from DICOM format into the Brain Imaging Data Structure (BIDS; https://bids.neuroimaging.io/) using the Dcm2Bids tool (https://github.com/cbedetti/Dcm2Bids). Quality of structural and functional data was assessed using the MRI Quality Control tool (MRIQC) (Esteban et al., 2017) and compared to a set of quality metrics from the MRIQC web API (Esteban et al., 2019) using the MRIQ-Ception tool (https://github.com/elizabethbeard/mriqception). Out of 55 scanned participants, data from 6 were classified as outliers based on the following quality metrics: AFNI's outlier ratio and quality index, intensity changes (DVARS) and frame-wise displacement (FD). A full report of image quality metrics is available on OpenNeuro.

2.2.5 MRI preprocessing

Results included in this manuscript come from preprocessing performed using fMRIPrep version 1.5.1rc2 (Esteban et al., 2019; RRID:SCR_016216), a Nipype (Gorgolewski et al., 2011, 2017; RRID:SCR_002502) based tool. Each T1w (T1-weighted) volume was corrected for INU (intensity non-uniformity) using N4BiasFieldCorrection v2.2.0 (Avants et al., 2008; RRID:SCR_004757) and skull-stripped using antsBrainExtraction.sh v2.2.0 (using the OASIS template). Brain surfaces were reconstructed using recon-all from FreeSurfer v6.0.1 (Dale et al., 1999; RRID:SCR_001847), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017 ; RRID:SCR_002438). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009; RRID:SCR_008796) was performed through nonlinear registration with the antsRegistration tool of ANTs v2.2.0 (Avants et al., 2008; RRID:SCR_004757), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-

matter (GM) was performed on the brain-extracted T1w using fast (FSL v5.0.9; Zhang et al., 2001; RRID:SCR_002823). Functional data was slice time corrected using 3dTshift from AFNIv16.2.07 (Cox, 1996; RRID:SCR_005927) and motion corrected using mcflirt (FSL v5.0.9; Jenkinson et al., 2002). Distortion correction was performed using an implementation of the TOPUP technique (Andersson et al., 2003) using 3dQwarp (AFNI v16.2.07; Cox, 1996). This was followed by co-registration to the corresponding T1w using boundary-based registration (Greve and Fischl, 2009) with six degrees of freedom, using bbregister (FreeSurfer v6.0.1). Motion correcting trans-formations, field distortion correcting warp, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using antsApplyTransforms (ANTs v2.2.0) using Lanczos interpolation. Physiological noise regressors were extracted applying CompCor (Behzadi et al., 2007). Principal components were estimated for the anatomical CompCor variants (aCompCor). A mask to exclude signal with cortical origin was obtained by eroding the brain mask, ensuring it only contained subcortical structures. Six aCompCor components were then calculated within the intersection of the subcortical mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run. Frame-wise displacement (FD) and DVARS (Power et al., 2014) are calculated for each functional run using their implementations in Nipype. Functional data were masked using the brain mask obtained from fMRIPrep and 14 fMRIPrep derived confounds (six motion parameters, FD, standardized DVARS and six aCompCor) were removed on a voxel- wise level using the Denoiser tool (https://github.com/arielletambini/denoiser). As a final step, functional data were spatially smoothed using a Gaussian kernel of 6 mm full-width at half-maximum.

2.2.6 Behavioral data analyses

We collected accuracy and reaction times (RTs) data. RTs were filtered for errors and outliers above 3 standard deviations from the subjects mean for each video type condition. One subject was excluded from the behavioral analysis because we did not collect all of his responses

due to a response pad issue. Anticipated responses (RTs < 200 ms) were absent. RTs were assessed with repeated measures ANCOVA with type of video as within-subjects factor and five trait scores as continuous predictors. Since accuracy data were not normally distributed (Shapiro-Wilk test $p < .05$), a paired Wilcoxon signed rank test was used to assess accuracy differences between the two types of videos, and Spearman rank correlation was used to assess the accuracy correlations with the five trait measures, separately for each type of video.

2.2.7 fMRI data analyses

First level GLM analysis

First level GLM analysis was performed using FSL FEAT (www.fmrib.ox.ac.uk/fsl). A GLM model was built for each participant and each run with the two experimental conditions (social and random) as regressors of interest and their temporal derivatives as regressors of no interest. Erroneous trials and those with "not sure" responses were not excluded from the model to preserve higher inter-individual variability, as we predicted that accuracy would vary based on agreeableness scores and our main interest was the participants' stimulus processing. The regressors were time locked to the onset and duration of the video and convolved with a double-gamma hemodynamic response function. FILM pre-whitening was used to correct for autocorrelation and low-frequency drifts were removed using a high-pass filter with a 100 s cutoff. All following analyses were performed on subject-level social and random beta maps, averaged over the four runs (fixed effects).

Group level GLM analysis

To localize mean group effects, whole-brain group level analysis was performed with mixed effect (FLAME 1), as implemented in FSL, for the social - random contrast. Statistical map was assessed with a cluster- based threshold of $Z > 3.1$ and corrected at $p = 0.05$ (family-wise error correction).

Representational similarity analysis

We conducted a within subject representational similarity analysis (RSA) to test if the participants' level of agreeableness will be related to the dissimilarity between the neural patterns of activity during social and random content processing. As anticipated in the introduction, we expected to observe a positive correlation between the two measures in brain regions belonging to the mentalizing network (Molenberghs et al., 2016 ; Schurz et al., 2014), indicating that subjects low in agreeableness will have less distinct neural representations of social and random content, and vice versa. However, since agreeableness has been associated also with brain regions outside the mentalizing network (e.g., Cai et al., 2020 ; Liu et al., 2019 ; Riccelli et al., 2017), we opted for a whole-brain parcellation scheme in our analyses, since it offers a nice middle ground between the restrictive ROI- based approach and the heavily corrected and computationally expensive searchlight approach. As a first step, all beta maps were divided into 200 non-overlapping regions using a whole-brain parcellation derived from the meta-analytic functional coactivation of the Neurosynth database (https://identifiers.org/neurovault.image:39711). Next, representational dissimilarities were computed for each participant by correlating social and random response patterns from each parcel and subtracting it from 1 (i.e., higher correlation is reflected as lower dissimilarity). The participants' dissimilarities within each parcel were then correlated with their agreeableness scores using Spearman's rank-order correlations. The statistical significance (p-value) of the resulting correlations was obtained by computing the same correlations after permuting 30,000 times the agreeableness scores and calculating the proportion of permuted correlations that exceeded the ones yielded from non-permuted data. The same representational dissimilarity and behavioral score correlation was repeated for the remaining four trait dimensions. All resulting permuted p-values were corrected for multiple comparisons by dividing the alpha (0.05) by the number of behavioral measures tested (five traits and one accuracy score) and the number of parcels (200).

Univariate analysis

Alternatively, we wanted to assess whether the hypothesized neural dissimilarity associated with agreeableness could be explained by univariate activation differences between social and non-social content. Therefore we calculated the difference between the average levels of activity (averaged betas across all voxels within each parcel) for each condition (social –random contrast), and correlated it with the agreeableness score. All statistical procedures and subsequent multiple comparison corrections were same as in the RSA.

Meta-analytic decoding

The correlational maps from the RSA and the Z maps from the GLM analysis were correlated with the meta-analytic activation maps of the terms in the Neurosynth database in order to determine what cognitive terms are most commonly associated with these brain regions in the literature. Since the Neurosynth decoder compares whole-brain images, both comparisons were performed on unthresholded data that contain additional information in the form of continuous values at all voxels (Gorgolewski et al., 2015). For the RSA maps, we retained the top 10 terms with the highest positive correlations referring to cognitive functions (i.e., anatomical terms were excluded), and afterwards we compared their correlations with the correlations observed for the GLM maps, in order to explore what cognitive terms associated with the RSA maps have similar associations with the social-random contrast.

2.3 Results

2.3.1 Behavioral results

The accuracy differed significantly between the two types of video content (social accuracy mean: 97.6%, SD: 15.5; random accuracy mean: 88.4%, SD: 32; Wilcoxon signed rank test V = 353.5, p = .0006, r = 0.5), but there were no significant correlations with trait measures (all p's > 0.25; see Supplementary Table 2.1 for all results). For the RTs, there was no main effect of video type (social RTs mean: 839 ms, SD: 374; random RTs mean: 837 ms, SD: 477; $F(1, 42) = 3.64$, p

= .06, η_p^2= 0.08) and no main effects of or interactions with trait measures (all *p*'s > 0.09; see Supplementary Table 2.1 for all results).

2.3.2 Neuroimaging results

Whole-brain group results

The social-random content contrast at the group level showed widespread bilateral activations in regions typically associated with social content processing, comprising the fusiform gyrus, inferior and middle temporal regions, inferior frontal gyrus and superior medial frontal cortex (Figure 2.1; Table 2.1). The results from the random-social contrast are included in the Supplementary material (Supplementary figure 2.1).

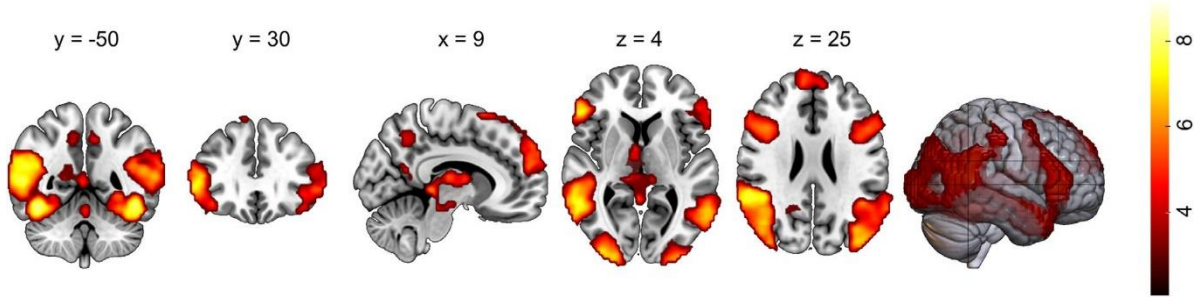


Figure 2.1. Social-random contrast activations. Brain activations from the whole brain voxelwise social – random content contrast. The statistical map was assessed with a cluster-based threshold of $Z > 3.1$, corrected at $p = 0.05$ (family-wise error correction) and projected onto a Montreal Neurological Institute (MNI) template. The color bar indicates Z values. The values above the slices indicate the coordinates in MNI space. Right is right.

Anatomical region	MNI			Peak Z	Cluster level	
	x	y	z		size	p
R. Inferior Temporal Gyrus	43	−55	−15	9.34	5933	<0.001
R. Fusiform Gyrus	42	−42	−15	8.54		
R. Middle Temporal Gyrus	60	−45	9	8.47		
R. Amygdala	21	−4	−13	8.46		
L. Fusiform Gyrus	−33	−42	−15	8.65	3089	<0.001
L. Inferior Temporal Gyrus	−36	−36	−15	8.58		
L. Inferior Occipital Gyrus	−26	−98	−10	8.26		
L. Hippocampus	−24	−12	−12	7.47	1062	<0.001
L. Amygdala	−27	0	−18	6.5		
L. Superior Temporal Pole	−42	18	−24	6.31		
L. Inferior Frontal Gyrus (operc.)	−48	15	24	5.92		
L. Inferior Frontal Gyrus (triang.)	−51	30	0	5.63		
L. Superior Medial Frontal Gyrus	−6	57	30	6.55	415	<0.001
R. Superior Medial Frontal Gyrus	6	54	30	6.27		
R. Supplementary Motor Area	9	15	66	5.35		
L. Middle Temporal Pole	−54	0	−18	6.57	210	<0.001
L. Precuneus	−9	−51	45	4.7	114	<0.001
R. Precuneus	3	−60	42	4.22		
L. Middle Frontal Gyrus	−39	3	54	4.87	61	0.0013
Vermis	0	−48	−21	4.87	38	0.0176

Table 2.1. Significant clusters from the whole brain voxelwise social – random content contrast.

Representational similarity results

The representational similarity analysis testing whether the degree of agreeableness correlates with the dissimilarity in neural representations of social and random content showed a significant positive correlation in the dorsomedial prefrontal cortex (dmPFC; $r = 0.596$, $p < .008$; Figure 2.2). Importantly, no other personality trait correlated either positively or negatively with the social-random dissimilarity measure (Supplementary figure 2.2).

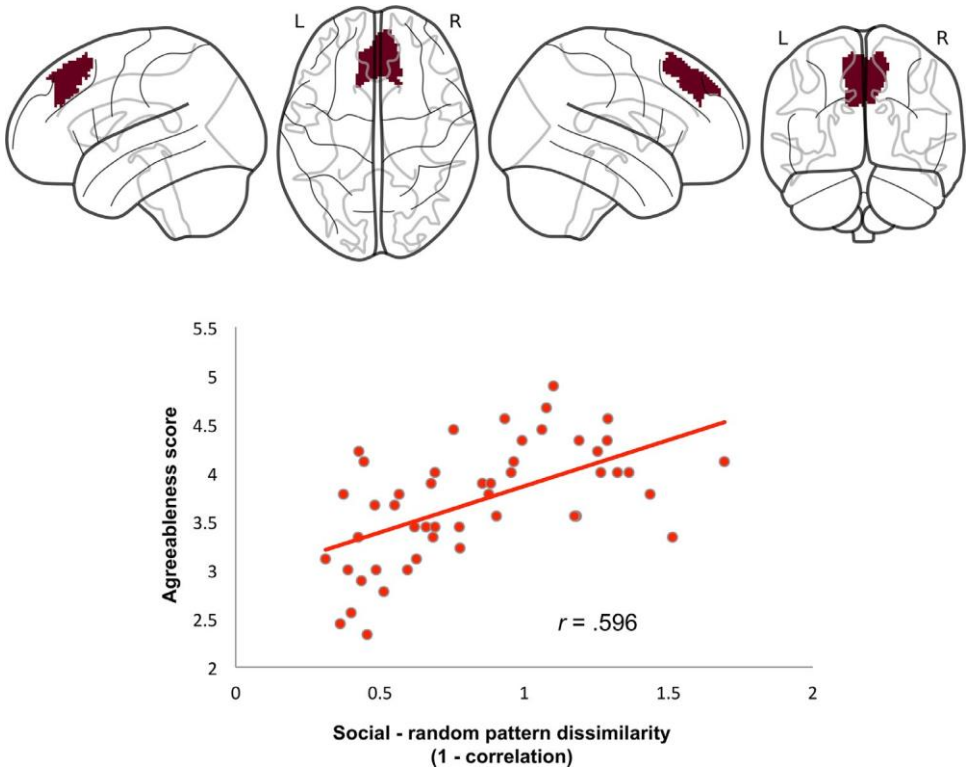


Figure 2.2. Agreeableness correlation with social random pattern dissimilarity. Results from the whole brain representational similarity analysis assessing the correlation between the degree of agreeableness and the dissimilarity in neural representations of social and random content. The upper panel shows the dorsomedial prefrontal cortex (dmPFC), the only region showing a significant positive correlation (permutation based *p*-values were Bonferroni-corrected over all tested regions - 200). The scatterplot in the lower panel shows the correlation between the participants' agreeableness trait scores and their social – random content pattern dissimilarity. The latter was calculated by correlating social and random response patterns and subtracting it from 1 (i.e., higher correlation is reflected as lower dissimilarity).

As an additional control analysis, we included age, IQ (assessed with an abbreviated Raven test (Bilker et al., 2012) and gender as covariates in the model predicting dmPFC pattern dissimilarity from agreeableness. Pattern dissimilarity in the dmPFC remained uniquely associated with agreeableness [$F(1, 44) = 21.62$, $p < .0001$, $\eta_p^2 = 0.329$]. To examine whether the representational similarity of social content, and separately of random content, in the dmPFC varied with levels of agreeableness, we computed the average similarity between social videos and between random videos, and correlated those with the agreeableness scores. Although neither was significant (social: $r = -0.039$; random $r = -0.067$), we show the representational similarity matrices (RSM) within and between conditions in the dmPFC with a median split on agreeableness data (HA: high agreeableness, LA: low agreeableness; median = 3.7) for visualization purposes (Figure 2.3A), to display the differences in social-random similarities between high and low agreeable individuals. We assessed the within and between conditions similarities between the two groups with a 2-way ANOVA (Figure 2.3B), and found a significant difference between the two groups only for the between condition similarity (group × condition: $F(2, 50) = 11.14$, $p < .001$, $\eta_p^2 = 0.31$). As expected, the low agreeableness (LA) group had significantly higher representational similarity between social and non-social content than the high agreeableness group (HA) (Tukey adjusted $p < .0001$). Finally, the dissimilarity correlation with task accuracy showed no significant results (Supplementary figure 2.2).

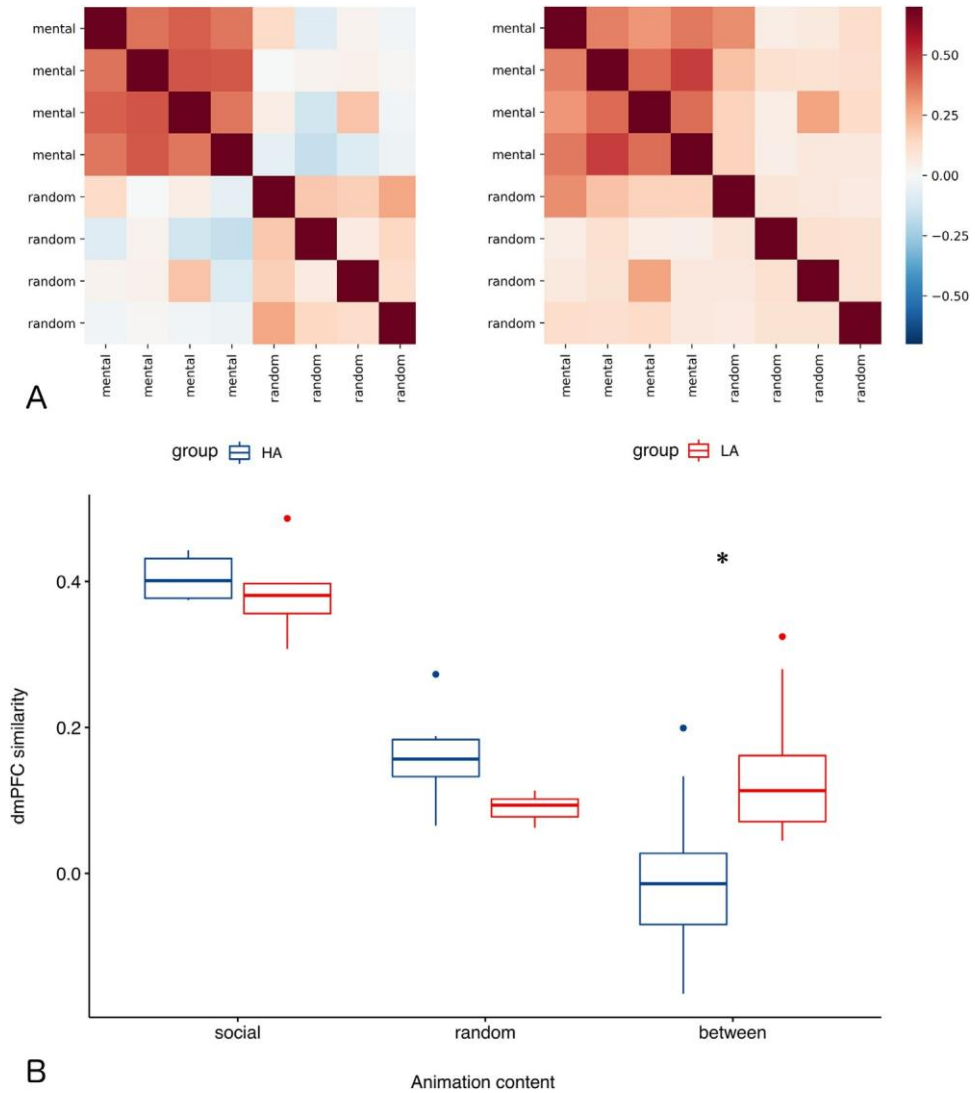


Figure 2.3. High agreeableness (HA) and low agreeableness (LA) individuals' representational similarity in the dmPFC. (A) The representational similarity matrices (RSM) within and between conditions in the dmPFC with a median split on agreeableness data (median = 3.7) for visualization purposes. The color bar indicates the Pearson's *r*. (B) Average within and between conditions similarities (Pearson correlation) in the HA and LA groups ($* p < .0001$).

Univariate results

The univariate analysis assessing whether the observed neural pattern dissimilarity associated with agreeableness could also be explained by univariate activation differences between social and non-social content showed no significant results (Supplementary figure 2.3).

Meta-analytic decoding results

The meta-analytic decoding was performed to provide an empirical interpretation of the pattern observed in the RSA unthresholded map. Out of the top 10 terms referring to cognitive constructs, five were related to social cognition, and all five had similarly high correlations for the GLM unthresholded map (Figure 2.4), suggesting that the cognitive terms associated with variations in agreeableness are also correlated with activity patterns during social vs. random processing.

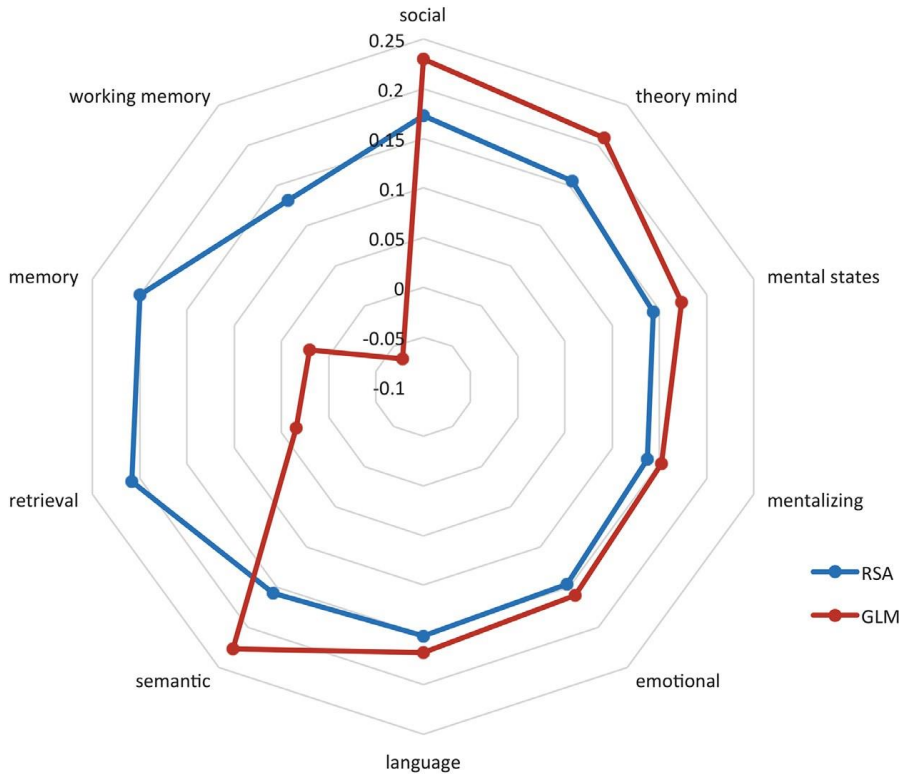


Figure 2.4. Meta-analytic decoding. Unthresholded RSA and GLM maps show similar associations with social cognitive terms from the Neurosynth database.

2.4 Discussion

Agreeableness trait is one of the five major dimensions of personality indexing individual variations in empathy, altruism and cooperation. Previous work has linked agreeableness with

socio-cognitive abilities and, in particular, with the ability to infer and reason about others' mental states, also known as ToM or mentalization (Allen et al., 2017; Nettle and Liddle, 2008). Recently, a new piece of evidence has strengthened this hypothesis by showing that individual differences in both agreeableness and ToM are related to variation in the same underlying neural network (Udochi et al., 2020). Here we complement and extend these findings by investigating how neural encoding of social information is related to agreeableness, therefore bridging the gap between neural and behavioral features of this trait. A straightforward prediction is that individuals high in agreeableness encode social and non-social information in a more dissimilar fashion with respect to individuals low in agreeableness, and primarily in brain regions involved during mentalization. To test this prediction we adopted a representational similarity analysis (RSA) approach that derives the degree of similarity between patterns of activation for different stimuli (Kriegeskorte et al., 2008). Since we were interested in assessing whether the degree of similarity between social and non-social content neural representation can predict the degree of agreeableness, we compared on an individual basis patterns of brain activation during a classic ToM animation task in which different shapes could interact in socially meaningful way, or randomly (Castelli et al., 2000; Wheatley et al., 2007). Consistent with our predictions, we found that more agreeable individuals had more distinct encodings of social and random animations, and correspondingly these encodings were more similar in individuals with low agreeableness. Critically, this positive association between agreeableness and representational dissimilarity of social and random content was significant only in the dmPFC, a brain region consistently related to attributions of mental states (Molenberghs et al., 2016; Schurz et al., 2014). This assumption was also confirmed by the meta-analytic decoding of our RSA results, which showed strongest associations with socio-cognitive topics. The dmPFC has been largely acknowledged as one of the key regions in the mentalizing network, although there is still no consensus regarding its functional role. Indeed, numerous meta-analyses have delineated the dmPFC involvement within different social, affective and cognitive functions (Binder et al., 2009; Kober et al., 2008; Kogler et al., 2020; Molenberghs

et al., 2016; Schurz et al., 2014; Spreng et al., 2009; Van Overwalle, 2009), making it even more difficult to reconcile all the observed data into a single picture. A different approach gathering evidence from causal and reverse inference studies, which are more suited for testing structure-function relationships, showed the dmPFC strongest functional link with social cognition and mental state inference processes (Lieberman et al., 2019). In particular, there have been a growing number of studies adopting multivariate pattern analyses that have characterized and/or decoded the neural representation of social information within the dmPFC (e.g., Corradi-Dell'Acqua et al., 2014; Dungan et al., 2016; Skerry and Saxe, 2015; see Wagner et al., 2018 for a revew). In a more comprehensive study, Tamir et al. (2016) investigated what psychological dimensions that organize the understanding of mental states shape also their neural representations in the brain, and one specific component, "rationality", loading highly on dimensions such as emotion, experience and warmth, predicted reliably the neural patterns in the dmPFC. Although in our study we did not consider different dimensions of social information that would allow us to capture what particular dimension is represented differently in individuals with different levels of agreeableness, we speculate that social videos conveyed emotional information that could be inferred from the inter- actions between the shapes (e.g., fear, surprise, compassion, happiness, irritation, etc.). Highly agreeable individuals might have spontaneously inferred these emotional states, increasing the pattern dissimilarity in the dmPFC between social and non-social videos, contrarily to individuals low in agreeableness. This speculation would also be in line with the supramodal representations of perceived emotions observed in the dmPFC (Peelen et al., 2010; Skerry and Saxe, 2014). This hypothesis should however be tested with an appropriate task design, in particular by modulating the emotional content across stimuli and assessing its impact on neural representations in individuals with different levels of agreeableness. Previous research on inter-individual differences in mentalization ability has typically focused on neurological and psychiatric conditions, evidencing its importance in every day functioning (Baron-Cohen, 1995; Kerr et al., 2003; Richell et al., 2003; Snowden et al., 2003; Stuss et al., 2001). Nevertheless, it is

well acknowledged that there are substantial mentalization differences also in healthy adults, yet these are difficult to capture behaviorally, mainly due to ceiling effects observed on standard laboratory tasks (Koster-Hale and Saxe, 2013). To overcome this issue, one possible strategy is to assess neural variations to ambiguous stimuli that induce spontaneous mentalization, which we implemented in our study. Few other studies have adopted this approach to investigate inter-subject variability in mentalization (Moessnang et al., 2017; Moriguchi et al., 2006; Udochi et al., 2020; Wagner et al., 2011), and most of them observed modulation of activity in the dmPFC in relation to trait-like scores. Specifically, higher empathy (Wagner et al., 2011) and perspective-taking scores (Moriguchi et al., 2006), and lower autism trait (Moessnang et al., 2017) were associated with a greater recruitment of dmPFC, suggesting a more enhanced processing of social information. In line with this evidence, our finding provides additional understanding regarding the distinctiveness of social information in the dmPFC and how it underlies individual differences. In particular, by comparing patterns of activity during social and non-social content processing, we observed that the degree to which the two information are distinguished in the dmPFC predicts individual variations in agreeableness, a trait related to mentalizing ability (Allen and DeYoung, 2017). It is important to point out that this result can reflect a different encoding of social and/or non-social information in individuals with different levels of agreeableness, or alternatively it can reflect differences in processes related to the task at hand (e.g., emotion recognition but also attentional processes, cognitive control, etc.). However, given the greater neural similarity for social videos observed in the dmPFC (Figure 2.3), and the frequent involvement of this region in processing social interactions (e.g., Wagner et al., 2016), we are confident that differences in social process are a more plausible interpretation. Echoing previous correlational fMRI studies, we addressed a potential alternative account according to which the agreeableness trait could be predicted by average activation differences between social and non-social content. Results from this analysis showed no significant trait prediction, suggesting that individual variations in agreeableness are better reflected within representational differences of social and non-social

information than the overall difference in activity they evoke. Still, we do not know if these differences in social content representation are functionally relevant, since we did not observe similar relations with task performance, probably due to an insensitive behavioral measure we adopted. However, it would be plausible to expect more dissimilar encoding of social vs. non-social information in relation to better understanding of the events depicted in the animations. Future studies should implement more detailed task performance measures to investigate the functional importance of the underlying neural representations. One potential limitation of this study, which probably reduced the power of our main brain-behavior correlation analyses, is the moderate sample size. Although we have adopted a sampling strategy (i.e., we selected our participants from a larger sample based on their personality scores), which was found to increase the power to detect moderate effect sizes in small samples without inflating the false positive rate (de Haas, 2018), future studies should try to replicate this finding in larger samples. Lastly, but equally important, the observed task-related representational difference underlying agreeableness argues for a different, behavior-based approach in studying the cognitive and neural mechanisms behind personality. In personality neuroscience, agreeableness has been one of the least studied traits (Allen and DeYoung, 2017) and until recently its neurobiological substrates have been explored only among stable brain features like structural or resting-state measures (e.g., Cai et al., 2020; Riccelli et al., 2017). Although these studies were rarely underpowered in terms of sample size, they reported very divergent findings, or even null-results (e.g., Avinun et al., 2020; Dubois et al., 2018). One possible explanation for these inconsistent findings may involve the complexity of behaviors associated with agreeableness and its link with various processes such as empathy, self-regulation and motivation (Graziano and Tobin, 2016). This could result in multiple brain systems mediating variability in the agreeableness trait, which would therefore be hardly identifiable with structural and resting-state brain measures. This explanation is in line with a recent functional connectivity study (Liu et al., 2019) in which the agreeableness trait was found to be correlated to a widely distributed connectivity pattern comprising the largest number of

connections, almost four times that of other traits. Therefore our finding linking agreeableness to mentalization probably captures only one aspect of agreeableness and future studies should implement more itemized personality measures to identify other cognitive and neural determinants of agreeableness.

2.5 Conclusion

Despite its limitations, the present study represents a step forward in characterizing the neural determinants of agreeableness, a personality trait highly related to socio-cognitive abilities. Our results suggest that neural representations of social content in the dmPFC can vary among individuals with different levels of agreeableness in a functionally relevant way: more similar representations of social and non-social content predict lower agreeableness scores, revealing the link between neural and behavioral mechanisms underlying agreeableness. Additionally these sorts of connections between personality traits and specific cognitive abilities provide new opportunities for the development of more objective personality measures.

CHAPTER 3

Study 2: Personality and sex differences in social interaction perception: an ERP and source imaging study

---

Pisanu, E.*, Arbula, S.*, & Rumiati, R. I. (under review). *Personality and sex differences in social interaction perception: an ERP and source imaging study*

Abstract

Agreeableness is one of the five personality traits which is associated with socio-cognitive abilities. In the present study, we assessed the impact of agreeableness on the perception of social interaction while taking into account sex differences that might moderate this relation. Sixty-two young adults, previously selected from a larger sample screened for agreeableness, underwent EEG recording while performing a social interaction detection task. Behavioral results revealed that higher agreeableness scores predicted better ability to detect social interactions only in males. ERP analysis led to a significant correlation between agreeableness and contrast for social and non-social conditions, both within the entire sample and separately in females and males. Specifically, higher levels of agreeableness were associated with a deeper neural processing of social versus non-social stimuli within all three samples. This modulation occurred relatively early in time, however, in males, it was more extensive, and also predictive of their performance. Source estimation procedures suggested the engagement of right fronto-parietal regions for the ERP-agreeableness association in all three samples. Our findings complement and extend previous research on personality and social skills interrelation, evidencing the impact of agreeableness on the neural mechanisms underlying the detection of social interactions, and revealing the involvement of sex.

3.1 Introduction

Agreeableness is one of the five personality traits defined by the Five Factor model (Costa and McCrae, 1992a; Goldberg, 1990) which has been associated with socio-cognitive abilities including empathy and theory of mind (ToM; Allen and DeYoung, 2017; Nettle and Liddle, 2008). As all personality traits, agreeableness is characterized by two extreme poles and each person positions herself between them depending on her own individuality. Individuals high in agreeableness tend to be altruistic, cooperative, and understanding of others' perspectives. In contrast, those low in agreeableness may exhibit poorer social skills, self-focus, and even

aggression in extreme cases (Allen and DeYoung, 2017; DeYoung, 2010; Graziano et al., 2007; John et al., 2008). Neuroimaging studies have strengthened the association between agreeableness and social cognition by showing an overlap of their neural networks (Allen et al., 2017; Udochi et al., 2022). In our previous fMRI study, (Arbula et al., 2021), we investigated how neural variations during social content encoding are related to differences in agreeableness. Specifically, by using a classic ToM task, in which social and socially neutral animations were shown (Castelli et al., 2000), we observed a significant relationship between the similarity in neural representations for the two types of content and the levels of agreeableness. We found that individuals with lower agreeableness had more similar neural patterns for social and neutral content, while those with higher agreeableness had more distinct patterns, specifically in the dorsomedial prefrontal cortex (dmPFC), a brain region typically associated with ToM abilities (Molenberghs et al., 2016; Sabbagh and Bowman, 2018; Saxe and Baron-Cohen, 2006; Schurz et al., 2014, 2021). This finding suggests that individual differences in agreeableness are associated with differences in social information processing. However, due to the complexity of the stimuli, we could not disentangle the different dimensions of social information that might be distinctly represented in individuals with different levels of agreeableness.

To overcome this limitation, in the present study we focused on the detection of social interactions, a pivotal aspect of social information processing. The ability to perceive and understand social interactions emerges in early childhood and is shared among primates (Sliwa and Freiwald, 2017). Social stimuli, especially those depicting social interactions or groups of individuals, enjoy privileged access to visual processing and awareness, aiding the brain in information extraction (Isik et al., 2020; Su et al., 2016; Vestner et al., 2018). Social relationships are vital for our survival and well-being, as evidenced during periods of lockdown when enforced isolation and limited social interactions had a significant impact on mental health, leading to increased symptoms of anxiety and depression. This effect was observed across the general population and was particularly pronounced among patients and caregivers who had to endure

separation from their families to mitigate the risk of contagion (Ludwig-Walz et al., 2023; Musa et al, 2023; Pisanu et al., 2022). Considering the significance of social interactions, the brain spontaneously extracts information about them, albeit in a relatively slow manner (~300 ms), potentially relying on top-down computations needed to acquire relational information among multiple agents (Isik et al., 2020).

Based on these premises, we hypothesized that personality, and in particular the agreeableness trait, might influence the neural computations driven by top-down processes with a similar time dynamic. To test our hypothesis, we assessed the impact of agreeableness on social interaction detection by means of electroencephalography (EEG). The social interaction construct was operationalized as a relational process of two individuals acting while looking at each other (i.e., "mutual gaze") or by looking at the same object/point of interest (i.e., "joint attention"), as in Isik and colleagues (2020) MEG study. These two types of interactions are among the first to be identified by infants (Guellaï et al., 2020; Mundy, 2018), and as such are considered the bases of social engagement.

We predicted that variations in agreeableness would be correlated with differences in the sensitivity to detect social interactions, along with modulations in event-related potential (ERP) responses related to the social compared to non-social condition. Furthermore, previous studies reported sex-related differences in social processing, with females showing a better and faster performance in detecting social and emotional stimuli (Farris et al., 2008; Montagne et al., 2005), and higher responsivity of the action observation network (Alaerts et al., 2011). However, there is controversy in the literature in this regard, as the results depend on the specific task and process under investigation, and behavioural differences are not consistently observed (Isernia et al., 2020; Pavlova et al., 2010). Moreover, there is evidence of neuroanatomical (Gur et al., 2002) and neurobiological sources (Cahill, 2006; Perrin et al., 2009; for a review: Paletta et al., 2022; Proverbio, 2021) contributing to sex differences in social cognition. Therefore, we also anticipated potential inter-individual differences based on sex.

To examine individual differences in ERP responses, we opted for a massive univariate approach by using an EEGLAB plugin (i.e., LIMO toolbox) that allowed us to analyse and explain variance by considering both intra- and inter-subject variability, as well as the effect of the experimental condition (Pernet et al., 2011). We considered this method to be less restrictive compared to traditional approaches, as it does not require a priori selection of specific time points, electrodes, or components to analyze, which is more suitable when there is limited existing literature on the processes being investigated, as in our case.

3.2 Methods

3.2.1 Participants recruitment and personality assessment

Sixty-two young adults (27 males) with a mean age of 24.03 (SD = 4.34, range = 18-35 years) were selected from a larger sample (N = 462) based on their agreeableness scores to increase variability within the study: 20 subjects had scores above and 20 below 1.4 SD from the Italian population mean (Ubbiali et al., 2013). As personality traits are continuous variables, in order to ensure that our sample accurately captures measurements at either tail of the distribution, deliberate efforts were made to select individuals with diverse scores along the z-score continuum, encompassing both negative and positive values. The screening process involved administering the Italian adaptation of the Big Five Inventory (Ubbiali et al., 2013), a questionnaire comprising 44 items rated on a five-point Likert scale, assessing the five personality traits from the Five Factor model (Extraversion, Neuroticism, Openness, Conscientiousness, Agreeableness). The questionnaire was advertised through different social media and email. None of the participants reported a history of neurological or psychiatric disorders. They all provided written informed consent and received 20 euros for their participation. The study was conducted in accordance with the guidelines of the Declaration of Helsinki and approved by the SISSA's ethics committee.

3.2.2 Social interaction perception task

Participants viewed 48 visually matched target images depicting real-world scenes showing

two people either engaged in a social interaction or acting independently (24 stimuli per condition, namely social and non-social). Participants indicated whether each target image represented a social interaction by pressing a key button. The operational definition of social interaction aligned with Isik and colleagues (2020), who originally developed the stimuli: two agents with actions directed towards each other (mutual gaze) or a shared point of interest (joint attention).

Participants were seated 60 cm away from the screen in a dark and quiet room. Stimuli were presented at 9 × 5 degrees of visual angle on a gray background, in the center of the screen by using PsychoPy3 (v2020.1.1). The task began with an explanation of the definition of social interaction using 4 example images and 4 practice trials that were not included in the analyses. Participants completed 240 trials (120 per condition) interspersed with a self-paced pause every 60 trials. Stimuli were arranged in a pseudo-random order to avoid more than two repetitions of the same condition. Each trial began with a fixation cross (1000 ms), followed by the target image (500 ms) (Figure 3.1); after the target offset, a blank screen was shown (2000 ms) after which they had a maximum of 2000 ms to indicate their response by pressing either the "f" or "j" key on a QWERTY keyboard using the index fingers of both hands. The response keys were counterbalanced across participants.

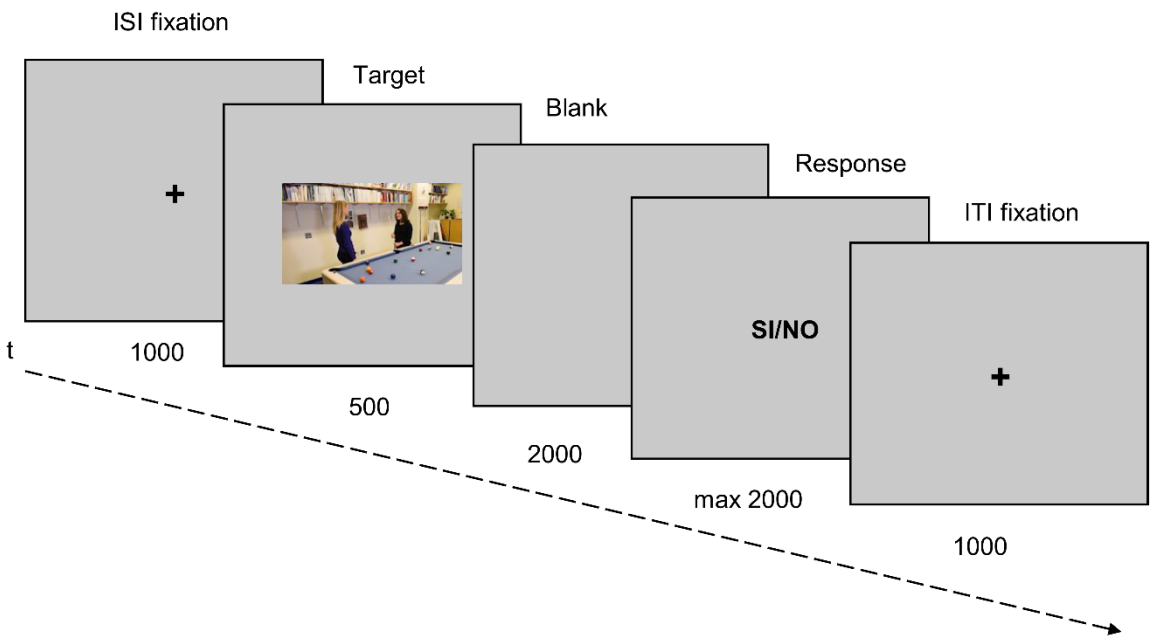


Figure 3.1. Schematic representation of the social perception interaction task. Stimuli succession is represented left to right, with the respective screen presentation time expressed in milliseconds below each stimulus. The image is one of the targets used representing social interaction.

3.2.3 EEG recording and preprocessing

EEG data was collected using a cap with 64 Ag/AgCl active electrodes based on the International 10-20 system (Klem et al., 1999). A BioSemi Active-Two amplifier system was used for signal amplification. The ActiView acquisition software (version 707, Biosemi, Amsterdam) was employed for data visualization and storage. Electrode offsets were maintained within ±50 mV, and the signal was sampled at a rate of 1024 Hz with a 24-bit digitization resolution. Eye movements and blinks were monitored using horizontal and vertical electrooculogram (EOG) recorded from additional electrodes placed below and above the left eye.

Data were pre-processed offline in EEGLAB 14.1.1b (Delorme and Makeig, 2004, https://sccn.ucsd.edu/eeglab), in Matlab R2021a (https://uk.mathworks.com/products/matlab.html). Continuous EEG recordings were resampled offline at 250 Hz, and band-pass filtered (cut-off frequencies: 0.25 and 45 Hz) using two zero-phase Hamming windowed sinc FIR filters (transition bandwidth: 0.5 and 10 Hz, passband edges: 0.5 and 40 Hz). Bad channels, poorly correlated with the others, were removed thorough an EEGLAB *clean_rawdata* plugin (autocorrelation parameter = 0.8, excluding on average 1.9 channels, SD = 2.1). Epochs ranging from 1000 ms before and 2500 ms after the stimulus onset were then extracted. An independent component analysis (ICA; Makeig et al., 1996) was performed to remove those components with more than 60% probability of being eye, motor, channel noise, heart, or line noise using ICLabel (Pion-Tonachini et al., 2019).

After the ICA, removed channels were replaced using spherical spline interpolation (Perrin et al., 1989) and data were re-referenced to a common average reference. Finally, all epochs were corrected to a pre-stimulus baseline of 200 ms and were inspected for artifacts based on different statistics (amplitude, linear trend, joint probability and kurtosis; Delorme et al., 2007); epochs

containing artifacts based on one or more of these criteria were excluded from further analyses. Overall, 9.7% percent of the trials were rejected for violating artifact criteria (with no asymmetry across conditions). The average number of epochs used for the analysis in both conditions was 108, ranging from 66 to 115 epochs. Two participants (one male and one female) were excluded from the EEG analysis due to technical problems during the recording.

3.2.4 Behavioral data analysis

We collected accuracy and reaction time (RT) data. RTs were filtered to exclude errors, anticipated responses (< 150 ms), and outliers exceeding 3 standard deviations from the subject's mean for each condition. To better address the research question of whether agreeableness may predict variations in sensitivity to detect social interactions, discriminating them from non-social interactions, we computed and incorporated signal detection theory indices into our behavioral models. The sensitivity index (*d'*) was used to quantify the participant's ability to discriminate social from non-social stimuli. The formula for calculating *d'* is: $d' = Z_{HIT} - Z_{FA}$, where Hit represents the proportion of social trials to which subjects responded "social" (hit rate) and FA represents the proportion of non-social trials to which subjects responded "social" (false-alarm rate). Both rates were Z-transformed. We also calculated the bias for social responses, indicating the subject's tendency to respond "social", using the formula: $-0.5*(Z_{HIT} + Z_{FA})$ (Macmillan and Creelman, 1990; Stanislaw and Todorov, 1999).

Next, we performed a linear mixed models analysis in R v.4.1.1 (R Core Team, 2020) to examine the impact of agreeableness on participants' sensitivity to differentiate social from non-social stimuli. The sensitivity index (*d'*) was used as the dependent variable, agreeableness was included as a fixed effect, and subject intercepts were considered as random effects. A second model was created to investigate whether the relationship between agreeableness and sensitivity differed between males and females, by adding sex as an additional fixed effect. The same analyses were repeated for the bias measure, examining the presence of a potential bias toward social

responses. To ensure that any observed effects were specific to agreeableness and not influenced by other personality traits, the other four traits were then included as additional predictors in the models. Assumptions of linearity, homoscedasticity, and normality were assessed by inspecting residual plots.

3.2.5 ERP analysis

ERP analysis was performed using Matlab R2021a along with EEGLAB 2020_0 and LIMO EEG plugin (v 2.0, https://github.com/LIMO-EEG-Toolbox/limo_tools). A massive univariate approach was used by running two levels of analysis (Pernet et al., 2011).

First-level GLM analysis

At the first level, parameters of a General Linear Model were estimated by using a standard ordinary least squares solution for each subject, at each time point and each electrode independently, and separately for the two experimental conditions, namely social and non-social. To assess the differential activity evoked by the two conditions, the corresponding ERP amplitudes were subtracted to create a contrast. The considered time window ranged from 0 to 1000 ms, time locked to stimulus onset. To maintain higher inter-individual variability, erroneous trials were not excluded from the model, since we expected accuracy to vary in relation to agreeableness scores.

Second level GLM analysis

At the group level, a robust regression analysis was conducted on the contrasts between social and non-social ERP responses, using agreeableness as a regressor. The same analysis was repeated separately for males and females to explore potential sex differences. An additional robust regression was performed for the same ERP contrasts and *d'* within the time windows in which we observed a significant effect of agreeableness, to investigate whether the effects associated with agreeableness might be predictive of task performance.

All reported results underwent correction for multiple testing using spatial-temporal clustering with a cluster forming threshold of $p = .05$ (Maris and Oostenveld, 2007; Pernet et al., 2015).

3.2.6 Source analysis

Cortical source activations were estimated using the Brainstorm software, which employs a distributed dipoles model for fitting (Tadel et al., 2011). The locations of the scalp-recorded electrical activity were estimated using the dynamic statistical parametric mapping (dSPM) method (Dale et al., 2000), which uses minimum-norm inverse maps. Single subject noise covariance matrices and individual noise standard deviations at each location were calculated and estimated by means of the pre-stimulus baseline intervals (-200 to 0 ms) from single trials (Hansen et al., 2010).

For the head model, the boundary element method (BEM) was employed with the default parameters in Brainstorm, using the ICBM152 anatomy since individual anatomies were not available. The BEM model provides three realistic layers and representative anatomical information (Gramfort et al., 2010; Stenroos et al., 2014). The source estimation was performed using the option of constrained dipole orientations, which models one dipole oriented perpendicular to the cortical surface for each vertex (Tadel et al., 2011).

For the estimation of active sources, the single-trial EEG data for each participant were averaged to obtain the subject average. Sources were then separately estimated for the two experimental conditions (social and non-social) and subtracted to create a contrast at the subject level. On each source, we applied a z-score normalization against the baseline period (-102, -4 ms) to reduce the influence of inter-individual fluctuations in neural current intensity caused by non-relevant anatomical or physiological differences and rectified them in absolute values. To investigate the cortical sources underlying the ERP effects associated with agreeableness, the cortex surface was divided into 50 non-overlapping regions using the surface clustering option in Brainstorm. We chose to adopt a whole-brain parcellation approach, which provides a suitable compromise between the limited ROI-based method and the overly corrected whole-brain approach.

After extracting the contrast values from the 50 regions within the time window of the ERP-

agreeableness significant effect, a robust correlation analysis was conducted between these contrast values and agreeableness scores. The correlation analysis was performed for the entire sample and separately for females and males. A skipped Pearson correlation (Wilcox et al., 2018) was performed by means of the Robust Correlation toolbox (https://github.com/CPernet/Robust-Correlations; Pernet et al., 2013), which considers the overall structure of the data to protect against bivariate outliers. The statistical significance is tested through the percentile bootstrap 95% confidence intervals and results are reported with the FDR correction for multiple comparisons.

3.3 Results

3.3.1 Behavioral results

The linear model examining the relationship between agreeableness and the ability to discriminate between social and non-social stimuli, as measured by the *d'* index, did not yield a significant effect ($F(1, 60) = .29$, $p = .59$, $\eta^2 = .004$). However, when sex was included as an additional fixed effect in the model, a trend emerged showing that higher agreeableness scores predicted better ability to detect social interactions differently across males and females ($F(1, 58) = 3.62$, $p = .062$, $\eta^2 = .06$; Figure 3.2). Post-hoc analysis examining the slopes of the agreeableness trend separately for males and females revealed a positive slope for males, although it was not significantly different from zero (males slope = 0.173, $CI_{95\%}$ (-0.016 0.363), $p = 0.073$; females slope =-0.0572, $CI_{95\%}$ (-0.2078 0.0935), $p = 0.451$). Noteworthy, when controlling for the other personality traits, the result found for agreeabeleness persists ($F(1, 50) = 3.75$, $p = .06$, $\eta^2 = .07$) while no other significant or nearly significant associations were found (all p's > .094).

An independent t-test was carried out to check whether the abovementioned results might be affected by differences in the level of agreeableness between males and females and no significant difference emerged ($t = 0.075$, $p = .94$), indicating a well-balanced sample in terms of agreeableness. Finally, no differences were found between males and females in terms of agreeableness-related modulations of the response bias measure for social responses ($F(1, 58) =$

.59, p = .44, η^2 = .01). Visual inspection of residual plots did not reveal any obvious deviations from linearity, homoscedasticity or normality.

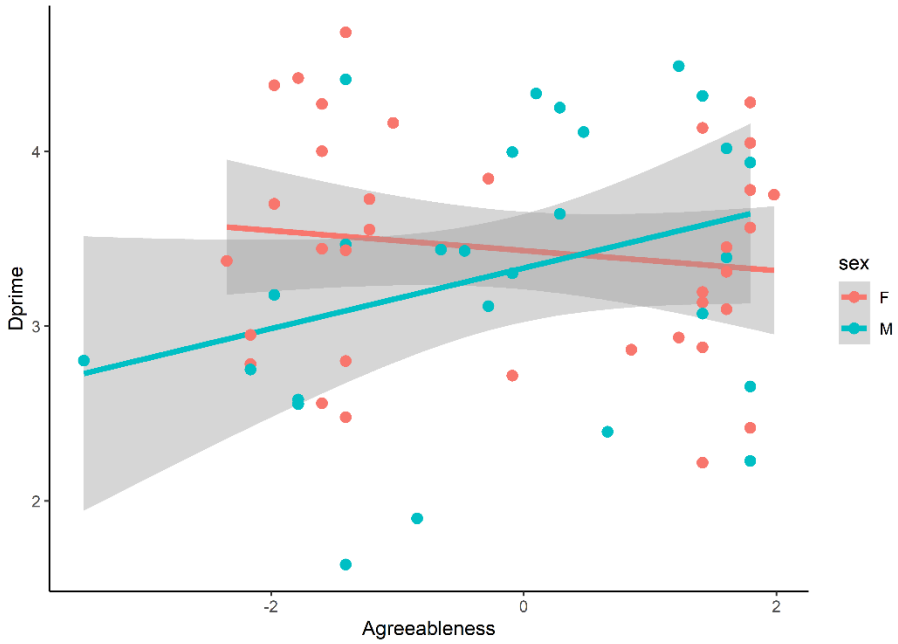


Figure 3.2. Scatterplot showing the relationship between agreeableness and the *d'* measure. Agreeableness is represented in z scores, the red line represents females, the blue line males; each spot is a subject. The grey shadow around the trendline represents the standard error (statistics and *p* values are reported in the text).

3.3.2 ERP results

Within the entire sample, we observed a significant correlation between agreeableness and the ERP social - non-social contrast in 2 clusters (cluster 1: 191-227 ms, max F value 13.41 at 215 ms on channel P10, corrected *p* value 0.045; cluster 2: 199-223 ms, max F value 9.82 at 211 ms on channel C3, corrected *p* value 0.003; Figure 3.3A). These results showed that agreeableness modulates the differential neural response such that subjects with higher agreeableness scores had more negative contrast values over the left fronto-central electrodes and, as a consequence of the dipole effect, more positive contrast values over the right parietal electrodes (Figure 3.3B). The opposite pattern was associated with low agreeableness scores. To inspect more carefully the direction of these contrasts across different levels of agreeableness, we divided the entire sample

in two groups (high A: high agreeableness, low A: low agreeableness) by means of the median-split of their agreeableness z-scores, and plotted their averaged ERPs (i.e., first level beta values) for the social and non-social condition (Figure 3.3C). The plot reveals that the high A group exhibited a larger negative amplitude for social vs. non-social trials over the left fronto-central electrodes while the low A group showed a greater negativity for the non-social stimuli. The same pattern, albeit reverse, was found over the right parietal electrodes reflecting mainly a greater positivity for non-social *vs.* social trials in the low A group.

As regards the female sample, a significant cluster was found (195-227 ms, max F value 19.03 at 199 ms on channel CP6, corrected *p* value 0.011; Figure 3.4A) which overlapped with cluster 1 of the entire sample. Accordingly, more agreeable females showed more positive contrast values reflecting a larger positive amplitude for social vs. non-social trials over the right centro-parietal electrodes, while less agreeable females showed more negative contrast values indexing a larger positive amplitude for non-social *vs.* social trials (Figure 3.4B, C).

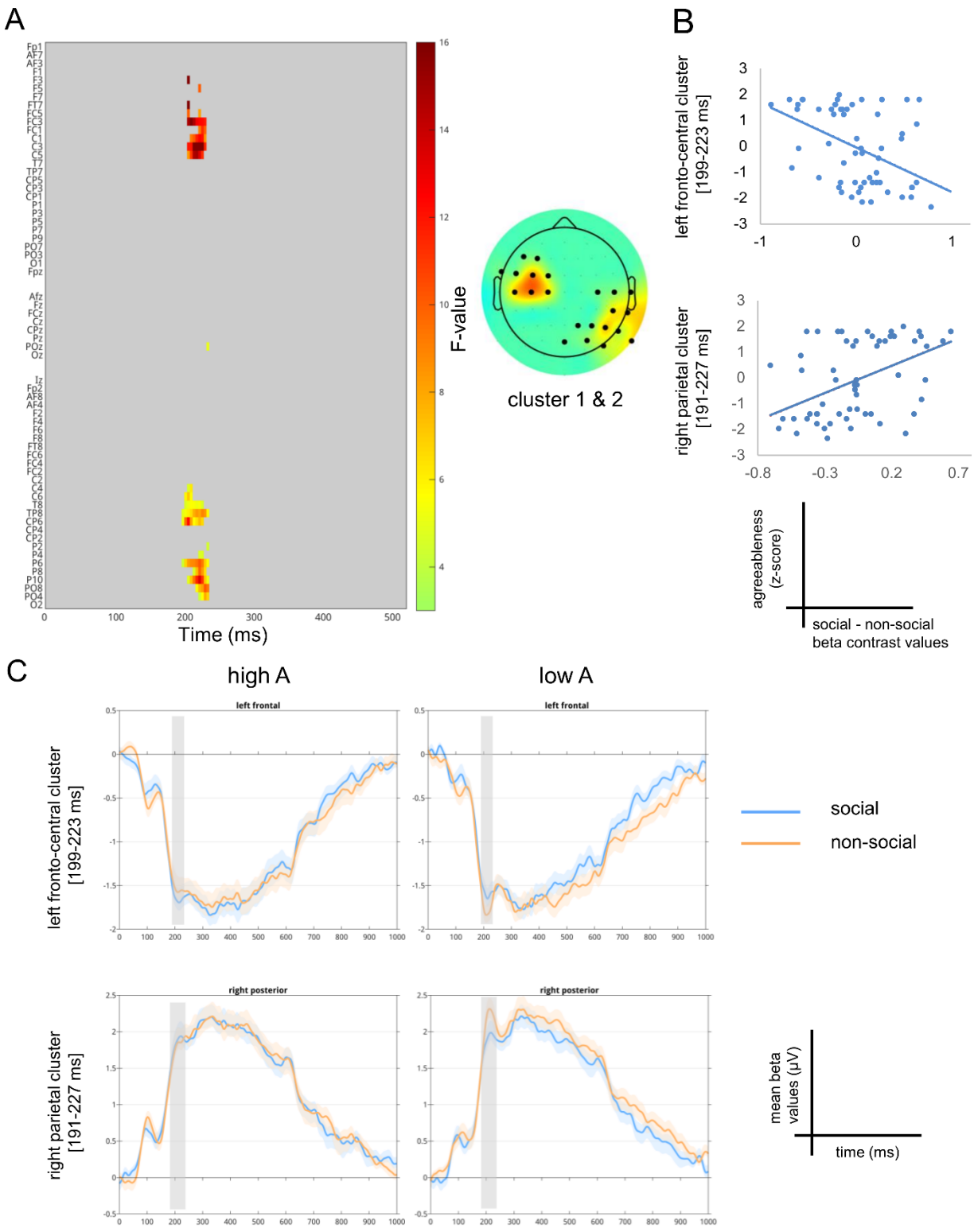


Figure 3.3. ERP-Agreeableness regression results for the entire sample. (A) Raster plot depicting significant channels and time-points for the agreeableness regression colored based on their F values. On the right the topographical representation of the F values averaged across the significant time-window (191-227 ms). (B) Scatterplots representing the correlation between the beta contrast values averaged across the significant time window and the agreeableness trait, separately for the two clusters. (C) Trace plots showing the mean beta values for the social (in blue) and non-social conditions (in orange) averaged over the significant electrodes from each of the two clusters, separately for the

high A and low A group.

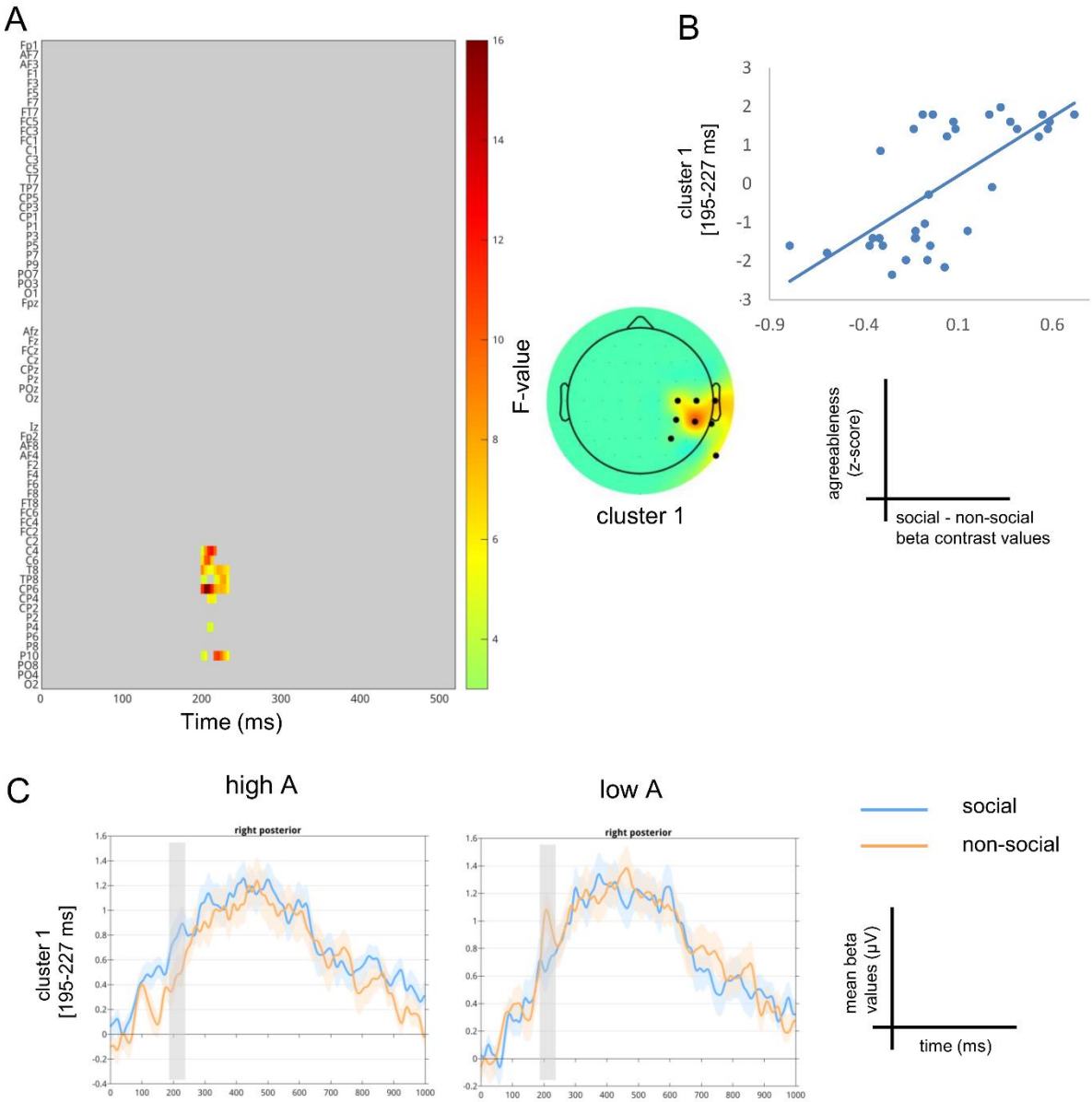


Figure 3.4. ERP-Agreeableness regression results for the female sample. (A) Raster plot depicting significant channels and time-points for the agreeableness regression colored based on their F values. On the right the topographical representation of the F values averaged across the significant time-window (195-227 ms). (B) Scatterplot representing the correlation between the beta contrast values averaged across the significant time window and the agreeableness trait. (C) Trace plots showing the mean beta values for the social (in blue) and non-social conditions (in orange) averaged over the significant electrodes, separately for the high A and low A group.

Within the male sample, three significant clusters were found within three different time windows (cluster 1:140-160 ms, max F value 15.58 at 152 ms on channel C3, corrected *p* value

0.008; cluster 2: 367-391 ms, max F value 11.78 at 383 ms on channel TP8, corrected *p* value 0.015; cluster 3: 703-730 ms, max F value 12.03 at 723 ms on channel FC2, corrected *p* value 0.006; Figure 3.5A). In cluster 1, comprising the left central electrodes, males with higher levels of agreeableness showed more positive contrast values than males with low agreeableness (Figure 3.5B). This result was led by a larger positive amplitude for social *vs.* non-social trials in more agreeable males which was inversed in less agreeable males showing a greater positivity for non-social stimuli (Figure 3.5C).

Within clusters 2 and 3, the pattern of the correlation between agreeableness and the ERP social – non-social contrast values matched the previous clusters found for the entire and the female sample: higher agreeableness corresponded to more positive contrast values over the right temporo-parietal electrodes (cluster 2) and to more negative contrast values over the left fronto-central ones (cluster 3) (Figure 3.5B). Likewise, these contrasts reflected greater positivity (cluster 1) and negativity (cluster 2) for social trials in highly agreeable male participants, and for non-social trials in male participants with low agreeableness (Figure 3.5C).

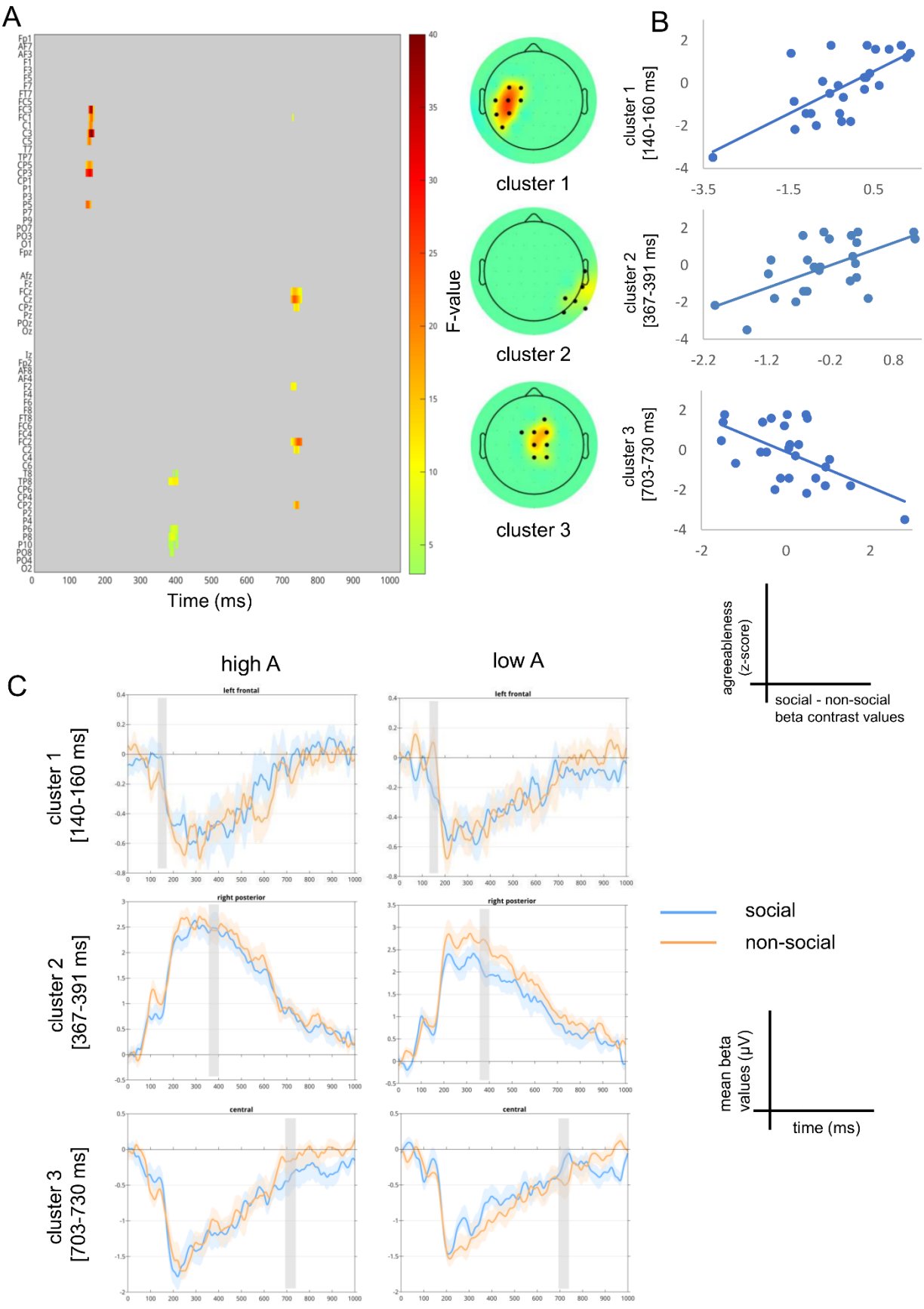


Figure 3.5. ERP-Agreeableness regression results for the male sample. (A) Raster plot depicting significant channels and time-points for the agreeableness regression colored based on their F values. On the right the topographical

representation of the F values averaged across the significant time-windows in the three clusters. (B) Scatterplots representing the correlation between the beta contrast values averaged across the significant time window and the agreeableness trait, separately for the three clusters. (C) Trace plots showing the mean beta values for the social (in blue) and non-social conditions (in orange) averaged over the significant electrodes from each of the three clusters, separately for the high A and low A group.

Additional regression analyses investigating whether the observed effects of agreeableness are predictive of task performance evidenced this association only within the male sample. A significant positive correlation (max F value 29.91 at 141 ms on channel P1, corrected *p* value <.001) between the *d'* measure and the ERP social – non-social contrast values was observed within the first cluster (140 – 160 ms) in which we found a positive association with agreeableness. In particular, these results pointed in the same direction as the correlation with agreeableness: higher *d'* was associated with more positive ERP contrast values (Figure 3.6), reflecting a larger ERP amplitude for social *vs.* non-social stimuli.

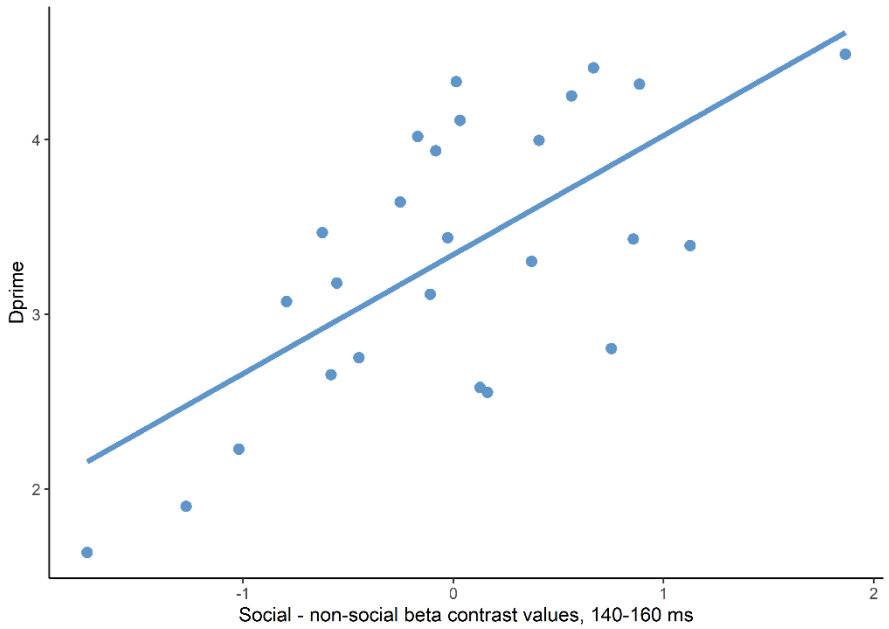


Figure 3.6. Scatterplot representing the correlation between the beta contrast values averaged across the significant time window (140-160 ms) and the *d'* measure within the male sample.

3.3.3 Source Localization

Within the entire sample, a positive correlation was found between agreeableness and the activation for the social – non-social contrast in the parcel placed in the right fronto-parietal cortex (rFPC; r = .35, corrected p = .05; Figure 3.7A, B) in the averaged time window of the ERP effect (191-227 ms). Thus, the source activation in the rFPC was greater for social *vs.* non-social stimuli in more agreeable subjects, which mirrors the association found with the right parietal ERP component. The same association in the rFPC was observed within the female sample after the robust regression was performed in the averaged time window of the female ERP effect (195-227 ms), which however did not survive the FDR multiple comparison correction (r = .43, uncorrected p = .017, corrected p = .42; Figure 3.7C), therefore this result should be taken with caution.

As regards males, a significant positive correlation emerged again in the same parcel (rFPC; r = .52, p = .05; Figure 3.7D), within the time window covering the second cluster (367-391 ms), reflecting a greater activation for social vs. non-social stimuli in more agreeable males. No other parcel survived the correction for multiple comparisons in the time windows corresponding to the first and third cluster.

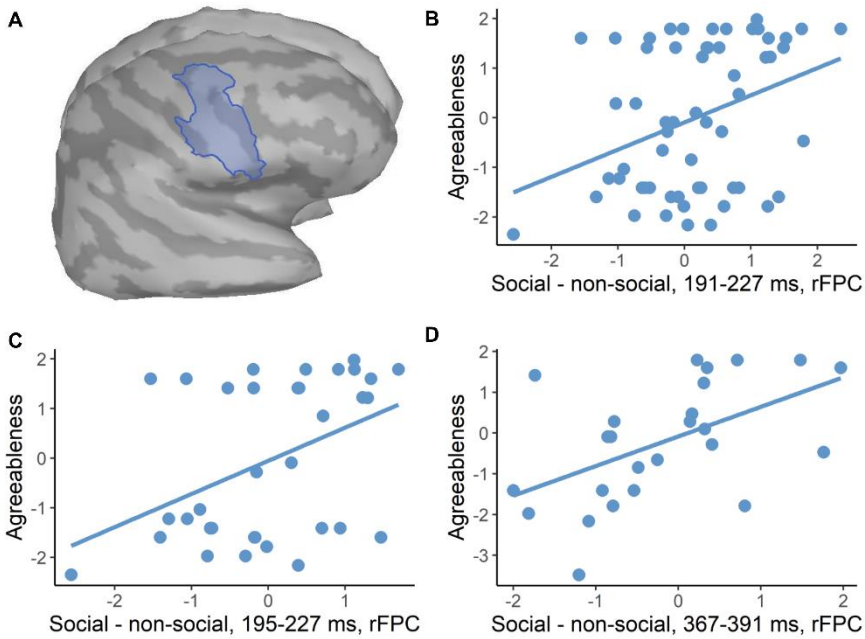


Figure 3.7. Results from the skipped Pearson correlations between agreeableness and source activations. (A) The

rFPC source. (B), (C), (D) Scatterplots representing the correlation between agreeableness z-scores and the social – non-social contrast values in the rFPC respectively for the entire, female and male samples.

3.4 Discussion

Agreeableness is one of the five broad dimensions of personality provided by the Five Factor model, mainly associated with social cognitive processes (Allen and DeYoung, 2017; Nettle and Liddle, 2008; Costa and McCrae, 1992a), as confirmed by recent neuroimaging studies evidencing an overlapping neural network between these two domains (Allen et al., 2017; Udochi et al., 2022). In our previous fMRI study (Arbula et al., 2021) we tested whether neural variations associated with agreeableness reflect differences in processing social information, by assessing the degree of similarity between patterns of activation for social and non-social content across individuals. Our results suggested that more agreeable individuals encode social and non-social information more distinctively compared with less agreeable individuals, whose encoding of the two contents is more similar. This neural-behavioral association was present only in the dmPFC, a brain region typically linked with ToM abilities.

In the present study, we addressed two important issues raised by previous research. First, we focused the social information content on a more circumscribed dimension encompassing social interaction detection, to examine whether variations in agreeableness affect more specific lower-level processes. Second, we aimed at identifying at which stage of the neural processing of social interaction agreeableness kicks in, to be able to distinguish the impact on early, perceptual processes from slower, inferential processing. Our hypothesis suggested a potential association between agreeableness and top-down, inferential neural computations involved in extracting information about social interactions over a relatively slow time window (see also Isik et al., 2020).

We expected that agreeableness would modulate the neural processes underlying social interaction detection, which would also be reflected behaviorally as a higher sensitivity to detect social interactions in more agreeable individuals. To this end, we first calculated the sensitivity

(*d'*) measure from the Signal detection theory (Macmillan and Creelman, 1990; Stanislaw and Todorov, 1999) to assess whether agreeableness affects participant's ability to discriminate social from non-social stimuli at the behavioral level, and then we carried out an ERP analysis to establish whether differential neural responses to social and non-social interactions are modulated by this personality trait. We also predicted potential interactions between these effects and sex due to known sex differences in social cognition (Di Tella et al., 2020; Li et al., 2020; Pavlova et al., 2010; Stake and Eisele, 2010; for a review see: Paletta et al., 2022; Proverbio, 2021).

Our findings support our predictions regarding the influence of agreeableness on social interaction detection. We observed that higher agreeableness scores were associated with better discrimination of social from non-social interactions at the behavioral level, although this effect was specific to males. In terms of neural responses, we found that agreeableness modulated ERP signals in an early time window for the entire sample and females (~200 ms), while in males, this modulation was observed across multiple time windows (~150 ms, ~380 ms, ~720 ms) throughout the stimulus presentation. Despite these latency differences, the effect of agreeableness on the neural electrical activity pointed in the same direction: more agreeable participants showed an increased responsivity to social compared with non-social stimuli, whereas less agreeable participants exhibited the opposite pattern.

The source estimation procedures identified the right fronto-parietal area as a potential brain region underlying the association between agreeableness and the social – non-social ERP effects. Specifically, highly agreeable individuals exhibited a stronger activation of this brain region in response to social stimuli compared to non-social stimuli. This result was confirmed for the significant clusters observed in the entire sample and in females (although for the latter, the results did not survive the multiple comparisons correction), and for the second cluster in males.

These findings are in line with previous research demonstrating that social stimuli elicit larger ERP amplitudes compared to socially neutral stimuli (Ding et al., 2017; Kosonogov et al., 2019; Olofsson et al., 2008; Schacht and Vrtička, 2018). Noteworthy, this effect is smaller in populations

with lower social skills, such as in patients with ASD (Benning et al., 2016; Cox et al., 2015; Jeste and Nelson, 2009; Leung et al., 2021; Stavropoulos and Carver, 2014). Several theories have addressed this reduction in the ERP response in patients with social processing impairments, none definitive (for a review: Jeste and Nelson, 2009). For instance, the well-regarded "social motivation theory" suggests that the reduced amplitude may index a shallow processing of the stimulus due to hyposensitivity of the neural system supporting motivation and attention to social stimuli (Chevallier et al., 2012; Dawson et al., 2005; Kikuchi et al., 2009). It is well-established that social stimuli have a processing priority in the human brain compared to non-social stimuli (Fletcher-Watson et al., 2008; Gliga et al., 2009; Isik et al., 2020; Kirchner and Thorpe, 2006; Ro et al., 2001), likely due to their rewarding nature and their relevance for survival (Adolphs, 1999, 2003; Schacht and Vrtička, 2018). Considering this, a larger ERP amplitude observed during early time windows may reflect the allocation of greater cognitive resources to process social stimuli, which can be attributed to a natural predisposition to prioritize and attend to social information due to its rewarding nature. In contrast, in patients with ASD this priority in processing social content appears to be diminished, potentially due to a reduced rewarding value associated with social stimuli (Chevallier et al., 2012). This special population can even show a greater ERP amplitude for neutral stimuli than social stimuli (Benning et al., 2016; Stavropoulos and Carver, 2014). Interestingly, we observed a similar pattern in individuals with lower agreeableness, where higher neural responses were found for stimuli depicting non-social interactions. This suggests that in these cases, the typical human preference for social stimuli is not operating as expected. Since social preference has been argued to be a predictor of attention orienting (Fletcher-Watson et al., 2008; Huguet et al., 1999), the attentional resources are allocated to other stimuli such as objects (Chevallier et al., 2012, 2013).

In this study, participants were drawn from the general population without any specific deficit that could directly affect their social processing abilities. This may partially explain the relatively weak modulation of agreeableness on the sensitivity scores measured behaviorally. At

the same time, agreeableness may still influence motivational and attentional processes related to social stimuli, even without affecting the ability to distinguish between social and non-social interactions. Indeed, increased ERP amplitudes observed during early time windows suggest the engagement of attentional mechanisms during the initial processing of stimuli, associated with the saliency and relevance of the same (Agam and Sekuler, 2007; Bayer et al., 2017; Fellrath et al., 2014; Ibanez et al., 2012; Kosonogov et al., 2019; Wynn et al., 2008). Consequently, we argue that agreeableness can exert a modulatory influence on these processes by enhancing the focus on social stimuli. On the other hand, individuals with low agreeableness may exhibit reduced processing of social stimuli that convey information about social interactions. This could be attributed to a lower processing priority placed on social content, leading to a greater orientation towards non-social stimuli.

This interpretation is consistent with the involvement of the right fronto-parietal regions that we observed to be engaged from both ERP and source estimation procedures. Indeed, these brain regions represent a pivotal hub of the dorsal attention network, which is responsible for selecting relevant sensory information by amplifying relevant input and suppressing irrelevant input (Corbetta and Shulman, 2002; Ptak 2012; Sturm et al., 2016). However, this interpretation should be taken with caution since for the sources' estimations, individual anatomies were not available. Noteworthy, Wilkowski and colleagues (2006) found an association between agreeableness and attention to social stimuli at the behavioral level; our findings broaden this evidence by adding the neural component. However, we would like to emphasize that the data acquisition for this study was conducted during the pandemic. Hence, it is important to acknowledge that this particular context may have influenced the processing of social interactions. Since to our knowledge, this is the first study assessing the impact of agreeableness on social cognition processes by means of EEG, our findings will need to be confirmed in future studies, also by examining the observed association in individuals with ASD and investigating the role that personality might have in this special population.

Notably, our study revealed significant sex differences in relation to agreeableness, both behaviorally and neurally. The effects of agreeableness on neural activity were more extensive and occurred later in males compared to the entire sample and females. Additionally, within the male sample, there was an overlap between agreeableness, performance, and the depth of social stimulus processing, suggesting that more profound social stimulus processing predicted higher agreeableness and better task performance.

Considering the extant literature about ERP neural components associated with late latencies (Codispoti et al., 2007; Ding et al., 2017; Ibanez et al., 2012; Schupp et al., 2000), our findings suggest that agreeableness not only affects the initial salience and relevance processing of social stimuli but may also play a role in more complex attentional processes related to the rewarding aspects of these stimuli. Since this later effect was not evident in females, it is likely that their motivational processes are less influenced by personality traits. This can be understood in the context of existing research on sex differences, which has shown that females tend to be more socially oriented (Baron-Cohen et al., 2005), more sensitive to non-verbal social and emotional cues, and quick learners of such cues compared to males (Hall and Halberstadt, 1981; Paletta et al., 2022; Proverbio, 2021; Stake and Eisele, 2010). Some electrophysiological studies have shown a greater ERP neural response in females associated with social stimuli as opposed to objects (Proverbio et al., 2008). Even when it comes specifically to social interaction, females have been found to anticipate social interactions by predicting the actions of people whereas males need to acquire more information (Pavlova et al., 2010). These inherent sex differences may contribute to the lesser impact of personality on social interaction detection processes in women. Overall, our results suggest that sex is an important factor to consider when trying to understand the neural underpinnings of inter-individual differences in personality traits.

3.5 Conclusion

The current study expands our understanding on personality and sex-related differences in social skills by revealing how neural mechanisms underlying social interaction perception are modulated differently among males and females in relation to the agreeableness trait. Our results show that agreeableness affects brain's activity during the processing of social interactions, but at different latencies in males and females. Since there has been limited investigation into this research topic thus far, future research should aim at explaining whether this modulation of agreeableness occurs for both social and emotional stimuli and whether it might depend on the degree of complexity of the task and of the processes analyzed. Overall, these findings provide valuable contributions to our understanding of individual differences in the social domain, with potential implications for various fields, including clinical interventions targeting social impairments and fostering tailored interventions that consider individual variations.

CHAPTER 4

Study 3: Agreeableness modulates mental state decoding: Electrophysiological evidence

---

Pisanu, E., Arbula, S., & Rumiati, R. I. (under review). *Agreeableness modulates mental state decoding: Electrophysiological evidence*

Abstract

Agreeableness is one of the five personality traits which is associated with Theory of Mind (ToM) abilities. One of the critical processes involved in ToM is the decoding of emotional cues. In the present study, we investigated whether this process is modulated by agreeableness while taking into account task complexity and sex differences that are expected to moderate the relationship between emotional decoding and agreeableness. Two tasks were employed and submitted to 62 participants during EEG recording: the Biological (e)motion task, involving the perception of basic emotional actions through point-light body stimuli, and the Reading the Mind in the Eyes task, requiring the decoding of complex mental states from eye expressions. ERP results showed a significant correlation between agreeableness and the contrast for emotional and non-emotional trials in a late time window only during the RME task. Specifically, higher levels of agreeableness were associated with a deeper neural processing of emotional versus non-emotional trials within the whole and male samples. In contrast, the modulation in females was negligible. The source analysis highlighted that this ERP-agreeableness association engages the ventro-medial prefrontal cortex. Our findings expand previous research on personality and social processing and confirm that sex modulates this relationship.

4.1 Introduction

The agreeableness personality trait of the Five Factor model (FFM) reflects individual differences in pro-social tendencies, cooperation, empathy, and interpersonal harmony (Costa and McCrae, 1992; Goldberg, 1990). Agreeableness has been associated with neural processes involved in emotional regulation, social cognition, reward sensitivity, and Theory of Mind (ToM; DeYoung and Allen, 2017; DeYoung and Blain, 2020; Nettle and Liddle, 2008; Udochi et al., 2022). Individual variations in agreeableness seem to correspond to differences in social-emotional content processing. For instance, during a social animation task, individuals with lower levels of agreeableness exhibited more similar neural patterns of activation for social and non-

social content, whereas more agreeable individuals displayed more distinct neural patterns for these two information in the dorsomedial prefrontal cortex (Chapter 2). This latter brain region has often been reported in ToM (Sabbagh and Bowman, 2018; Saxe and Baron-Cohen, 2006; Schurz et al., 2014, 2021).

Furthermore, to understand whether lower-order processes, corresponding to early sensory and perceptual processing stages, or higher-order processes encompassing slower inferential processes are affected, we designed an EEG study aimed at targeting the detection of social interactions as one of the processes that could underlie agreeableness (Chapter 3). We argued that agreeableness affects social stimulus processing in a relatively early time window (~ 200 ms) but that this modulation depends on the sex of the participants. Specifically, in male participants agreeableness had an ample effect on social stimulus processing that, from the beginning of the trial, extends beyond the end of the stimulus presentation that lasted 500 ms, and was also predictive of their performance. Over and above the sex effect, individuals with lower levels of agreeableness displayed reduced responsivity to social stimuli compared with non-social stimuli, while individuals with higher levels of agreeableness showed the opposite pattern.

What remains to be addressed is whether the modulation of agreeableness occurs for both social and emotional stimuli processing, and whether it is affected by the complexity of the task and of the neural processes analyzed.

To address these questions, in the present EEG study, the same participants as in Pisanu et al. (Chapter 3) performed two tasks of varying levels of complexity, specifically aimed at decoding emotional cues. The former is the Biological (e)motion (BM) task (Atkinson et al., 2012) that requires emotional perception processes to discriminate basic emotional actions (i.e., happiness, anger, and sadness) expressed in the point-light bodily movement stimuli. It is well established that bodily movements serve as a rich and immediate source of social information (Beall et al., 2008; De Gelder, 2006) that can be rapidly and effortlessly extracted even from sparse visual information, as effectively demonstrated in point-light bodily stimuli (Blake and Shiffrar, 2007;

Puce et al., 2015). Extracting social information from bodily movements of others has been proposed to rely on the sensorimotor system, suggesting the possible involvement of a simulation process. This system is the same that is involved in action execution (Del Vecchio et al., 2020; Gallese and Goldman, 1998; Jeannerod, 2001, 2006; Rizzolatti and Craighero, 2004). For instance, EEG recordings consistently reveal a decrease in alpha-band (8-12 Hz) oscillations over the sensorimotor cortex. This phenomenon, known as μ-alpha suppression, is argued to reflect the mental simulation of action and has been proposed to facilitate the internal representation of others' emotional states (see Hobson and Bishop, 2017; Fox et al., 2016 for a review). Using point-light body stimuli, Siqi-Liu et al. (2018) observed a more pronounced μ-suppression in response to emotional actions compared to neutral actions, indicating a heightened neural response to actions involving an emotional component from the middle to the end of the stimulus presentation. Furthermore, these authors found that individuals characterized by lower levels of autistic trait demonstrated a more pronounced degree of μ-suppression in comparison to those with higher levels of the same trait. The latter task we employed is the Reading the Mind in the Eyes (RME; Baron-Cohen et al., 2001) that, unlike the BM task, engages higher mental state recognition (i.e., compassionate, playful, confused) only through the expression of the eyes. It is a well-established task of mental state decoding, representing the first stage of ToM that refers to the identification from visible cues of the type of mental state without encompassing the subsequent stage of inferring the content associated with that mental state, such as the compassion for a specific event like the loss of one's dear friend (Baron-Cohen et al., 2001). As such it is thought to evaluate the subjects' ability to "put themselves into the mind" of the other person by attuning to their emotions.

In this study, using these two tasks enabled us to test the interplay between early-stage emotional perception and higher-level mental state inference. In doing so, we could then provide a comprehensive understanding of the relationship between personality and the neural mechanisms underlying emotional cue decoding, spanning from the perception of basic emotional actions to the recognition of complex mental states. We hypothesized that variations in the levels of

agreeableness would correlate with differences in the ability to detect emotions, along with modulations in event-related potential (ERP) responses in the emotional condition compared with the non-emotional condition. We also hypothesized a modulation of event-related spectral perturbations (ERSP) responses expressed as a stronger μ-suppression for more agreeable participants.

We also predicted that this modulation would depend on the type of task, with a more extensive and later influence of agreeableness for the RME task than the BM task, as the former involves more complex neural processes. Furthermore, as in our previous EEG study where we found a sex effect in agreeableness (Chapter 3), we expected that the emotional processes at study here should be modulated by sex.

Lastly, we measured the impact of agreeableness on the spontaneous processing of gaze stimuli. We reasoned that individuals with higher agreeableness may exhibit heightened sensitivity to eye stimuli even during passive observation, without engaging in any specific actions or decision-making processes. Thus, we predicted that individuals with high agreeableness scores would demonstrate similar neural responses when recognizing emotional states and when simply observing gaze stimuli. Conversely, we expected individuals with lower agreeableness scores to exhibit a greater disparity in their neural responses.

4.2 Methods

4.2.1 Participants recruitment and personality assessment

Sixty-two young adults (27 males, mean age 24.03, SD = 4.34, range = 18-35 years) were selected from a larger sample (N = 462) previously screened for agreeableness to increase in-sample variance by means of the Italian adaptation of the Big Five Inventory: 20 participants scored above and 20 scored below 1.4 SD from the Italian population mean (Ubbiali et al., 2013). As mentioned in the previous chapters, since personality traits are continuous variables, in order to ensure that our sample accurately captures measurements at either tail of the distribution,

deliberate efforts were made to select individuals with diverse scores along the z-score continuum, encompassing both negative and positive values. The questionnaire consists of 44 items rated on a five-point Likert scale aimed at evaluating the five personality traits from the Five Factor model (Extraversion, Neuroticism, Openness, Conscientiousness, and Agreeableness). Participants were recruited through social media and email. Exclusion criteria included a history of neurological or psychiatric disorders. All participants provided written informed consent and received 20 euros upon completion of the study. The study and informed consent procedures were approved by the SISSA's ethics committee and adhered to the guidelines of the Declaration of Helsinki.

4.2.2 Biological (e)motion task

Participants were presented with an Italian cue word followed by a target video. The cue could either represent one of the five basic emotions (happiness, sadness, fear, disgust, anger) or a common everyday action (knocking, hopping, digging, walking on spot, star jumping, touching toes). Targets were represented by 11 whole-body actions depicted in point-light displays videos, 5 representing an emotional action and 6 representing emotionally neutral actions. After each target video, participants were asked to indicate whether according to them the cue was congruent with the target by pressing a key button (yes/no). The stimuli were provided by Atkinson et al. (2012).

Participants were seated 60 cm from the screen in a dark and quiet room. The stimuli were presented at 9 × 5 degrees of visual angle on a black background, in the center of the screen by using PsychoPy3 (v2020.1.1). The task started with 4 practice trials containing emotional and neutral actions (these responses were not included in the analyses). Participants completed 100 trials (50 per condition) interspersed with a self-paced pause after 50 trials. Stimuli were arranged in a pseudo-random order to avoid more than two repetitions of the same condition. Each trial began with the presentation of the cue (500 ms), followed by a fixation cross (1000 ms), and the target video (2000 ms); after the offset of the target, participants were presented with a blank screen (500 ms) after which they had 2000 ms to give their evaluation via button press (Figure

4.1). To respond, participants pressed either the “f” or the “j” key on a QUERTY keyboard with the index finger of both hands; additionally, response keys were counterbalanced between subjects.

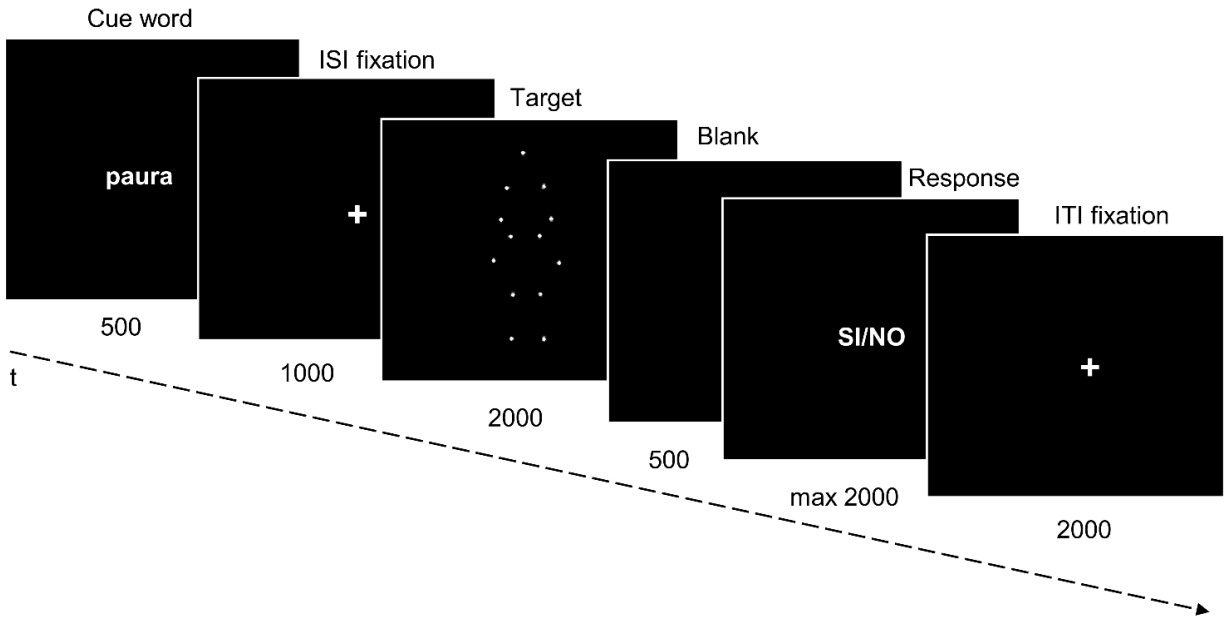


Figure 4.1. Schematic representation of the ‘biological (e)motion task’. Stimuli succession is represented left to right, with the respective screen presentation time expressed in milliseconds below each stimulus.

4.2.3 Reading the mind in the eyes task

This task is an adaptation of “The Reading the Mind in the Eyes” (Baron-Cohen et al., 2001) for use in an ERP paradigm. Participants were presented with an Italian cue word followed by a target image. The cue could either represent one of the complex emotional states from the original paper and pencil task (e.g., playful), an age band (e.g., 29-39 years) or the word “observe”. Targets were represented by 36 black and white photos of the eyes of male or female individuals from the original task. In both the emotion and age conditions, after each target, participants were asked to indicate whether according to them the cue was congruent with the target by pressing a key button (yes/no). In the “observe” condition, participants were instructed to simply watch the image without taking action.

Previous neuroimaging studies adopted the sex decoding as the control task (Sabbagh et al., 2004) nevertheless, this option entails a significant constraint, namely that the two conditions exhibit a

pronounced imbalance, with the control condition (sex decoding) being excessively easy in comparison to the experimental condition (emotional state decoding) affecting the interpretation of the results especially in terms of ERP components involved (Sabbagh et al., 2004). Therefore, we chose a more difficult condition, namely the age decoding, first piloted behaviorally in a sample of 90 subjects. Furthermore, we also added the "observe" condition to analyse possible individual differences emerging even simply by watching gaze stimuli without acting.

Participants were seated 60 cm from the screen in a dark and quiet room. The stimuli were presented at 9 × 5 degrees of visual angle on a grey background, in the center of the screen by using PsychoPy3 (v2020.1.1). The task started with 4 practice trials containing emotional states and age bands (these responses were not included in the analyses). Participants completed 150 trials (50 per condition: emotion, age, observe) interspersed with a self-paced pause every 50 trials. Stimuli were arranged in a pseudo-random order to avoid more than two repetitions of the same condition. Each trial began with the presentation of the cue (500 ms), followed by a fixation cross (1000 ms), and the target image (500 ms); after the offset of the target, participants were presented with a blank screen (2000 ms) after which they had 2000 ms to give their evaluation via button press (Figure 4.2). The response procedure is identical to that employed in the BM task. The evaluation was absent during the "observe" condition.

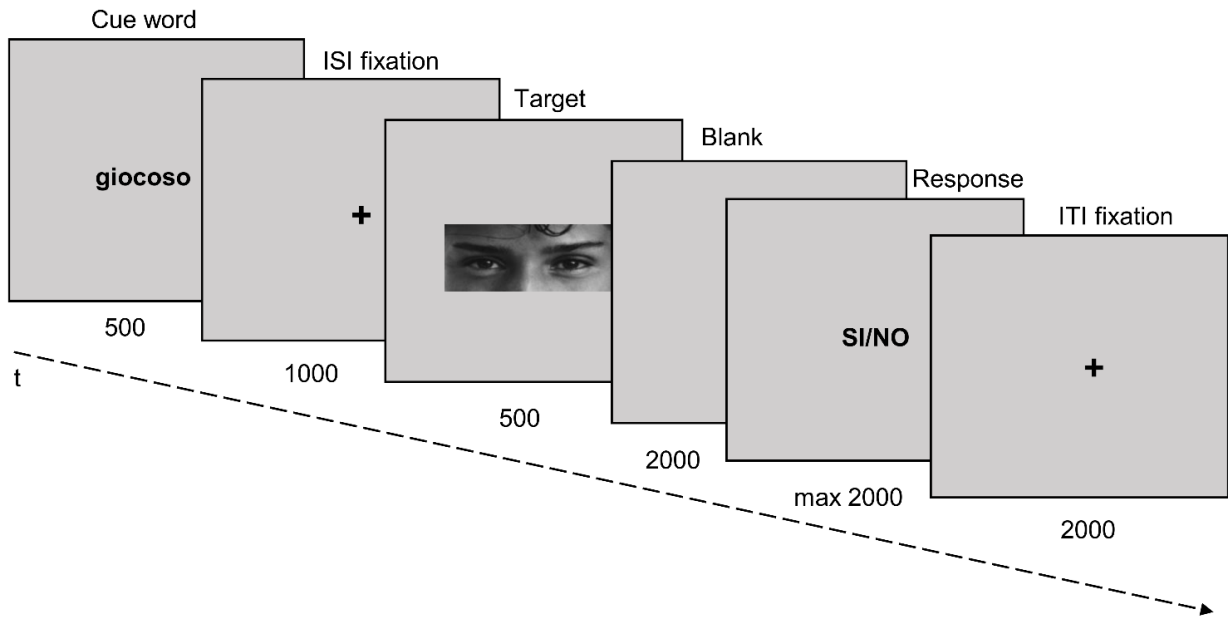


Figure 4.2. Schematic representation of the 'reading the mind in the eyes task'.

4.2.4 EEG recording and preprocessing

EEG data were recorded using a 64 Ag/AgCl active electrode setup based on the International 10-20 system (Klem et al.,1999), and signal amplification was achieved with a BioSemi Active-Two amplifier system. The ActiView acquisition software (ActiView 707, Biosemi, Amsterdam, The Netherlands) was adopted for visualization and data storage. Electrode offsets were maintained within ±50 mV, and the signal was sampled at 1024 Hz with 24-bit digitization resolution. Eye movements and blinks were monitored using horizontal and vertical electrooculogram (EOG) recorded from additional electrodes placed below and above the left eye.

Offline data preprocessing was performed using EEGLAB 14.1.1b (Delorme and Makeig, 2004, https://sccn.ucsd.edu/eeglab) in Matlab R2021a (https://uk.mathworks.com/products/matlab.html). The continuous EEG recordings were resampled at 250 Hz and band-pass filtered (0.25-45 Hz) using zero-phase Hamming windowed sinc FIR filters (transition bandwidth: 0.5-10 Hz, passband edges: 0.5-40 Hz). Poorly correlated channels were identified and removed using the EEGLAB *clean_rawdata* plugin (autocorrelation parameter = 0.8, resulting in an average exclusion of 1.9 channels, SD = 2.1). Epochs ranging from 1000 ms before to 2500 ms after target onset were extracted. Independent component analysis (ICA; Makeig et al., 1996) was employed to remove components with more than 60% probability of being eye, motor, channel noise, heart, or line noise using ICLabel (Pion-Tonachini et al., 2019).

Removed channels were replaced using spherical spline interpolation (Perrin et al., 1989), and data were re-referenced to a common average reference. Epochs were baseline-corrected to a 200 ms pre-stimulus baseline and inspected for artifacts based on different statistics (amplitude, linear trend, joint probability, and kurtosis) (Delorme et al., 2007). Overall, 8% of the BM trials and 10% of the RME ones were rejected for violating artifact criteria (with no asymmetry across

conditions). The average number of epochs used for the analysis was: 46 in both conditions for the BM, ranging from 37 to 50 epochs; 45 in the emotion and age conditions and 44 in the observe condition for the RME, ranging from 33 to 49 epochs. For the latter task, three participants (one male and two females) were excluded from the EEG analysis due to technical problems during the recording.

The ERSP were extracted from the whole epoch (-1000 - 2500 ms) by means of a Morlet wavelet convolution. We estimated 42 log-spaced frequencies, ranging between 4 and 45 Hz, using 4 cycles at lowest frequency to 22.5 at highest, with a temporal resolution of approximately 12 ms. The baseline correction was performed relative to the average power in the time interval between 900 and 400 ms prior to stimulus onset.

4.2.5 Behavioral data analysis

We collected accuracy and reaction times (RTs) data. RTs were filtered for errors, anticipated responses (< 150 ms) and outliers above 3 standard deviations from the subjects mean for each condition. We used R v.4.1.1 (R Core Team, 2020) to perform linear mixed models selecting accuracy and RTs separately as dependent variables, type of condition (i.e., emotional, non-emotional) and agreeableness as two fixed effects, and subject intercepts as random effects. To explore potential sex differences in the relationship between agreeableness and performance, two other models were developed (one for accuracy and one for RTs) by including sex as an additional fixed effect. To ensure that any observed effects were specific to agreeableness and not influenced by other personality traits, the other four traits were then included as additional predictors in the models. Assumptions for the models were checked through the visual inspection of residual plots (linearity, homoscedasticity, and normality).

4.2.6 ERP analysis

ERP analyses were carried out separately for the BM and RME task by using Matlab R2021a along with EEGLAB 2020_0 and LIMO EEG plugin (v 3.3, https://github.com/LIMO-

EEG-Toolbox/limo_tools). A massive univariate approach was adopted by performing two levels of analysis for both the two tasks (Pernet et al., 2011).

First-level GLM analysis

At the first level, parameters of a General Linear Model were estimated by using a standard ordinary least squares solution for each subject, at each time point and each electrode independently, and separately for all the experimental conditions. To assess the relationship between agreeableness and the differential activity elicited by the experimental conditions, the corresponding ERP amplitudes were subtracted to create a contrast (emotional action – neutral action for the BM task; emotion – age, and emotion – observe for the RME). The selected time windows ranged from 1000 to 2500 ms for the BM task, and from 0 to 1000 ms for the RME task, time locked to stimulus-target onset. For the former task, in which the stimulus had a longer duration being a video, in line with the findings of Siqi-Liu et al. (2018), we started the analysis from 1000 ms discarding the time points related to early neural processes which were not our target, and instead keeping the time points following the end of the stimulus as for the RME task. Erroneous trials were retained in the model to preserve higher inter-individual variability, as we predicted that accuracy would vary based on agreeableness scores.

Second level GLM analysis

At the second level, for both tasks, robust paired t-tests between emotional and non-emotional conditions were carried out across all subjects, all time points, and all electrodes in order to compare participants' ERP amplitudes across different conditions. Results are reported corrected for multiple testing using the maximum statistics thresholding $p = .01$ (Maris and Oostenveld, 2007; Pernet et al., 2015) that ensures robust control over the family-wise error rate, providing both reliable significance and precise spatio-temporal localization of the effect observed (Groppe et al., 2011a, 2011b). Afterwards, robust skipped Pearson correlations were performed (Wilcox et al., 2018) by means of the Robust Correlation toolbox (https://github.com/CPernet/robustcorrtool; Pernet et al., 2013) between the agreeableness scores and the ERP contrast values (emotional –

non-emotional) of each subject in the time points and electrodes that resulted significant from the abovementioned paired t-tests. The statistical significance was tested through the percentile bootstrap 95% confidence intervals and results are reported with the FDR correction for multiple comparisons. The same correlation analyses were repeated separately for the samples of males and females to unravel potential sex differences. Significant correlation effects that lasted <20 ms were not reported.

By running the robust correlation based on the results from the paired t-test, we were able to bypass the preselection of specific time intervals and electrodes, while still performing a targeted investigation which considers the most relevant and informative data for the relationship between agreeableness and neural responses. This approach minimized the likelihood of false positive findings and enhanced the interpretability of the results. Additionally, the use of robust correlations reduced the impact of outliers by considering the overall structure of the data and excluding bivariate outliers.

4.2.7 ERSP analysis

Since our hypothesis specifically targeted a possible association between agreeableness and μ-alpha suppression, ERSP analysis was performed only for the BM task by using Matlab R2021a along with EEGLAB 2021_0 and LIMO EEG plugin (v 3.3). As for the ERPs, a massive univariate approach was used by running two levels of analysis (Pernet et al., 2011).

First-level GLM analysis

At the first level, parameters of a General Linear Model were estimated for each subject, at each time point, each electrode and each frequency in the range between 8-20 Hz independently, and separately for the experimental conditions, namely emotional action and neutral action. The time window ranged from 0 to 2500 ms time locked to stimulus onset, excluding edge artifacts induced by the wavelet transform. Parameters were estimated by using a standard ordinary least squares solution.

Second level GLM analysis

At the second level, a robust paired t-test between the emotional and emotionally-neutral actions conditions was carried out across all subjects, all time points, all electrodes, and all frequencies. Results are reported corrected for multiple testing using spatial-temporal clustering $p = .01$ (Maris and Oostenveld, 2007; Pernet et al., 2015).

Afterwards, we performed a robust skipped Pearson correlation between the ERSP contrast values (emotional action – emotional-neutrally action) of each subject and their agreeableness scores in the time points, electrodes and averaged range of frequencies that resulted significant in the abovementioned paired t-test. Frequencies were averaged to reduce the dimensionality of the data and consequently the number of multiple tests. Correlations were performed within the entire, female and male sample. Results are reported corrected for multiple testing using the FDR.

4.2.8 Source analysis

The estimation of cortical source activations was performed using the Brainstorm software, which adopts a distributed dipoles model for fitting (Tadel et al., 2011). To determine the locations of the recorded electrical activity on the scalp, the dynamic statistical parametric mapping (dSPM) method was employed, which incorporates minimum-norm inverse maps (Dale et al., 2000). For each participant, single subject noise covariance matrices and individual noise standard deviations at each location were calculated and estimated based on the pre-stimulus baseline intervals (-200 to 0 ms) from single trials (Hansen et al., 2010).

To build the head model, the boundary element method (BEM) was utilized with default parameters in Brainstorm, using the ICBM152 anatomy as individual anatomies were unavailable. The BEM model provides three realistic layers and representative anatomical information (Gramfort et al., 2010; Stenroos et al., 2014). The source estimation was carried out adopting the option of constrained dipole orientations, where one dipole oriented perpendicular to the cortical surface was modeled for each vertex (Tadel et al., 2011).

To estimate active sources, single-trial EEG data for each participant were averaged to

obtain the subject average. Sources were then estimated separately for the experimental conditions (emotional and non-emotional) and subtracted to create a contrast at the subject level for both tasks. Z-score normalization was applied to each source, utilizing the baseline period (-102 to -4 ms), to mitigate the influence of inter-individual fluctuations in neural current intensity resulting from non-relevant anatomical or physiological differences, and rectified them in absolute values. Afterwards, we computed paired t-tests between emotional and non-emotional conditions across subjects in the time windows-averaged of the ERP effects. Accordingly, for the RME task, two paired t-test were performed: emotion – age and emotion – observe. We then selected the brain regions emerging from these statistics and used them as regions of interest (ROIs) to perform robust skipped Pearson correlations with agreeableness scores within the entire, female and male sample. The ROIs were defined manually by visual inspection to be as accurate as possible since no atlas provided could encompass them properly. Since we were interested in investigating the cortical sources underlying the ERP or ERSP effects associated with agreeableness within both tasks, statistics on sources were computed only when these effects were found.

4.3 Results

4.3.1 Behavioral results

RME. The linear mixed model testing whether agreeableness affects participants' performance in recognizing complex emotional states compared to age bands, captured by the accuracy measure, showed a trend suggesting that higher agreeableness scores predicted better ability to recognize emotions *vs*. age, although the *p* value did not reach the significance ($F(1, 60) = 3.23$, $p = .077$, $\eta^2 = .05$; Figure 4.3). The current trend persists also when controlling for the other personality traits ($F(1, 50) = 3.14$, $p = .08$, $\eta^2 = .05$). When including sex as a second fixed effect in the model, the interaction with agreeableness remained non-significant ($F(1, 58) = 1.61$, $p = .21$, , $\eta^2 = .03$). No significant results were observed in the linear mixed models considering reaction time measure as a dependent variable, either with or without the consideration of sex

(respectively: $F(1, 60) = 2.06$, $p = .16$, , $\eta^2 = .04$; $F(1, 58) = 3.04$, $p = .09$, , $\eta^2 = .04$). No significant or nearly significant associations with any of the dependent behavioral variables investigated were observed for any of the other personality traits (accuracy: all p's > .2; RTs : all p's > .7).

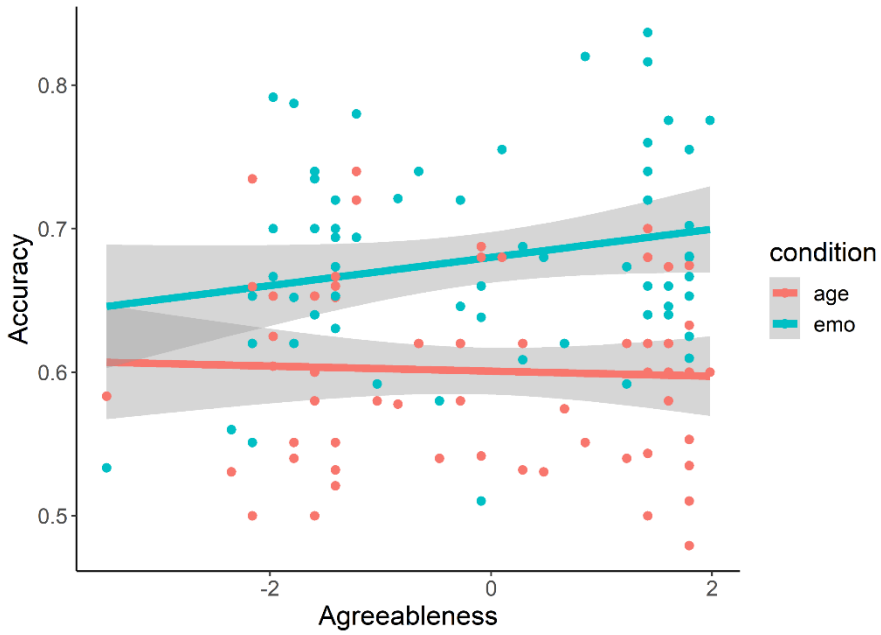


Figure 4.3. The scatterplot shows the relationship between agreeableness (represented in z scores on the x axis), and accuracy (on the y axis) in the RME task; the blue line represents emotion condition, the red line age's one; each spot is a subject. The grey shadow around the trendline represents the standard error (statistics and *p* values are reported in the text).

BM. No significant results emerged from the linear mixed-models carried out on accuracy (agreeableness model: $F(1, 60) = .69$, $p = .41$, $\eta^2 = 01$; agreeableness × sex model : $F(1, 58) = .77$, $p = .39$, $\eta^2 = .01$) and reaction times (agreeableness model: $F(1, 60) = .98$, $p = .33$, $\eta^2 = .02$; agreeableness × sex model: $F(1, 58) = .03$, $p = .87$, $\eta^2 = .0004$). No significant or nearly significant associations with any of the dependent behavioral variables investigated were observed for any of the other personality traits (accuracy: all p's > .3; RTs : all p's > .2).

Visual inspection of residual plots did not reveal any obvious deviations from linearity, homoscedasticity or normality in both tasks.

4.3.2 ERP results

RME. A significant difference emerged between the ERP neural response associated with

the emotion condition and the age condition from the paired t-test starting at 504 ms and ending at 1000 ms (max *t* value 10.73 at 746 ms on channel AFz, corrected *p* value .001; Figure 4.4A) within the entire sample. This difference was led by the emotional condition eliciting a higher neural response than the age condition in the abovementioned time window expressed by a larger late positivity over fronto-central electrodes and, as a consequence of the dipole effect, a larger negativity over parieto-occipital electrodes (Figure 4.4B).

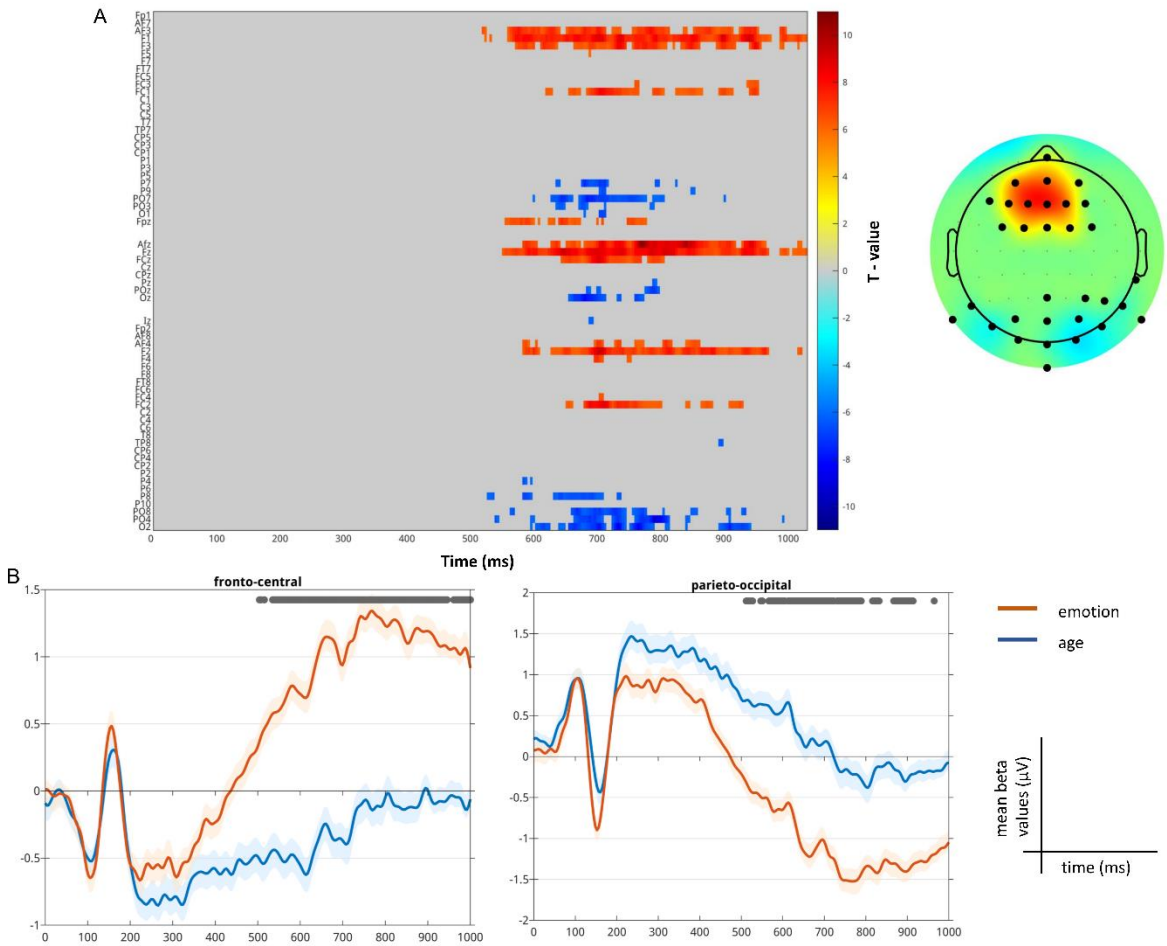
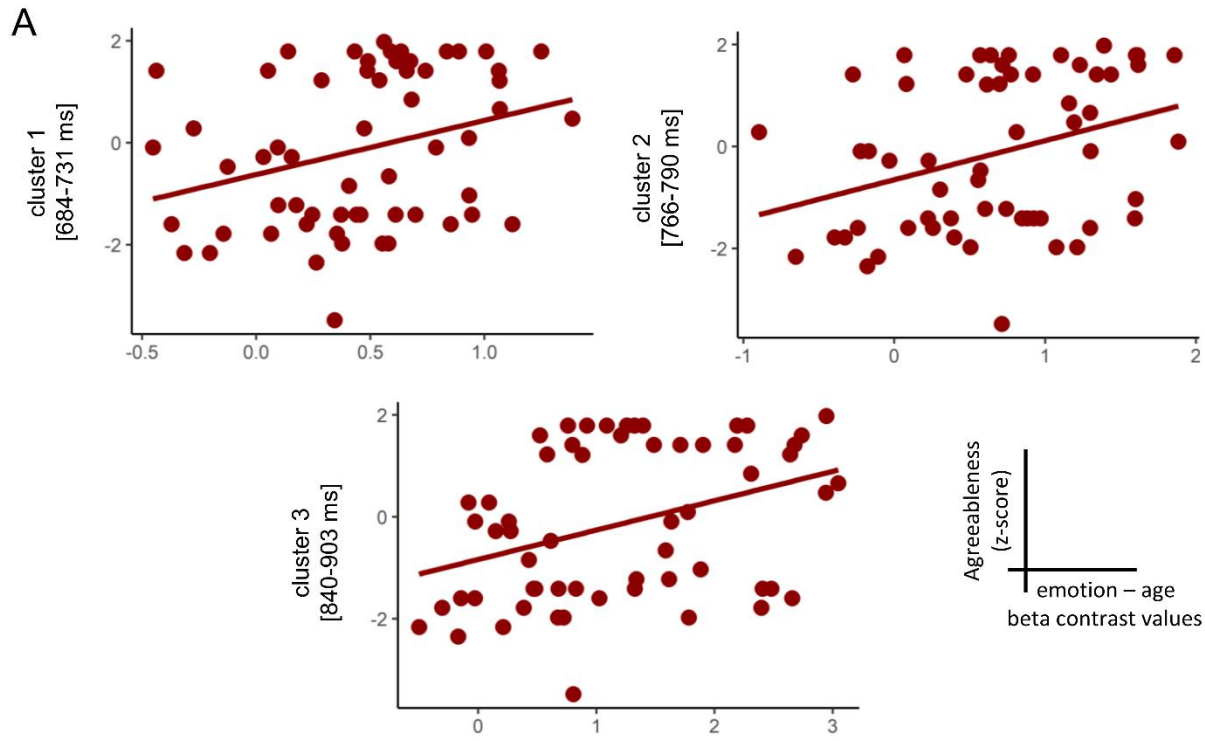


Figure 4.4. ERP results from the paired t-test (emotion – age), RME task. (A) Raster plot depicting significant channels and time-points for the paired t-test colored based on their *t* values. The electrodes on the y-axis are arranged topographically: left and right sides of the head are organized respectively on the top and the bottom part of the axis, midline electrodes are displayed in the middle. Thus, overall the y-axis represents the scalp anterior to posterior. On the right the topographical representation of the *t* values averaged across the significant time-window (504-1000 ms). (B) Trace plots showing the mean beta values for the emotion (in red) and age conditions (in blue) averaged over the significant electrodes. The grey line at the top indicates the significant time window.

After performing the skipped Pearson correlation between agreeableness and the ERP

emotion – age contrast values within the entire sample, we found 3 significant temporal clusters. Cluster 1 starts at 684 ms and ends at 731 ms on channels F1, Fz, F2, FC2, FCz (median r = .35, range = .28 to .46; median lower bound of the $CI_{95\%}$ = .14, range = .11 to 21; median upper bound of the $CI_{95\%}$ = .54, range = .51 to .60; all corrected p value < .05); cluster 2 starts at 766 and ends at 790 ms on channels F1 and Fz (median r = .47, range = .42 to .48; median lower bound of the $CI_{95\%}$ = .12, range = .09 to .14; median upper bound of the $CI_{95\%}$ = .56, range = .54 to .58; all corrected p value < .05); cluster 3 starts at 840 ms and ends at 903 ms on channels F1, Fz, F2 (median r = .40, range = .24 to .54; median lower bound of the $CI_{95\%}$ = .16, range = .12 to .23; median upper bound of the $CI_{95\%}$ = .58, range = .53 to .63; all corrected p value < .05). These results showed that agreeableness modulates the differential neural response such that subjects with higher agreeableness scores had more positive contrast values over the fronto-central electrodes (Figure 4.5A). To inspect the direction of these contrasts across different levels of agreeableness, we divided the entire sample in two groups (high A: high agreeableness, low A: low agreeableness) by means of the median-split of their agreeableness z-scores, and plotted their averaged ERPs (i.e., first level beta values) for the emotion and age condition (Figure 4.5B). The plot reveals that the high A group exhibited a larger positive amplitude for emotion *vs.* age trials over the fronto-central electrodes and a larger difference between the two conditions with respect to the low A group. This pattern was the same for all the three temporal clusters, although the modulation of agreeableness was more extensive within the third and latest cluster where we observed the highest differentiation between the two conditions in the high A group with respect to the low A group.


A
cluster 1
[684-731 ms]
cluster 2
[766-790 ms]
cluster 3
[840-903 ms]
Agreeableness
(z-score)
emotion – age
beta contrast values

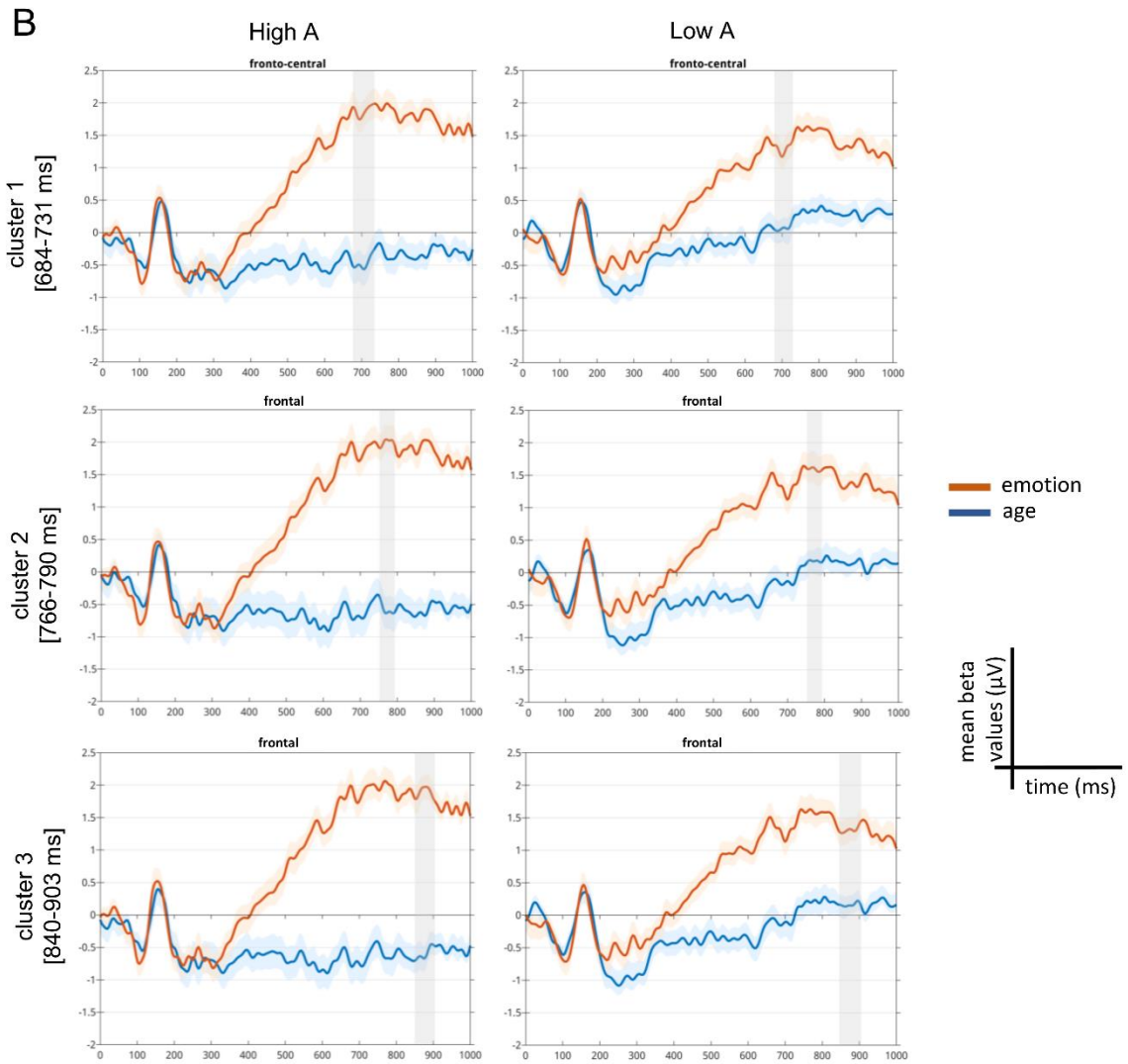

B
High A
Low A
fronto-central
frontal
cluster 1
[684-731 ms]
cluster 2
[766-790 ms]
cluster 3
[840-903 ms]
emotion
age
mean beta
values (μV)
time (ms)

Figure 4.5. ERP-Agreeableness skipped Pearson correlation results for the entire sample. (A) Scatterplots representing the correlation between the beta contrast values averaged across the significant time window and the agreeableness trait, separately for the three temporal clusters. (B) Trace plots showing the mean beta values for the emotion (in red) and age conditions (in blue) averaged over the significant electrodes from each of the three clusters, separately for the high A and low A group.

As regards the female sample, no significant correlation effects that lasted >= 20 ms were observed.

Within the male sample, the robust correlation unveiled one significant temporal cluster starting at 844 ms and ending at 891 ms over channels AF3, F1, Fz (median r = .51, range = .43 to .57; median lower bound of the $CI_{95\%}$ = .18, range = .14 to .26; median upper bound of the $CI_{95\%}$ = .65, range = .63 to .76; all corrected p value < .05). Noteworthy, this cluster matched the third cluster found within the entire sample. Particularly, males with higher levels of agreeableness showed more positive contrast values over frontal electrodes than males with low agreeableness (Figure 4.6A). This result was driven by a greater difference between the ERPs associated with the two conditions, particularly guided by a larger positive amplitude for emotion *vs.* age trials in more agreeable males compared to less agreeable ones (Figure 4.6B).

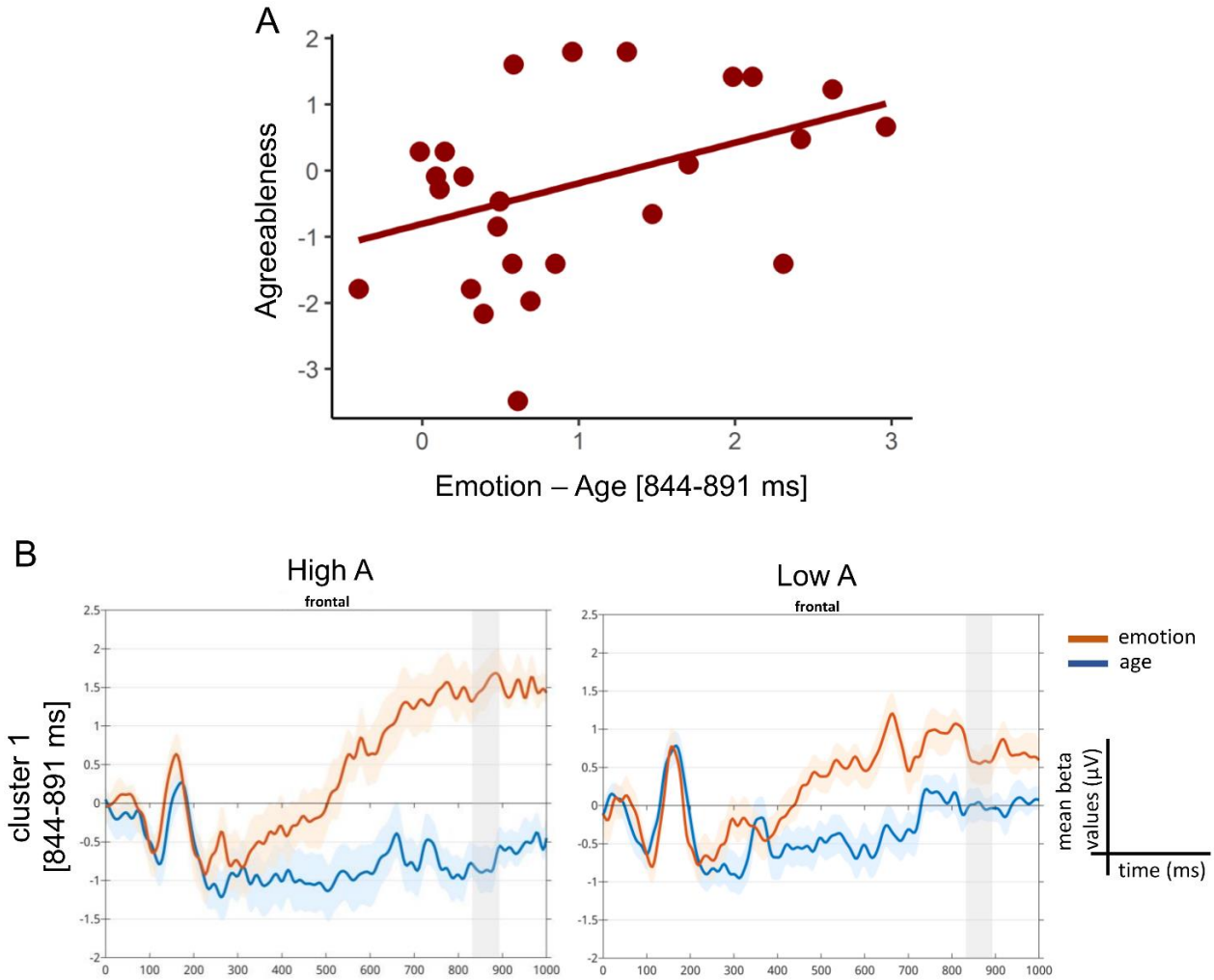


Figure 4.6. ERP-Agreeableness skipped Pearson correlation results for the male sample. (A) Scatterplot representing the correlation between the beta contrast values averaged across the significant time window and the agreeableness trait for the significant temporal cluster. (B) Trace plots showing the mean beta values for the emotion (in red) and age conditions (in blue) averaged over the significant electrodes, separately for the high A and low A group.

From the paired t-test comparing the emotion and observe conditions, a significant neural difference emerged within the entire sample starting at 367 ms and ending at 1000 ms (max *t* value 11.37 at 664 ms on channel Fz, corrected *p* value .001; Figure 4.7A). This difference was explained by the emotional condition eliciting a higher neural response (larger late positivity) than the observe condition in the abovementioned time window over fronto-central electrodes, and, as a consequence of the dipole effect, a larger negativity over the parieto-occipital electrodes (Figure 4.7B).

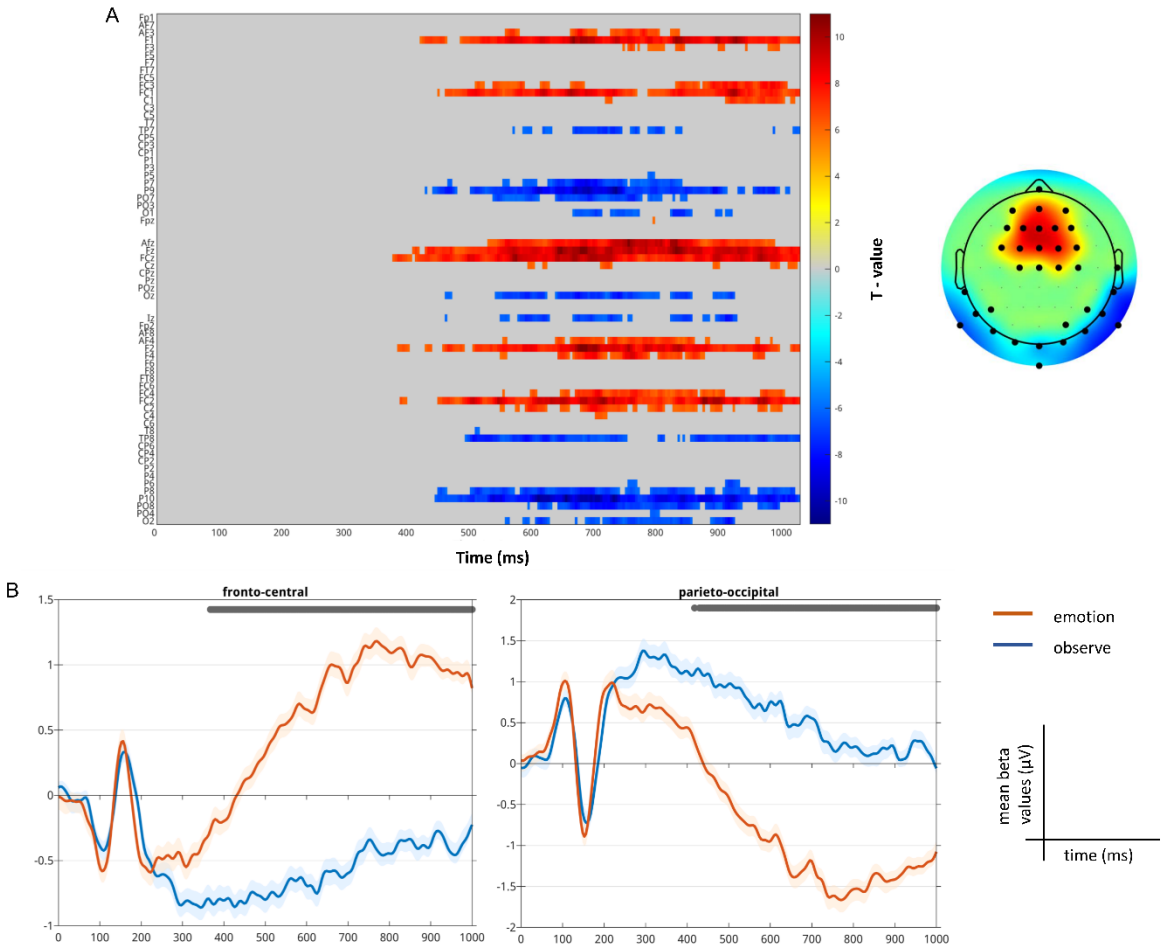


Figure 4.7. ERP results from the paired t-test (emotion – observe), RME task. (A) Raster plot depicting significant channels and time-points for the paired t-test colored based on their *t* values. On the right the topographical representation of the *t* values averaged across the significant time-window (367-1000 ms). (B) Trace plots showing the mean beta values for the emotion (in red) and observe conditions (in blue) averaged over the significant electrodes.

The skipped Pearson correlation between agreeableness and the ERP emotion – observe contrast did not reveal significant results when correcting for multiple comparisons within the entire sample, and either in females or males. However, we report the uncorrected results of all subjects since they align with our hypothesis that agreeableness might affect the processing of emotional stimuli also when subjects are asked to simply observe without taking action. Indeed, subjects with higher agreeableness scores had contrast values approaching to zero while those with lower agreeableness showed more positive contrast values over the fronto-central electrodes and, as a consequence of the dipole effect, more negative contrast values over the parieto-occipital

electrodes, mostly in a time window between 809 and 833 ms (Figure 4.8A). These results were led by a reduced difference between the emotion and observe condition within the high A group with respect to the low A one in which the distinction between the two conditions was higher (both in fronto-central and parieto-occipital electrodes. Figure 4.8B).

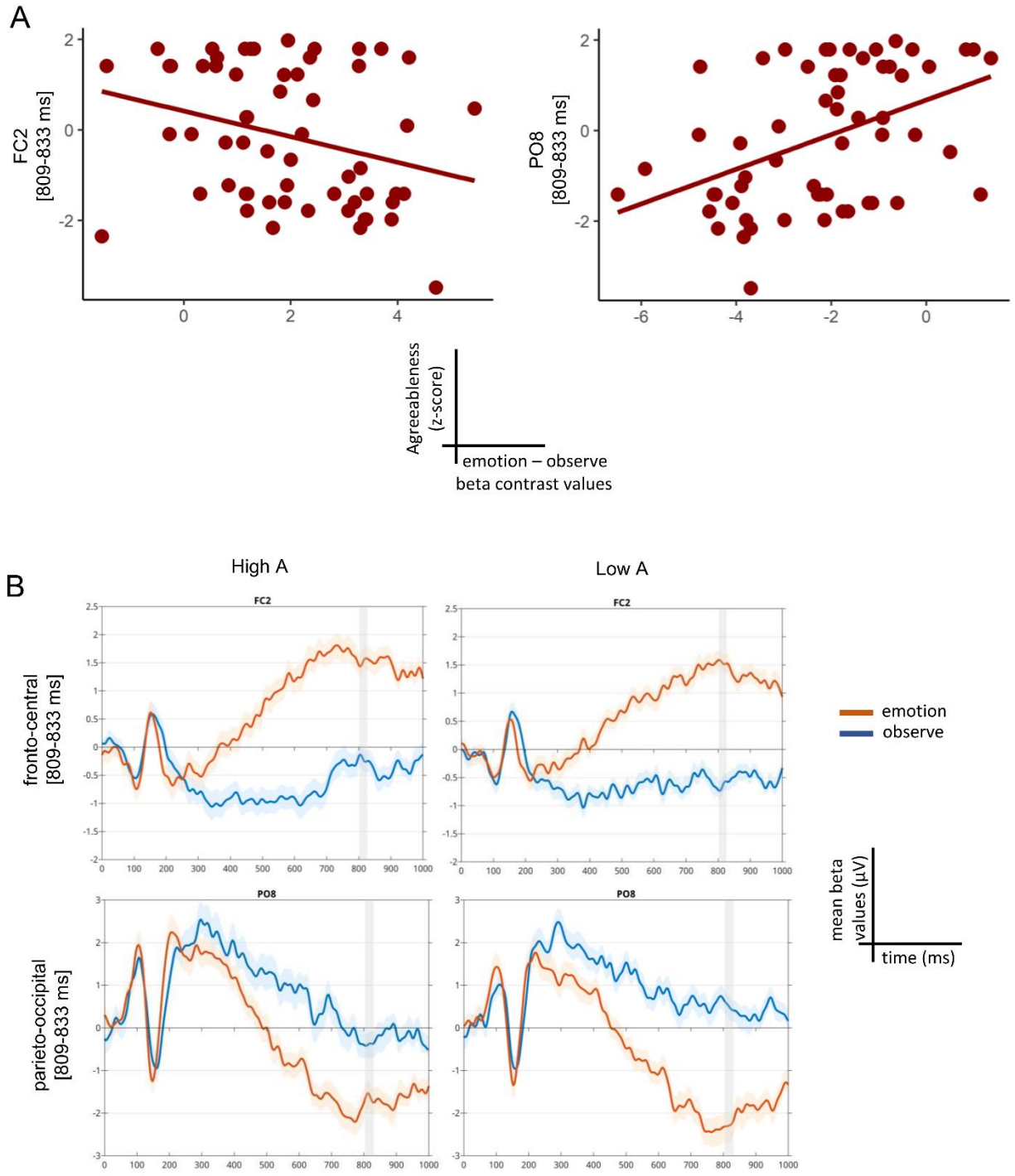


Figure 4.8. ERP-Agreeableness skipped Pearson correlation results for the entire sample. (A) Scatterplots representing the correlation between the beta contrast values averaged across the significant time window (809-833

ms) and the agreeableness trait, separately for the most significant electrodes (FC2: max r = -.46, uncor p = .02 at 813 ms; PO8: max r = .54, uncor p = .001 at 817 ms). (B) Trace plots showing the mean beta values for the emotion (in red) and observe conditions (in blue) from each of the two electrodes, separately for the high A and low A group.

BM. Two significant temporal clusters emerged from the paired t-test between the ERP responses associated with emotional action and neutral action (cluster 1 starts at 1117 ms and ends at 1731 ms, max t value 7.45 at 1465 ms on channel F1, corrected p value <.0001; cluster 2 starts at 1813 ms and ends at 2371 ms, max t value 7.2 at 2175 ms on channel C3, corrected p value <.0001; Figure 4.9A). This difference was driven by the emotional condition eliciting a higher neural response (larger late positivity) than the neutral condition in the abovementioned time window over fronto-central electrodes (Figure 4.9B).

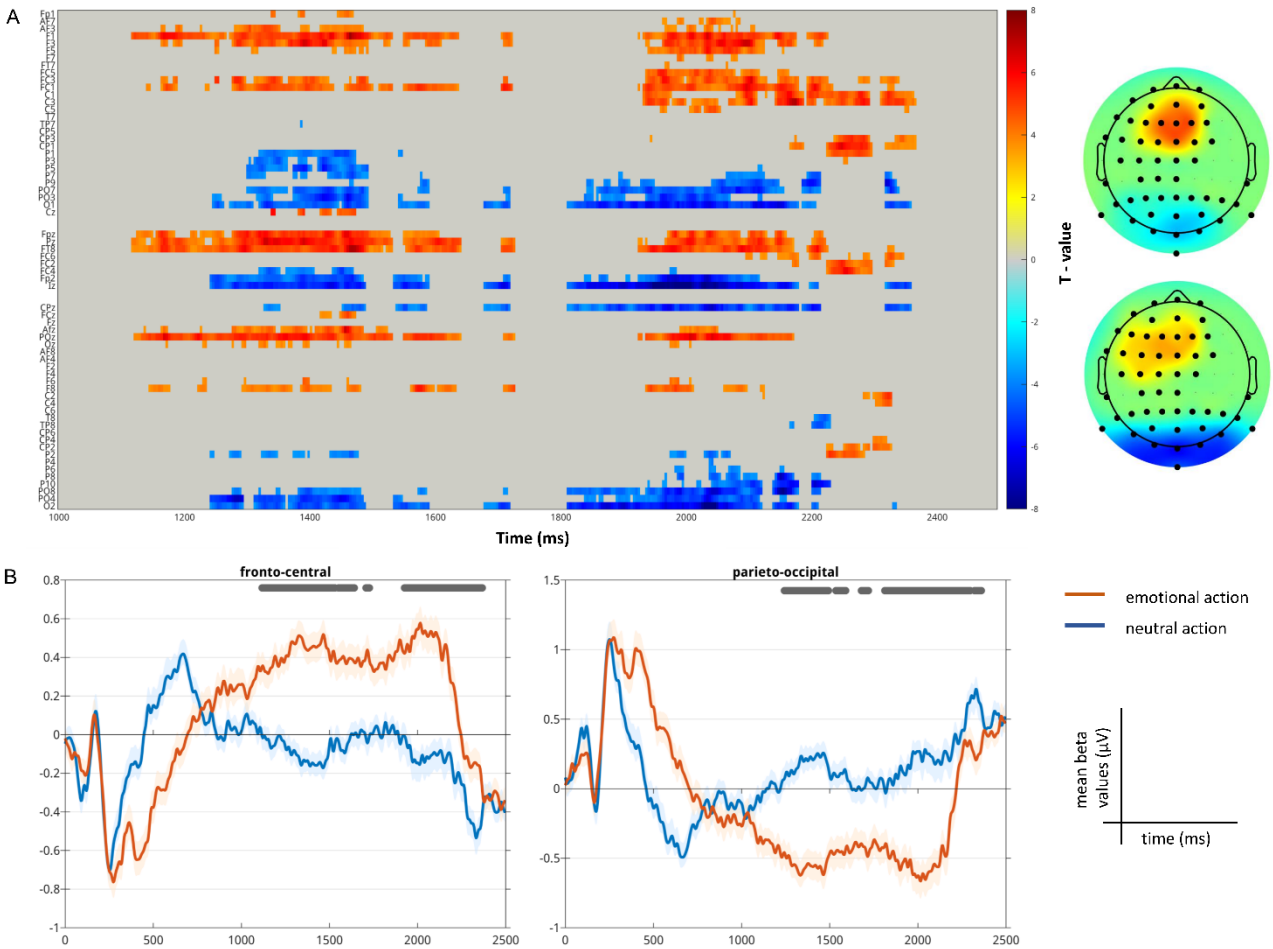


Figure 4.9. ERP results from the paired t-test (emotional action – neutral action), BM task. (A) Raster plot depicting significant channels and time-points for the paired t-test colored based on their t values. On the right the topographical representations of the t values averaged across the significant time-window for the two temporal clusters (the first on

the top, the second on the bottom). (B) Trace plots showing the mean beta values for the emotional action (in red) and neutral action conditions (in blue) averaged over the significant electrodes.

However, when we performed the skipped Pearson correlation between agreeableness scores and the ERP emotional action – emotional-neutrally action contrasts, no results survived after controlling for multiple comparisons within the entire sample, and either in female or male samples.

4.3.3 ERSP results

When performing the paired t-test between the ERSP responses associated with emotional action and emotional-neutrally action conditions, we replicated the results of Siqi-Liu et al. (2018) showing a μ-alpha suppression in the canonical 8-12 Hz over the centro-parietal-occipital electrodes (CP5, P3, P7, PO7, PO3, POz, PO4, O2) in a late time window between 1700-2000 ms (max *t* value – 6.88, at 19141 ms, on channel PO4, corrected *p* value .0061. Figure 4.10).

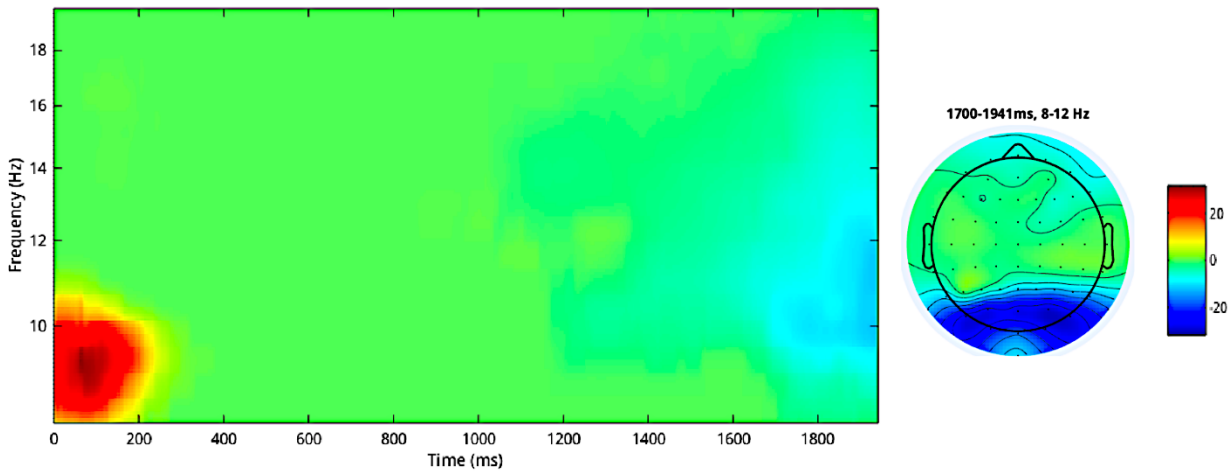


Figure 4.10. ERSP results from the paired t-test (emotional action – neutral action), BM task. Time-frequency plot of averaged frequencies from 8 to 20 Hz across all the 64 channels. On the right, the topographical representation of the *t* values averaged across the significant time-window, frequencies and electrodes.

However, same as for the ERP data, we did not find a modulation of agreeableness on the ERSP contrast values since no results from the skipped Pearson correlation survived after controlling for multiple comparisons within the entire sample, and either in female or male

samples.

4.3.4 Source Localization

Sources were estimated only for the RME task since no modulation of agreeableness was found either for the ERP or the ERSP data of the BM task.

Since, within the entire sample, three different clusters emerged from the ERP effects associated with agreeableness, we analysed separately the sources underlying these clusters and their correlation with agreeableness. To this purpose we carried out three paired t-tests (one for each ERP cluster), and used the significant resulting brain areas as ROIs to perform the skipped Pearson correlation with agreeableness scores. Among all the ROIs that emerged from the first two paired t-tests (ERP cluster 1 and 2), none survived the correction for multiple comparisons after the skipped Pearson correlation with agreeableness. Conversely, a significant positive correlation emerged between agreeableness and the activation for the emotion – age contrast in the ventro-medial prefrontal cortex (vmPFC; r = .444, corrected p = .018; Figure 4.11A,B) in the averaged time window of 840-902 ms (ERP cluster 3). Thus, the source activation in the vmPFC was greater for emotion *vs.* age stimuli in more agreeable subjects.

Noteworthy, we found a significant positive correlation between agreeableness and the activation for the sources contrast values (emotion – age) in the same ROI (vmPFC; r = .53, corrected p = .036; Figure 4.11B) within the male sample in a similar time window (848-887 ms), corresponding to the one of the ERP effects that were found to be associated with agreeableness. In line with the whole sample results, the source activation in the vmPFC was greater for emotion vs. age stimuli also in more agreeable males.

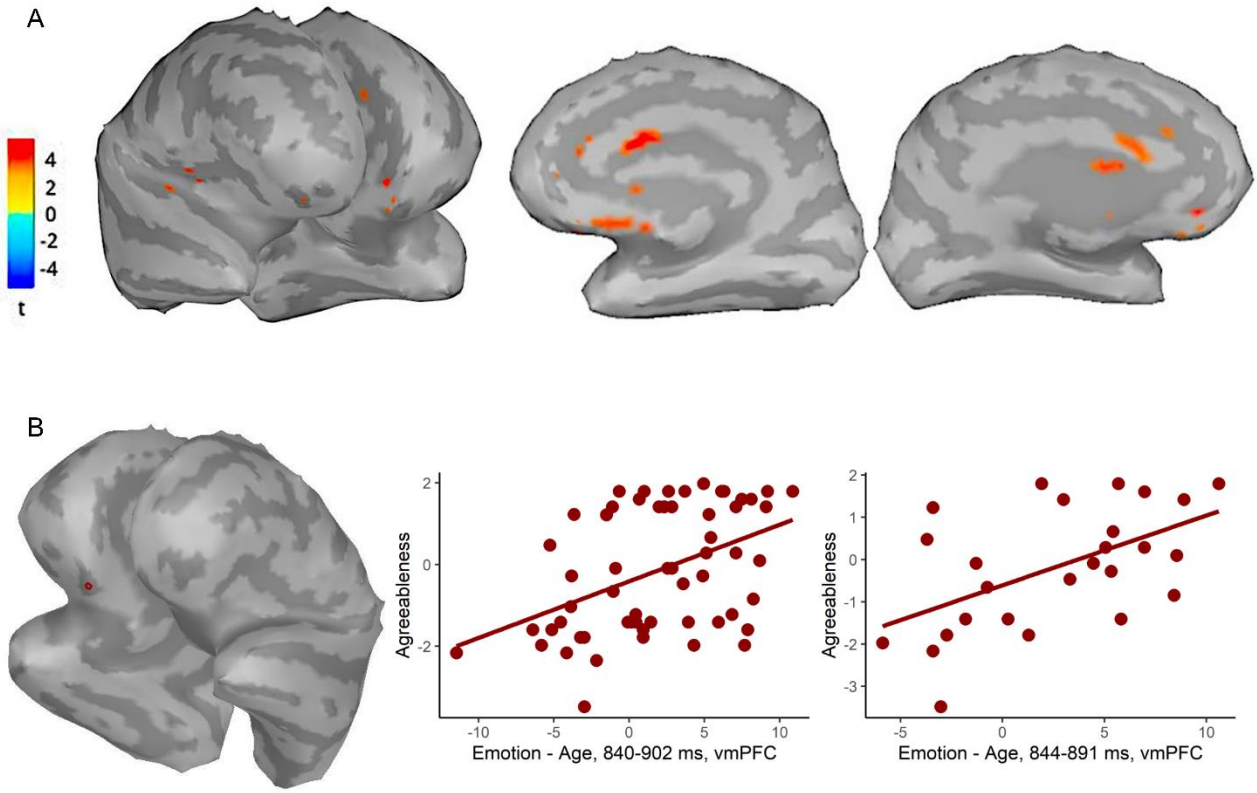


Figure 4.11. (A) Results from the paired t-test (emotion – age) performed on sources within the entire sample, RME task. (B) Results from the skipped Pearson correlations between agreeableness and source activations. On the right the vmPFC source. On the left scatterplots representing the correlation between agreeableness z-scores and the emotion – age contrast values in the vmPFC for the entire and male samples (respectively left to right).

4.4 Discussion

Agreeableness is a personality trait that has consistently been associated with social cognitive processes including ToM (DeYoung and Allen, 2017; Nettle and Liddle, 2008), as evidenced by recent neuroimaging studies revealing shared neural networks (Allen et al., 2017; Udochi et al., 2022). Individual variations in agreeableness were observed to be linked to differences in social information processing within the dmPFC, a key region of ToM. More agreeable individuals encoded social and non-social information in a more dissimilar fashion than less agreeable individuals, whose encoding of the two contents was more alike (Chapter 2).

To better understand what social information components are related to agreeableness, in

a previous EEG study we focused on the perception of social interactions, examining whether and at which stage agreeableness affects the underlying neural processes. We demonstrated that agreeableness was associated with differences in processing of social and non-social stimuli, occurring relatively early in time. This modulation was more consistent in male participants, and was also predictive of their performance (Chapter 3).

Since social stimuli are often intertwined with emotional aspects, in the present EEG study, we set out to investigate the role of agreeableness in modulating emotional information processing and its potential interaction with the complexity of the task and the underlying neural processes. To this purpose, we selected the BM task and the RME task, both targeting emotional cues decoding, albeit with different degrees of complexity. As the former requires emotional perception processes for discriminating basic emotions from point-light bodily stimuli, and the latter the decoding of higher mental states from eye expressions, agreeableness was expected to modulate the neural responses associated with emotional versus non-emotional conditions more strongly in the RME task and also later than in the BM task. Further predictions involved the role of sex in differentially modulating the relationship between agreeableness and emotional information processing (Chapter 3), and heightened sensitivity to gaze stimuli in individuals higher in agreeableness, even during passive observation.

The main results of the present study are the following. First, agreeableness modulated higher processes of mental state decoding, as tapped by the RME task, but not the emotional perception processes as assessed by the BM. In particular, relative to the control condition, higher agreeableness scores were associated with better accuracy in recognizing mental states, and with a modulation of ERP signals in a later time window, after the end of the stimulus presentation (~ 700-900 ms). However, when we analyzed male and female samples separately, no significant effects that lasted >= 20 ms were observed in females, suggesting that the modulation of agreeableness on the neural response in our female sample was negligible. In contrast, a later and larger cluster (844-891 ms) emerged in the male sample.

Both as a whole, and in the male sample, there was a consistent pattern in the effect of agreeableness on neural electrical activity. Participants with higher agreeableness scores exhibited a significantly higher difference between the ERPs associated with the two conditions, particularly driven by a heightened neural responsivity to emotional trials compared to non-emotional trials, which was reduced in less agreeable participants. This is consistent with the literature reporting that a late slow ERP component over frontal regions is associated with ToM processing (e.g., Liu et al., 2005; Meinhardt et al., 2011; Sabbagh and Taylor, 2000; Tesar et al., 2020; Wang et al., 2008; Zhang et al., 2009).

Second, source estimation procedures identified the vmPFC as a brain region candidate underlying the association between agreeableness and the emotional – non-emotional ERP effects. Specifically, highly agreeable individuals exhibited a stronger activation of this brain region when decoding mental states than in the control task (age decoding). This result was confirmed for the significant clusters observed both in the entire sample and in males.

Third, and in line with our hypotheses, individuals with high agreeableness exhibited similar neural responses during both passive observation of gaze stimuli and decoding of mental states. In contrast, a greater differential neural response between the two conditions emerged in less agreeable participants. This result should be interpreted with caution, as it does not survive multiple correction comparisons, however it might suggest that individuals with higher levels of agreeableness may decode mental states spontaneously, even during passive exposure to gaze stimuli. Future studies are needed to confirm this result.

Taken together, these findings suggest that personality modulates the decoding of emotional cues and that this modulation seems to depend on task complexity. Indeed, the RME task constitutes the first stage of ToM processing (Baron-Cohen et al., 2001), requiring the recognition of considerably more complex mental states, solely through static stimuli capturing individuals' gaze; in contrast, BM task requires discriminating basic emotions, primarily engaging neural perceptual processes from point-light bodily movements that allow a quick perception and

categorization of social information and emotions (Atkinson et al., 2004; Dittrich et al., 1996). In light of this evidence, we could assume that lower-order processes tapped by the BM might not be involved in personality, since they are more automatic and require fewer cognitive resources. On the other hand, agreeableness seems to foster a marked distinction in neural responses associated with the decoding of mental *vs.* non-mental states. This evidence is consistent with a higher responsivity and a more in-depth processing of emotional stimuli for effective mental states decoding in more agreeable subjects, as indicated by the increased late slow wave (LSW) observed during emotional trials, and even during passive observation of gaze stimuli. This wave has traditionally been recognized as one of the ERP components associated with cognitive aspects of ToM (see Sabbagh, 2013 for a review), however our study provides evidence of its involvement even in affective aspects of ToM. Furthermore, the LSW has been associated with emotional encoding by facilitating the integration of emotional and cognitive information, emotional intensity (Hajcak et al., 2009; Olofsson et al., 2008; Schupp et al., 2000; Tesar et al., 2020; Yick et al., 2015; see Hajcak et al., 2010 for a reiew), and also used as an index of abnormal emotional responding (Dennis and Hajcak, 2009; Horan et al., 2010; Marissen et al., 2010).

Our interpretation is consistent with the involvement of the vmPFC observed from source estimation procedures, although it should be taken with caution since individual anatomies were not available. This brain region, which is highly connected to the amygdala, together with the TPJ, represents the core region of self-other distinction and affective ToM, and in particular, it has been associated with the decoding of others' emotional states, thus with the RME task (Sabbagh et al., 2004; Schurz et al., 2014, 2021; see the entire volume by Baron-Cohen et al., 2016).

As a fourth and last point, the sex of participants seems to modulate agreeableness, similarly to what we observed in our previous EEG study on the decoding of social interactions wherein the modulation of agreeableness lasted longer in males than in females (Chapter 3). This finding supports our hypothesis that personality may play a role in processing social stimuli differently in males and in females. More specifically, our results align with the sex differences

observed in the social domain, where females tend to be more socially oriented and are often described as possessing inherent sensitivity to non-verbal social and emotional cues, (Baron-Cohen et al., 2005; Hall and Halberstadt, 1981; Paletta et al., 2022; Pavlova et al., 2010; Proverbio, 2008, 2021; Stake and Eisele, 2010). This inherent sensitivity likely contributes to the lacking impact of personality on emotional cue decoding observed in females. Therefore, our findings renew the critical importance of considering the sex of participants when investigating the neural mechanisms underlying inter-individual variations in personality traits.

However, when the behavior was considered, only a weak modulation of agreeableness emerged in the sample as a whole. This may be partially due to the fact that we recruited healthy participants from the general population, without any overt deficit affecting their emotional processing abilities. Agreeableness may still influence the neural process underlying mental state recognition, even without affecting the ability to recognize them, as observed also in our previous EEG study, wherein the modulation of agreeableness at the behavioral level was weak (Chapter 3). Although a similar null result between agreeableness and RME performance was reported by Nettle and Liddle (2008), which led them to conclude that this personality trait is unrelated to the affective component of ToM associated with mental state decoding, findings from our current study challenge this assertion, suggesting a different perspective.

4.5 Conclusion

The current study expands our understanding of the role of personality in social skills by revealing how neural mechanisms underlying emotional cues decoding are modulated by the agreeableness trait as a function of task difficulty and sex. Our results show that agreeableness affects brain's activity during mental state decoding in male participants, while in females this modulation seems to be negligible. Considering the limited body of research on this particular topic, further replication studies are needed to better understand the role of personality in emotional processing in both healthy and clinical populations and its effects on behavior. Overall,

these findings offer significant insights into the field of individual differences within the social domain, holding potentially useful implications for clinical interventions aimed at addressing social impairments, developing customized interventions that account for individual variations, thus contributing to the advancement of tailored and effective therapeutic approaches.

CHAPTER 5

Study 4: Does ASD individuals' personality affect the decoding of mental states?

Abstract

In the previous chapters, I hopefully provided evidence that personality modulates different socio-emotional processes associated with ToM in healthy individuals. In the present chapter, I will report a study in which I investigated the impact of personality on the same processes in a sample of young- adults with ASD without language and intellectual impairment, collecting behavioral and EEG data. I compared their personality profiles with those of normally developed control participants. Results showed that individuals with ASD obtained significantly lower scores in all the Big Five except for openness. Noteworthy, considerable variability persisted, reflecting a heterogeneity in personality traits even within the ASD group. When the agreeableness trait was considered, given its interplay with social and emotional processes, individuals within the ASD spectrum exhibited comparable accuracy to less agreeable controls while decoding mental states. However, their accuracy was significantly lower compared to controls with higher levels of agreeableness. ERP results showed a delayed and reduced N170 during stimulus processing for ASD participants compared with controls, in addition to a smaller LSW only when decoding mental states. This latter difference was more extensive and sustained over time for ASD vs. more agreeable controls than ASD vs. less agreeable ones. We suggested the presence of compensatory processes linking behavioral and neural responses. Overall, these findings advocate for a holistic understanding of individuals' unique attributes and potential strengths, paving the way for more effective interventions tailored to the specific needs of each individual along the autism spectrum.

5.1 Introduction

In the previous chapters, I hope I demonstrated that personality exerts an influence on how our brain encodes social stimuli. Specifically, I showed that agreeableness shapes social and emotional information processing associated with Theory of Mind (ToM), including the detection of social interactions and the decoding of complex mental states, through focused attention and deeper encoding of the stimuli (Chapters 2-4).

While in the previous studies participants without specific challenges were recruited, in the present study I have involved individuals with Autism Spectrum Disorder (ASD), a clinical label used to characterize individuals with a variety of early-onset developmental difficulties in social communication and restricted repetitive behaviors (Lai et al., 2013; Lord et al., 2020; Warrier et al., 2018).

In the DSM-5 (American Psychiatric Association, 2013) the previous subcategories of pervasive developmental disorders (autism, Asperger, and pervasive developmental disorders not otherwise specified) were merged into this spectrum facilitating these people's sense of identity and service access. However, this single diagnostic label encompasses a wide heterogeneity that makes it difficult to understand and capture the different underlying neurobiological mechanisms of these people and to provide the spread of differentiated services in the society (Bertelsen et al., 2021; Mottron, 2020; Vincent et al., 2019). This has also influenced research, which has predominantly focused on low-functioning ASD, neglecting other categories within the spectrum (Mottron, 2020). Particularly evident is the case of the former Asperger's Syndrome, whose distinct characteristics underscore the necessity for more personalized approaches in intervention design to facilitate better and positive outcomes (Bertelsen et al., 2021; Mottron, 2020; Vincent et al., 2019). Indeed, the literature highlights the precarious social and occupational status, as well as the compromised overall quality of life in young adults with this diagnosis now officially called ASD without language and intellectual impairment (Vincent et al., 2019), rendering them more susceptible to psychiatric disorders (Barnhill, 2007) with considerable difficulties also for their families.

Several theories have been developed to date trying to explain ASD difficulties in social interaction but the exact nature of their social-emotional deficits is still not well understood (for a review: Barak and Feng, 2016; Brighenti et al., 2018). However, in several studies ASD individuals were found to have intact behavioral ability as well as electrophysiological differences in social information processing, suggesting the possibility of compensatory strategies, especially

in ASD with high functioning like those without language and intellectual impairment (Coben et al., 2013; Fan et al., 2014; Kang et al., 2018; Lerner et al., 2013; Sokhadze et al., 2015; Wong et al., 2008; Yang et al., 2011). The variability in these studies that investigated social and communication processes in ASD could be due to differences in personality, especially in patients without language and intellectual impairment who, despite facing social challenges, demonstrate a greater ability to adapt within their environment and establish relationships compared to those with lower functioning ASD (Mali and Singhi, 2015; Solomon et al., 2004; Woodbury-Smith and Volkmar, 2009).

The role of personality and individual differences in autism is an emerging research topic that has received limited attention to date. Individuals with ASD tend to exhibit lower levels of openness, conscientiousness, extraversion, agreeableness, and emotional stability (Lodi-Smith et al., 2019) However, the observed correlations are small to moderate and this evidence is not consistent across all studies in the literature (Schwartzman et al., 2016). This variability in personality traits might help us in understanding the achievements and difficulties experienced by individuals with ASD. For instance, in children with ASD, neuroticism, extraversion, and agreeableness have been related to internalizing and externalizing problems (De Pauw et al., 2011; Schriber et al., 2014). Similarly, maladaptive personality characteristics among adults with ASD appear to negatively affect the quality of life (Ratner and Burrow, 2017; Rodgers et al., 2018; Schwartzman et al., 2016).

However, up to now, researchers focused only on behavior, mainly using self-reports or other-reports with conflating results from samples that differ in age (i.e., children and adults) and diagnosis (i.e., high and low functioning), therefore making it difficult to interpret the results (Lodi-Smith et al., 2019; Robinson et al., 2020; Schwartzman al., 2016). Furthermore, behavioral studies are insufficient to unveil eventual compensatory strategies, as they only show the final output of complex processes not accessible from the outside. Therefore a neural investigation is needed, and EEG is a very suitable technique for this purpose thanks to its noninvasive nature

which allows us to investigate the neural mechanisms underlying the interaction between personality and ToM. Results are expected to explain the neural mechanisms underlying these processes in ASD and to foster tailored interventions to improve skills based on their personality profile focusing on their assets for positive intervention strategies and social inclusion.

More specifically, the goal of the present study is to explore the role that personality might have in modulating the different social-emotional processes underlying ToM investigated in the previous EEG studies by means of the same experimental tasks, in subjects with ASD without language and intellectual impairment by comparing them with data from healthy controls. Specifically, the present study aims to address three distinct research questions:

1. Do individuals with ASD without language and intellectual impairment have a characteristic personality profile? Moreover, does intersubject variability persist within this sample, or do all subjects display markedly negative scores across all Big Five traits?

2. Could young adults with ASD without language and intellectual impairment resemble healthy controls with lower agreeableness scores?

3. During the abovementioned experimental tasks, could ASD participants exhibit similarities to less agreeable healthy controls, while also showing significant differences from more agreeable controls in terms of both behavior and neural processing of social and emotional stimuli?

First, we hypothesized a lower overall personality profile of young adults diagnosed with ASD compared to controls. However, due to the unique nature of their diagnosis - ASD without language and intellectual impairment - and their social adaptation, we also expected to observe considerable intersubject variability within this group. It is noteworthy that these individuals differ significantly from those with low-functioning ASD in terms of environmental adaptation (Kamp-Becker et al., 2010; Liss et al., 2001; Renty et al., 2006). With respect to agreeableness, we predicted the group of ASD to be comparable to less agreeable healthy controls.

Concerning our third research question, we predicted to observe differences both in

behavioral performance and neural responses between the ASD group and more agreeable healthy controls. In contrast, we did not expect differences between ASD and less agreeable controls. Specifically, we hypothesized worse accuracy and reduced ERP neural responses in social-emotional conditions, as opposed to neutral ones, for the ASD group relative to more agreeable controls. However, we did not expect such differences when comparing ASD to less agreeable controls. For ERSP neural responses, we hypothesized a significant difference between ASD and more agreeable controls expressed as a stronger μ-suppression for more agreeable participants, while no differences are expected between ASD and less agreeable subjects.

5.2 Methods

5.2.1 Participants recruitment

Forty-two Italian young adults with ASD without language and intellectual impairment (14 females, mean age = 26, SD = 8.98, range = 17-45 years) were recruited from a specialized clinical centre for autism placed in Trieste, namely the Centro Socio Sanitario Integrato per i Disturbi dello Spettro Autistico e le Disabilità Cognitive e/o Intellettive, Azienda Sanitaria Universitaria Giuliano Isontina (ASU GI). Four subjects did not complete the entire experiment, therefore their data were excluded leading to a final sample of 38 patients (13 females, mean age = 25, SD = 8.14, range = 17-44 years) The diagnosis was assessed by the clinicians of the clinical centre both qualitatively and quantitatively according to DSM-5 criteria. Exclusion criteria included a history of head injury, epilepsy, neurological or motor disorders, and QI < 70. All participants provided written informed consent and received 30 euros upon completion of the study. Informed consent was obtained from the parents of minor participants, with assent from the children prior to inclusion. The study and informed consent procedures were approved by the Regional Ethics Committee of Friuli Venezia Giulia and adhered to the guidelines of the Declaration of Helsinki. Participants from the two previous EEG studies (Chapters 3-4) were included as healthy controls for comparative analysis. Particularly, in those studies, we initially collected personality scores by

means of the Big Five Inventory (Ubbiali et al., 2013) from 462 participants, then, from this larger group, we specifically selected 62 subjects based on their agreeableness scores to increase in-sample variance. In the present study, we initially address our first research question by comparing the personality profiles of the new ASD sample with the initial pool of 462 participants. Following this comparison, we focused on the subsequent research questions using the subset of healthy controls (N = 62) that were specifically selected for further analysis.

5.2.2 Procedure and experimental tasks

Data from ASD subjects were collected during two experimental sessions. In the first session, participants completed a series of psychological measures: the Big Five Inventory (BFI; Ubbiali et al., 2013), the 10-item short forms of the Autism Spectrum Quotient (AQ-10), Empathy Quotient (EQ-10), Systemizing Quotient-Revised (SQ-R-10), and the Sensory Perception Quotient (SPQ-10) (Greenberg et al., 2018). In the second session, participants engaged in three experimental tasks during EEG recording: the Social interaction task (SIT), the Biological (e)motion task (BM), and the Reading the mind in the eyes task (RME). The order of the tasks was counterbalanced across subjects. For their detailed descriptions, please refer to Chapters 3-4.

5.2.3 EEG recording and preprocessing

Data collection and preprocessing procedures mirror those employed in Chapters 3 and 4. The average number of epochs within the ASD group used for the analysis was: 108 in both conditions for the SIT, ranging from 80 to 118 epochs; 46 in both conditions for the BM, ranging from 26 to 50 epochs; 46 in all the conditions for the RME, ranging from 36 to 50 epochs.

5.2.4 Behavioral data analysis

To explore the first research question, the scores of each personality trait assessed by means of the BFI, were transformed into z-scores utilizing the mean scores within the Italian population as a reference. This approach facilitated an examination of the participants' positioning relative to the Italian population's normative values. Furthermore, we conducted Welch's t-tests to compare

the z-scores between the ASD group and the wide healthy control sample (N = 462). Afterward, since among the personality traits under investigation, special attention is directed towards agreeableness, we divided our previous sample of healthy control (N = 62) selected for agreeableness by means of the median-split of their agreeableness score (median = -.09) obtaining two groups: the high agreeableness (HA: mean = 1.33; SD = .54) and the low agreeableness group (LA: mean = -1.37, SD = .75). Then, we performed a linear mixed-effects model to test differences of agreeableness across groups. Particularly, to test our hypothesis that the ASD group might significantly differ from the HA group but be quite similar to the LA group, we first used agreeableness as dependent variable of our model and "group" as a fixed categorical predictor with three levels: HA, LA and ASD group. Significant interactions were decomposed using post hoc comparisons.

To analyze the research question 2, we collected accuracy and reaction time (RT) data with the experimental tasks. RTs were filtered to exclude errors, anticipated responses (< 150 ms), and outliers exceeding 3 standard deviations from the subject's mean for each condition. For the behavioral analyses, we examine the impact of group on different behavioral measures.

For the SIT, Signal detection theory indices were computed for accuracy measures. The sensitivity index (d') was used to quantify the participant's ability to discriminate social from non-social stimuli. The formula for calculating d' is: d' = ZHIT - ZFA, where Hit represents the proportion of social trials to which subjects responded "social" (hit rate) and FA represents the proportion of non-social trials to which subjects responded "social" (false-alarm rate). Both rates were Z-transformed (Macmillan and Creelman, 1990; Stanislaw and Todorov, 1999). Next, we performed a linear mixed-effect model analysis to compare groups' sensitivity to differentiate social from non-social stimuli. The sensitivity index (d') was used as the dependent variable, group was included as a fixed effect, and subject intercepts were considered as random effects.

For the BM and RME tasks, we performed linear mixed-effects models separately for the emotional and non-emotional conditions. Accuracy and RTs were used separately as dependent

variables, group as fixed effect, and subject intercepts as random effects. Assumptions for all the models were checked through the visual inspection of residual plots (linearity, homoscedasticity, and normality). All analyses were performed in R v.4.1.1 (R Core Team, 2020).

5.2.5 ERP analysis

ERP analyses were carried out separately for each task by using Matlab R2021a along with EEGLAB 2020_0 and LIMO EEG plugin (v 3.3, https://github.com/LIMO-EEG-Toolbox/limo_tools). A massive univariate approach was adopted by performing two levels of analysis for both the two tasks (Pernet et al., 2011).

First-level GLM analysis

At the first level, parameters of a General Linear Model were estimated by using a standard ordinary least squares solution for each subject, at each time point and each electrode independently, and separately for all the experimental conditions. The considered time window ranged from 0 to 1000 ms for the SIT and the RME task, time locked to stimulus-target onset. For the BM task, in which the stimulus had a longer duration being a video, in line with the findings of Siqi-Liu et al. (2018) and with our previous EEG study (Chapter 4), we started the analysis from 1000 ms discarding the time points related to early neural processes which were not our target, and instead keeping the time points following the end of the stimulus as for the other two tasks. Erroneous trials were retained in the model to preserve higher inter-individual variability, as we predicted that accuracy would vary based on agreeableness scores.

Second level GLM analysis

At the second level, robust two samples t-tests were carried out for each task across all time points, and all electrodes testing the differences between groups within each experimental condition (ASD – HA; ASD – LA; HA – LA). Results are reported corrected for multiple testing using spatial-temporal clustering $p = .05$ (Maris and Oostenveld, 2007; Pernet et al., 2015).

5.2.6 ERSP analysis

Since our hypothesis specifically targeted a possible association between agreeableness and μ-alpha suppression, as for Chapter 4, ERSP analysis was performed only for the BM task by using Matlab R2021a along with EEGLAB 2021_0 and LIMO EEG plugin (v 3.3). As for the ERPs, a massive univariate approach was adopted by performing two levels of analysis (Pernet et al., 2011).

First-level GLM analysis

At the first level, parameters of a General Linear Model were estimated for each subject, at each time point, each electrode, and each frequency in the range between 8-12 Hz independently, and separately for the experimental conditions, namely emotional action and neutral action. The time window ranged from 0 to 2500 ms time locked to stimulus onset, excluding edge artifacts induced by the wavelet transform. Parameters were estimated by using a standard ordinary least squares solution.

Second level GLM analysis

At the second level, robust two samples t-test tests were carried out across all time points, frequencies, and electrodes testing the differences between groups within each experimental condition (ASD – HA; ASD – LA; HA – LA). Results are reported corrected for multiple testing using spatial-temporal clustering $p = .05$ (Maris and Oostenveld, 2007; Pernet et al., 2015).

5.3 Results

5.3.1 Behavioral results

Personality traits assessment. Figure 5.1 shows the positioning of our sample of ASD subjects with respect to the Italian population mean for each personality trait whose z-scores are represented on the y-axis and with respect to our large sample of healthy controls ($N = 462$). Welch's t-test revealed that ASD scored significantly higher in neuroticism ($t = 2.71$, df = 42.6, p value = .009) and lower in agreeableness ($t = -7.34$, df = 45.54, p value < .0001), extraversion ($t =$

-5.61, df = 43.82, *p* value < .0001), and conscientiousness (*t* = -3.25, df = 40.76, *p* value = .002) than healthy controls. No significant differences were found for the openness trait (*t* = -.68, df = 45.77, *p* value = .5).

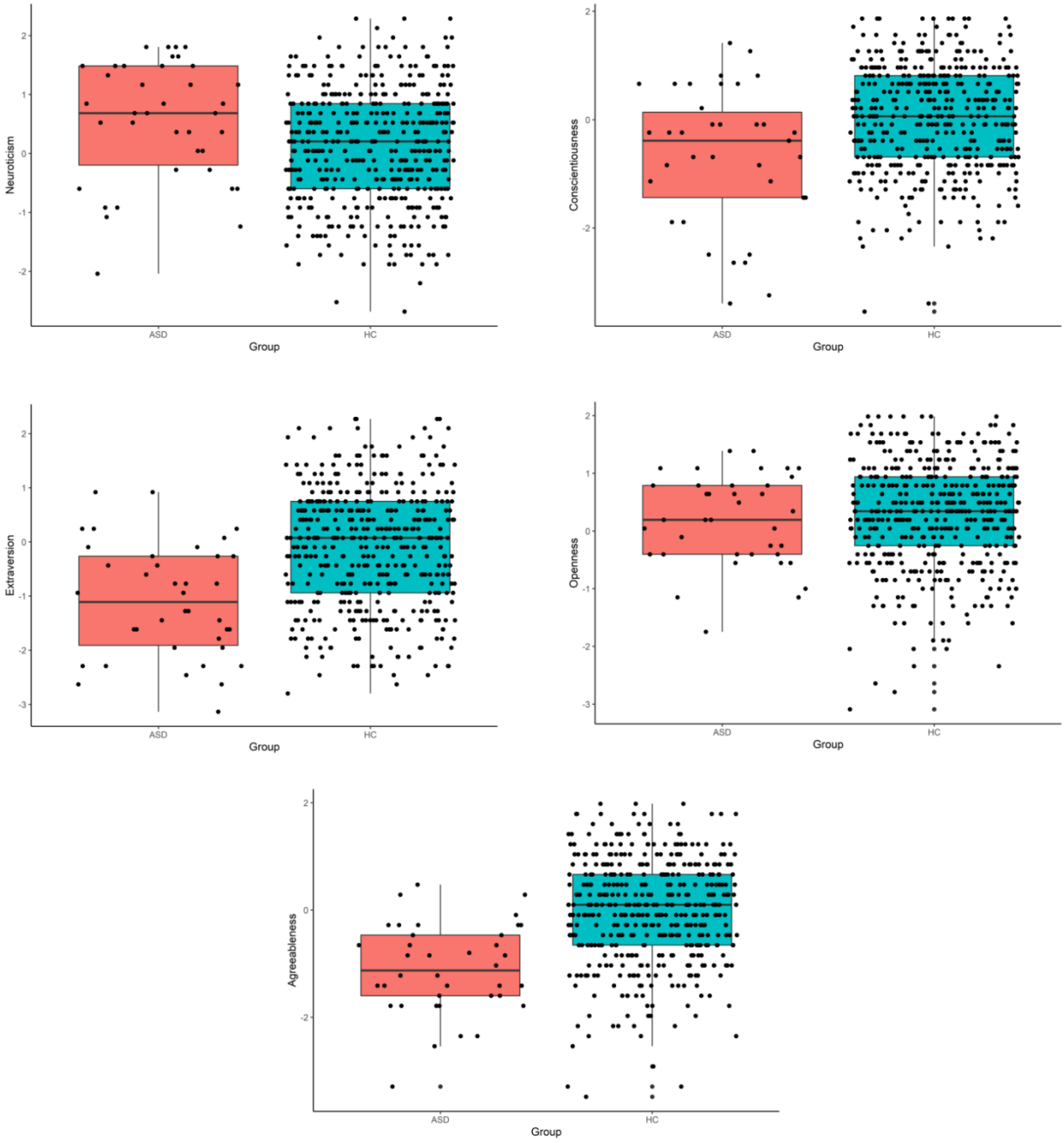


Figure 5.1. Personality trait comparisons in ASD and HC. Each trait is represented on the y-axis in z-scores of the Italian population mean, ASD and wide HC groups are represented on the x-axis, each spot represents one subject. Each box represents the interquartile range (IQR), with the median indicated by a horizontal line.

The larger difference was the one associated with the agreeableness trait, indeed the ASD group seemed to overlap with those healthy controls having scores below the mean. As a matter of fact, when dividing the selected control sample (N = 62) into two groups by the median-split of their agreeableness scores and performing the agreeableness × group model, significant results emerged (F(2, 35) = 130.2, p value < .0001, η^2 = .73). Post hoc comparisons revealed that the ASD group was significantly lower in agreeableness than the HA group (p value < .0001), but quite similar to the LA group (p value = .19); HA and LA significantly differed (p value < .0001) (Figure 5.2).

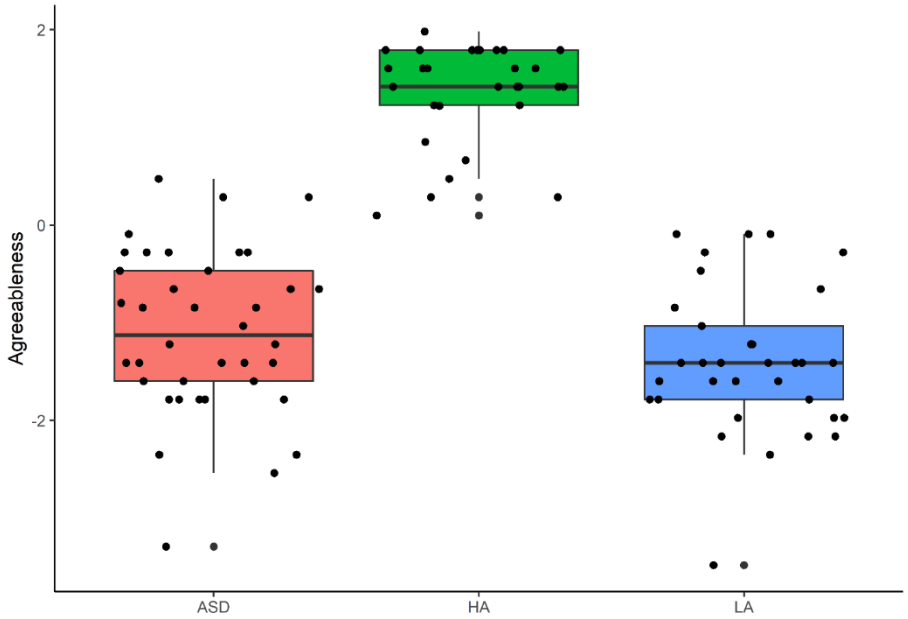


Figure 5.2. Agreeableness z-scores comparison across three groups: ASD, HA, and LA. Each box represents the IQR, with the median indicated by a horizontal line.

SIT. No significant differences across groups emerged from the linear mixed-effects model comparing the ability to discriminate between social and non-social stimuli, as measured by the d' index (F(2, 35) = .69, p = .51, η^2 = .01; Figure 5.3).

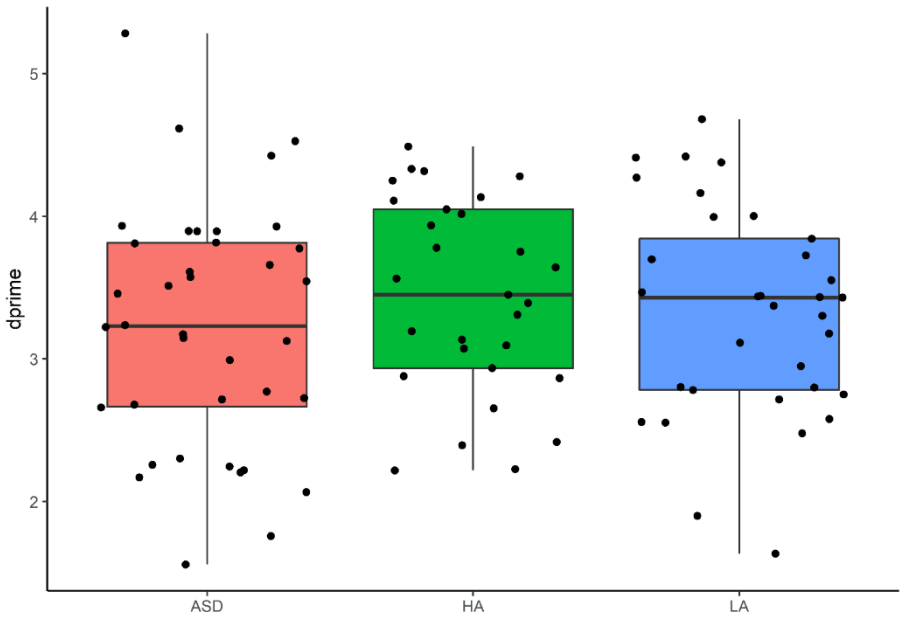


Figure 5.3. Dprime scores comparison across ASD, HA, and LA groups. Each box represents the IQR, with the median indicated by a horizontal line.

BM. No significant results emerged from the linear mixed-effects models carried out on accuracy (Figure 5.4) and reaction times in either the emotional actions condition (accuracy × group: $F(2, 35) = 2.1$, $p = .13$, $\eta^2 = .04$; RTs × group: $F(2, 35) = .83$, $p = .45$, $\eta^2 = .02$) or emotionally-neutral actions condition (accuracy × group: $F(2, 35) = 1.5$, $p = .24$, $\eta^2 = .03$; RTs × group: $F(2, 35) = 1.02$, $p = .37$, $\eta^2 = .02$), therefore the three groups did not significantly differ in their behavioral performances.

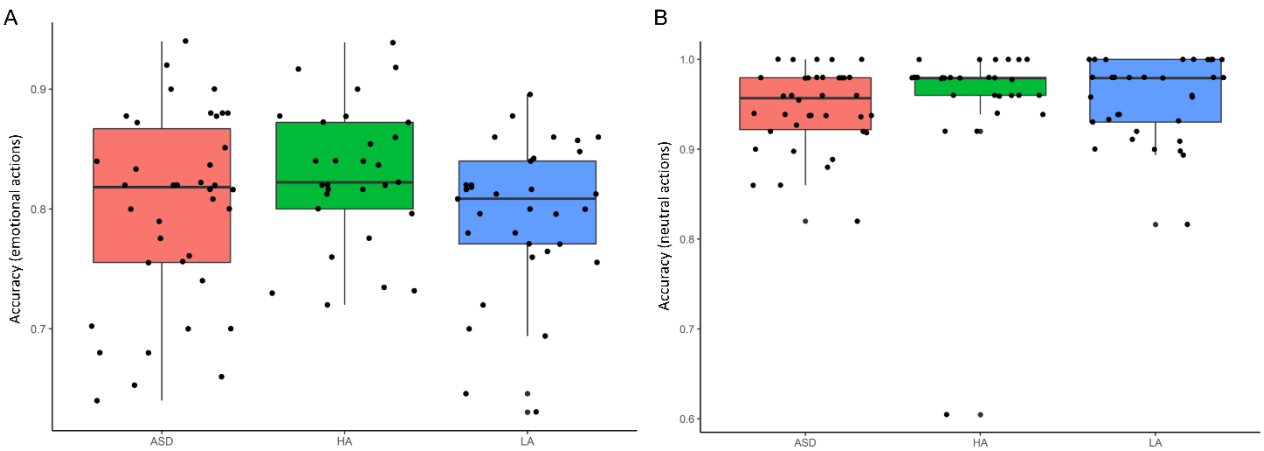


Figure 5.4. Accuracy scores comparison across ASD, HA, and LA groups within the (A) emotion and (B) neutral condition. Each box represents the IQR, with the median indicated by a horizontal line.

RME. The linear mixed-effects model testing accuracy differences across groups in recognizing complex emotional states revealed significant results ($F(2, 35) = 3.74$, $p = .033$, $\eta^2 = .07$; Figure 5.5A). Post hoc comparisons showed a significant difference between ASD and HA group (p value = .025) and no difference for ASD and LA group (p value = .19) as expected. No significant difference where observed between HA and LA groups (p value = .25). No differences emerged within the age condition ($F(2, 35) = 2.08$, $p = .14$, $\eta^2 = .04$; Figure 5.5B). Furthermore, the groups did not differ in terms of RTs in both the experimental conditions (emotion: $F(2, 35) = .73$, $p = .49$, $\eta^2 = .01$; age: $F(2, 35) = .55$, $p = .58$, $\eta^2 = .01$).

Visual inspection of residual plots did not reveal any obvious deviations from linearity, homoscedasticity, or normality in all the three tasks.

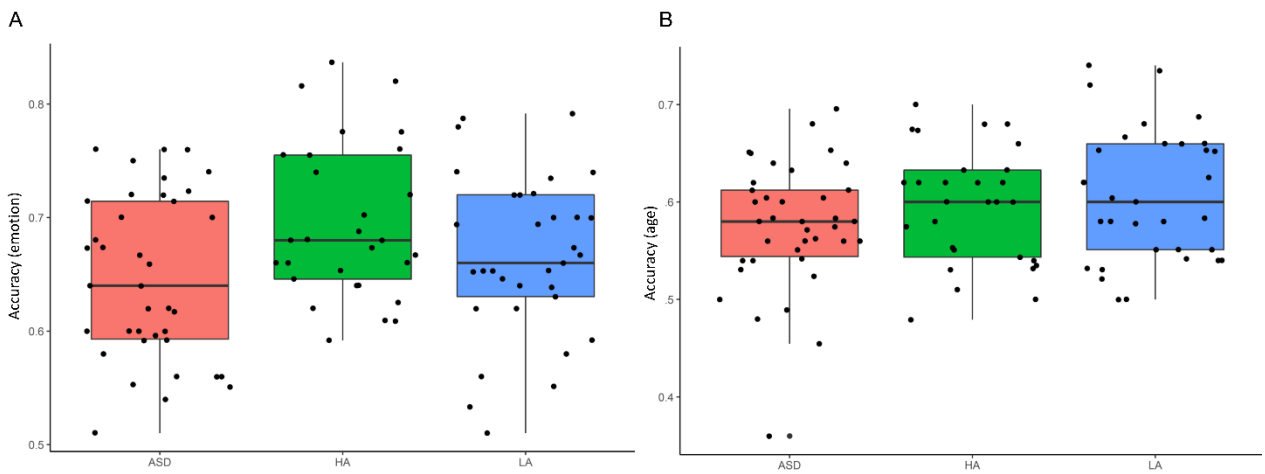


Figure 5.5. Accuracy scores comparison across ASD, HA, and LA groups within the (A) emotion and (B) age condition. Each box represents the IQR, with the median indicated by a horizontal line.

5.3.2 ERP results

SIT. From all the two samples t-tests, no differences survived the correction for multiple comparisons, therefore the three groups did not seem to process differently the stimuli of this task, mirroring the behavioral results.

BM. Also for this task, no significant differences emerged across groups from the analysis

performed in any of the experimental conditions, as in the behavioral results.

RME. When considering the emotional condition, as for the behavioral results, a significant difference emerged between the ASD and HA group in 2 clusters (cluster 1: 121-168 ms, max *t* value 6.09 at 141 ms on channel P10, corrected *p* value .031; cluster 2: 453-1000 ms, max *t* value -6 at 867 ms on channel Fz, corrected *p* value .003; Figure 5.6A). The difference in the first cluster was led by a delayed and reduced negative amplitude for the ASD group over the parieto-occipital electrodes (Figure 5.8A). On the other hand, the difference in the second cluster was driven by a smaller positive amplitude for the ASD group than the high agreeableness over the fronto-central electrodes (Figure 5.8A).

A significant difference emerged also when comparing ASD with LA group in 2 clusters (cluster 1: 113-176 ms, max *t* value 6.4 at 133 ms on channel P10, corrected *p* value .007; cluster 2: 402-590 ms, max *t* value -4.68 at 570 ms on channel FCz, corrected *p* value .007; Figure 5.6B). The first cluster matched the first one found for the t-test ASD – HA, therefore it was explained by a delayed and reduced negative neural response for the ASD group compared to the LA group (Figure 5.8A). As regards the second cluster, although the time window was earlier than the one observed for ASD – HA, it pointed in the same direction with a smaller positive amplitude for the ASD compared to the LA group over fronto-central electrodes (Figure 5.8A).

When considering the age condition, no significant differences emerged from the contrast ASD – HA. However, when comparing ASD with LA, one cluster emerged, perfectly overlapping with the ones found within the emotional condition starting at 121 ms and ending at 168 ms (max *t* value 5.77 at 148 ms on channel P9 correct *p* value .024) driven by a reduced negative amplitude for the ASD group over the parieto-occipital electrodes.

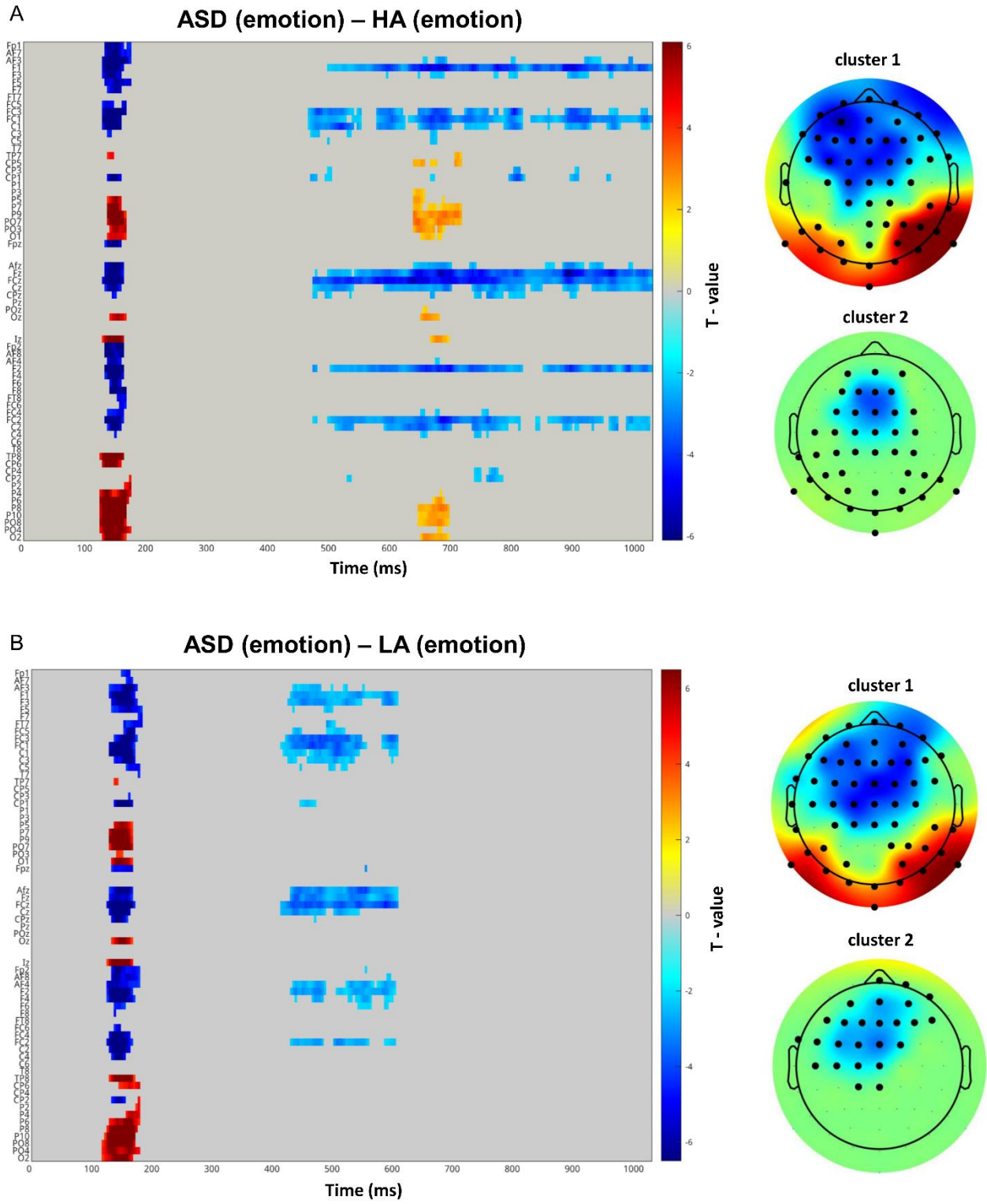


Figure 5.6. ERP results from the two sample t-tests (A) ASD – HA, and (B) ASD – LA within the emotion condition. On the left, raster plots depicting significant channels and time-points for the two sample t-tests colored based on their *t* values. The electrodes on the y-axis are arranged topographically: left and right sides of the head are organized respectively on the top and the bottom part of the axis, midline electrodes are displayed in the middle. Thus, overall the y-axis represents the scalp anterior to posterior. On the right the topographical representations of the *t* values averaged across the significant time-window of each cluster.

Finally, when considering the observe condition, for both the two samples t-tests, namely the ASD – HA and the ASD – LA, a very similar cluster emerged overlapping with the first found within the other two experimental conditions (cluster ASD – HA: 125-164 ms, max *t* value 5.3 at 152 ms on channel P9 correct *p* value .04 – Figure 5.7A; cluster ASD – LA: 125-168 ms, max *t* value 6.51 at 145 ms on channel Iz correct *p* value .013 – Figure 5.7B). As for the previous results regarding this cluster, the differences across groups were driven by a reduced and delayed neural response over parieto-occipital electrodes for the ASD compared to the other two groups (Figure 5.8B). Therefore, this early effect related to this cluster, always found from the two samples t-tests, seemed not to be related to agreeableness but reflected a difference between the ASD and healthy controls in processing the stimuli independently from the experimental condition. This effect seems clearly to refer to the N170 ERP component, extensively associated with face processing, generally right-lateralized and involving parieto-occipital sites (Feuerriegel et al., 2014; Kappenman et al., 2021; Rossion et al., 2003) that has been demonstrated to be delayed and/or reduced in ASD (see Kang et al., 2018 for a meta-analysis).

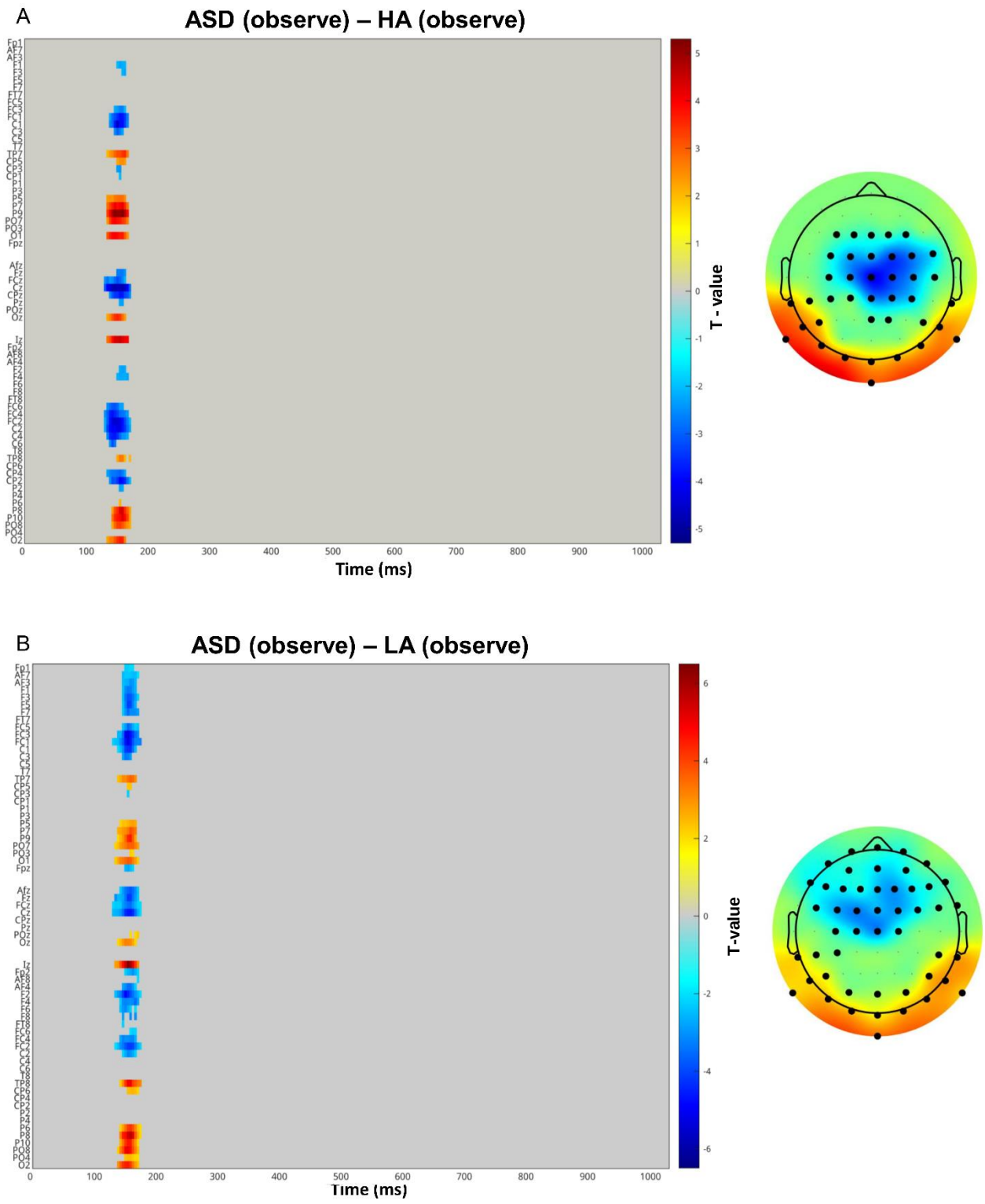
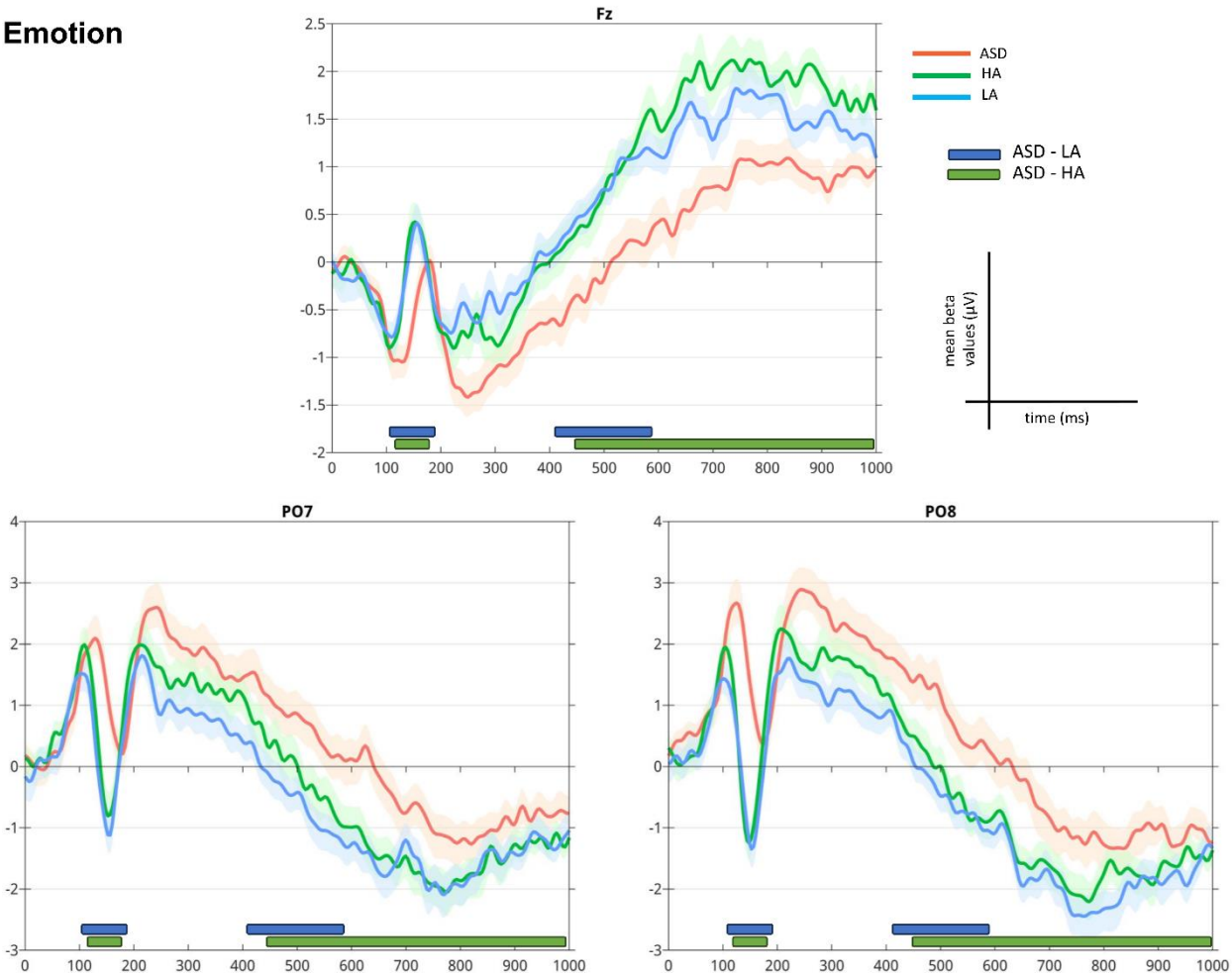


Figure 5.7. ERP results from the two sample t-tests (A) ASD – HA, and (B) ASD – LA within the observe condition. On the left, raster plots depicting significant channels and time-points for the two sample t-tests colored based on their *t* values. The electrodes on the y-axis are arranged topographically: left and right sides of the head are organized respectively on the top and the bottom part of the axis, midline electrodes are displayed in the middle. Thus, overall the y-axis represents the scalp anterior to posterior. On the right the topographical representations of the *t* values averaged across the significant time-window of each cluster.

No significant differences emerged between the HA and LA group in any of the three experimental conditions, although within the emotional condition the figure 5.8A shows a slightly smaller neural response for the LA than the HA group leading the LA group to have an intermediate position between the ASD and HA group. However, this difference did not survive the correction for multiple comparisons.

A

Emotion



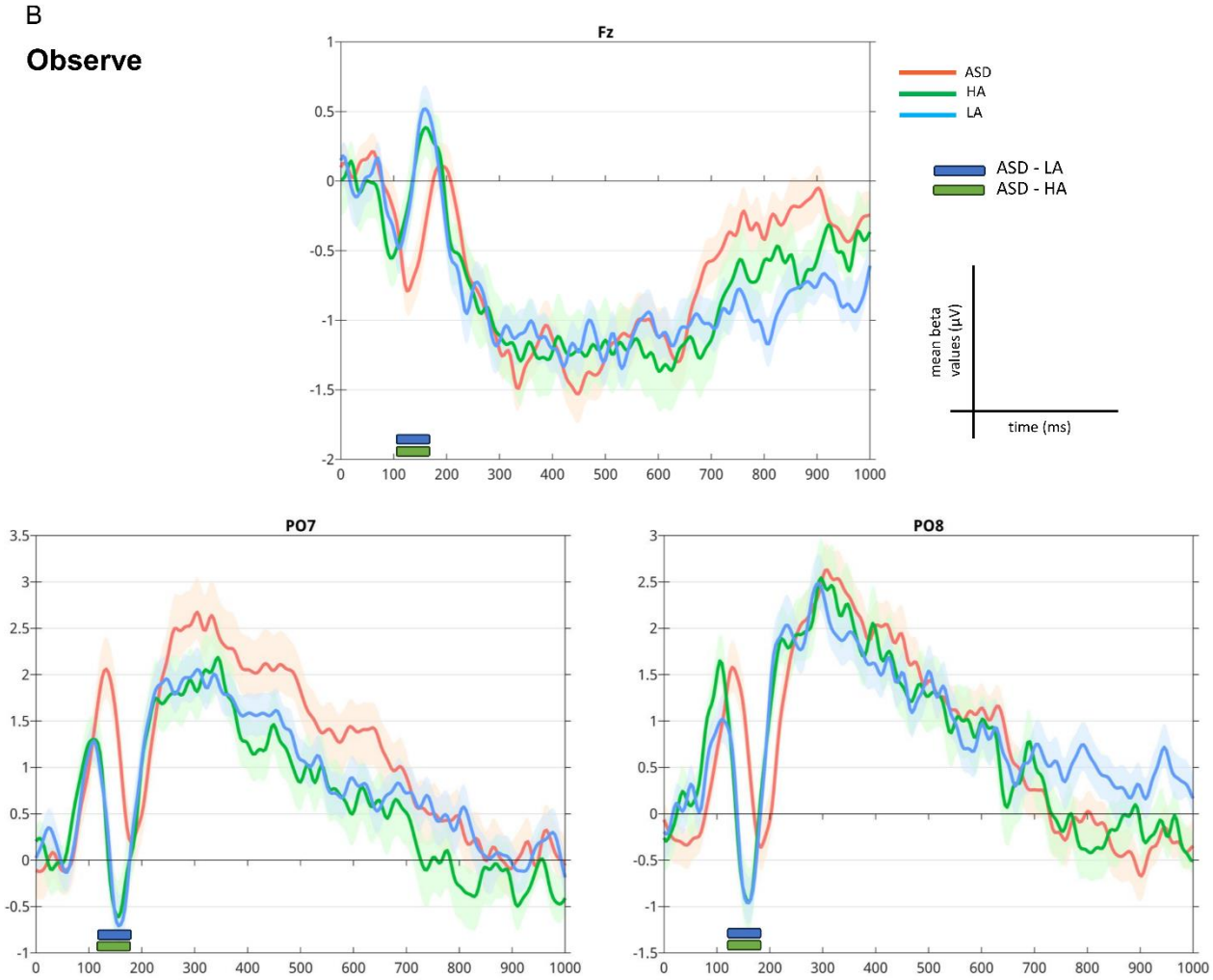


Figure 5.8. Trace plots showing the mean beta values within the ASD (red), HA (green), and LA (light blue) groups over the Fz, PO7, and PO8 electrodes for (A) the emotional condition, and the (B) observe condition. Fz is the most significant electrode of the second cluster within the emotional condition, while PO7 and PO8 are electrodes involved in the first cluster of both emotion and observe conditions and are typically referred to the N170 (references are reported within the text); the right-lateralization is evident for the emotional condition. Blue and green rectangles on the x-axis indicate the significant time windows of the clusters found from the t-tests: ASD-LA (blue), and ASD-HA (green).

5.3.3 ERSP results

The robust two sample t-tests performed on ERSP neural responses revealed no significant differences across groups in terms of μ-alpha suppression in the canonical 8-12 Hz for any of the experimental conditions.

5.4 Discussion

In preceding chapters, I demonstrated how agreeableness influences some of the socio-emotional processes associated with ToM in individuals without specific difficulties. In the current study, I investigated the modulation of personality on the same processes in a sample of young adults with ASD without language and intellectual impairment, adopting the same experimental tasks as in previous EEG studies. In the following I will discuss the results in response to the different research questions outlined in the introduction of this chapter.

As in response to the research question 1, I observed a personality profile tending towards lower scores in our ASD sample, in line with Lodi-Smith's et al. meta-analysis (2019). This profile revealed significantly lower scores in agreeableness, extraversion, conscientiousness, and higher scores in neuroticism compared to a large control group. Notably, I found no differences in openness scores, likely due to the nature of the high-functioning sample, as opposed to the predominantly low-functioning individuals in the published meta-analysis. Importantly, our findings indicate that, referring to the Italian population mean values, the group of ASD individuals did not all fall beyond the average with extreme negative scores; rather, there existed inter-subject variability within each personality trait where only some individuals exhibited extreme scores, similar to healthy subjects. These findings underscore the variability within this population, challenging the notion of equating extremely low personality profiles with this diagnosis and providing practical implications. They also suggest the need for personalized assessments to target individual strengths and weaknesses, enabling tailored interventions and more appropriate social integration pathways based on personality characteristics. Having the same diagnosis does not necessarily imply uniformity among individuals with that diagnosis. These distinctions can have a significant impact on their daily lives and overall well-being.

Moving to research question 2, agreeableness turns out to be a key factor in distinguishing between the ASD and control groups. As hypothesized, ASD subjects overlapped with controls

exhibiting lower agreeableness scores, while differentiating from those with higher agreeableness scores.

This evidence led us to pursue our research question 3, investigating similarities and differences between ASD participants and controls with varying levels of agreeableness in socio-emotional processes underlying ToM explored in our previous EEG studies. Our results indicated no group differences in the detection of social interaction and basic emotions, as measured respectively by the SIT and BM tasks, either at the behavioral or neural level. However, significant differences emerged in the RME task. Here, ASD individuals exhibited lower accuracy in decoding mental states compared to highly agreeable controls, with no such differences observed in relation to less agreeable ones. Neural responses showed differences in ERPs between ASD and both high and less agreeable controls, albeit two distinct effects were found. The first finding pertained to a distinction between groups across all conditions, not limited solely to the mental state decoding. Indeed, this difference was also evident when participants were tasked with assessing the age of presented stimuli or merely observing them. Consequently, it represents an early effect associated with stimulus processing, consistently revealing a reduced and slightly time-shifted ERP amplitude in parieto-occipital electrodes among individuals with ASD in comparison to both high and low agreeableness healthy controls. This effect can be attributed to the well-documented N170 component associated with facial stimuli processing, generally right-lateralized and involving parieto-occipital sites (Feuerriegel et al., 2014; Kappenman et al., 2021; Rossion et al., 2003) which is delayed and/or reduced in ASD individuals (see Kang et al., 2018 for a meta-analysis). A second effect strictly related to the mental state decoding was found, reflecting in behavioral results. Particularly, subjects with ASD exhibited a reduced amplitude of the LSW component - already found involved in this task in the previous chapter where its functional significance was discussed - during mental state decoding compared to both high and low agreeable controls. Notably, the traceplots of the ERPs reveal an interesting pattern: aligning with our prior research and hypotheses, individuals with high levels of agreeableness exhibited a

higher LSW during mental state decoding. Conversely, individuals with ASD showed a smaller amplitude, and those with low levels of agreeableness fell between the former two groups, displaying an intermediate amplitude. The discrepancy between ASDs and individuals with low agreeableness was less pronounced and extended over time as well as anticipated in time compared to that between ASD and subjects with high levels of agreeableness (Figure 5.6; ASD - HA: ~ 450-1000 ms; ASD - LA: ~ 400-600 ms). This could potentially account for the absence of this discrepancy between individuals with ASD and those with lower agreeableness in terms of behavioral outcomes. To sum up, neural differences emerged between ASDs and healthy controls, not attributed to agreeableness but rather linked to facial stimulus processing: ASDs consistently exhibit a diminished N170. Yet, this discrepancy did not seem to hinder performance significantly, which is instead impaired only during mental state decoding due to divergence in the LSW, aligning with the hypotheses of our previous study. Therefore, it emerges that even with the initial challenge posed by the N170, distinguishing ASD from controls, and the subsequent challenge tied to the LSW during mental state decoding, compensatory processes within ASD individuals might let them to achieve similar performance to less agreeable controls although without reaching the level of high agreeable ones. The present findings reinforce the role of agreeableness in these processes.

Regarding the null results in the previous two tasks, two considerations arise. First, individuals with ASD face challenges in social interactions and emotional expression, yet this does not imply an inability to recognize such stimuli. The evident challenges in their social skills do not necessarily signify a comprehensive impairment in the realm of social cognition. This is the rationale behind our approach of dissecting this domain, evaluating social and emotional aspects individually, while also accounting for task complexity. This holds true for ASD individuals, particularly those without language and intellectual impairment who exhibit a certain degree of adaptability to environmental demands, including relational aspects (Kamp-Becker et al., 2010; Liss et al., 2001; Renty et al., 2006). Second, as regards the recognition of basic emotions through

point-light bodily stimuli (BM task), this stimuli's extended duration and motion might contribute to ASD individuals' compensatory response, providing them the time to adjust to possible difficulties. A prospective avenue of exploration involves investigating an early time window starting from stimulus presentation within the BM task.

Noteworthy, our tasks were not designed to weigh on reaction times, and this is probably the reason why we failed to observe behavioral differences with this measure. Exploring swifter tasks that demand instantaneous responses could provide an interesting avenue to uncover distinctions, possibly not in terms of accuracy, but rather regarding delayed responses among individuals with ASD that could represent further proof of active compensatory mechanisms.

Finally, to fully characterize the influence of personality on the examined social-emotional processes within this population, within-subjects analysis should be performed to analyze whether specific autistic traits such as AQ and EQ (as described in the general introduction of the thesis) could moderate the behavioral and neural responses within the ASD group, potentially accounting for variability not captured by the Big Five traits. I am currently in the validation stage for these analyses.

5.5 Conclusion

The study I reported in this chapter delved into the complex interplay between personality and socio-emotional processes in individuals with ASD. We expanded our understanding beyond the borders of a personality profile, often associated with ASD without language and intellectual impairment, by showing that this is not consistent as there is an inter-subject variability within this population. Our investigation underscored that while young adults with this diagnosis may exhibit lower scores in the Big Five Questionnaire, they do not uniformly fall into the extreme ends of all personality dimensions. This evidence provides important implications by emphasizing the need for tailored interventions and support strategies that acknowledge individual strengths and weaknesses promoting positive outcomes and effective social integration. These findings further

highlighted the role of agreeableness in shaping socio-emotional processing, particularly in the context of ToM tasks. Notably, ASD individuals showcased a capacity for compensatory mechanisms, achieving behavioral performance akin to less agreeable healthy controls despite differences in neural responses. Overall, this study offers valuable insights into the multifaceted nature of personality within the ASD context.

CHAPTER 6

General discussion

Personality influences a wide variety of emotional, motivational, and cognitive states, with implications for outcomes in our occupational, relational, and daily life spheres. As such, personality plays a critical role in a wide range of research fields that aim at explaining the complexity of human nature. In the present thesis, I focused on agreeableness, one of the personality traits described within the Five Factor Model (FFM; Costa and McCrae, 1992a; Goldberg, 1990). This trait has been linked to socio-cognitive abilities such as empathy and Theory of Mind (ToM; Allen and DeYoung, 2017; Nettle and Liddle, 2008). Similarly to other personality traits, agreeableness spans across a continuum. Individuals characterized by high levels of agreeableness tend to demonstrate qualities of altruism, cooperation, and empathy with the viewpoints of others. Conversely, those with lower levels of agreeableness might display challenges in social interactions, self-centered tendencies and, in some cases, even aggressive behaviors (Allen and DeYoung, 2017; DeYoung, 2010; Graziano et al., 2007; John et al., 2008).

The experimental work included in this thesis was designed to establish the extent to which agreeableness influences social and emotional neural mechanisms associated with ToM in the general population and in young adults with Autism Spectrum Disorder (ASD) without language and intellectual impairments. To reach this goal, I carried out four studies, each with the aim of targeting one specific neural process underlying ToM and involving computer-based tasks combined with neuroscientific methodologies. Given that in personality neuroscience the number of subjects involved is often inadequate, in the studies reported in my thesis I pre-selected healthy subjects based on their personality scores to increase in-sample variance. The preselection was made by means of the Big Five Inventory (Ubbiali et al., 2017), a self-report questionnaire associated with the FFM validated within the Italian population. Across my studies, agreeableness consistently emerged as a pivotal trait influencing social cognitive processes.

More specifically, in Chapter 2, I report an fMRI study investigating whether variations in the way social information is encoded in the brain are associated with differences in individuals' levels of agreeableness. This was accomplished by using a classic ToM task in which social and

socially-neutral animations were shown (Castelli et al., 2020). I adopted a multivariate representational similarity approach to capture within single individuals the activation pattern similarity of social and non-social content and I tested its association with the agreeableness trait in a hypothesis-driven manner. Results showed individuals with lower agreeableness exhibited similar activation patterns for both social and non-social content, while those with higher agreeableness displayed more distinct patterns. Importantly, this connection between agreeableness and the similarity of neural activation patterns for social and neutral content was statistically significant specifically within the dmPFC - a brain region typically associated with ToM abilities (Molenberghs et al., 2016; Sabbagh and Bowman, 2018; Saxe and Baron-Cohen, 2006; Schurz et al., 2014, 2021). This finding suggests that individual differences in agreeableness are associated with differences in social information processing. However, due to the complexity of the stimuli, results did not allow us to disentangle the different dimensions of social information that might be distinctly represented in individuals with different levels of agreeableness.

To overcome this limitation, in Study 2 (Chapter 3) I targeted the detection of social interactions, a pivotal aspect of social information processing. ERP analyses were used to identify the stage of the neural processing at which agreeableness kicks in, in order to distinguish the impact on early, perceptual processes from slower, inferential processing. Results indicated that agreeableness modulates the neural processes underlying the detection of social interaction. Behaviorally, this modulation was reflected as a higher sensitivity to detect social interactions in more agreeable males only. At the neural level, agreeableness modulated ERP signals in a relatively early time window for the entire sample and females (~200 ms), while in males, this modulation was observed across multiple time windows (~150 ms, ~380 ms, ~720 ms) throughout the stimulus presentation also predicting their performance. Despite timing differences, the effect direction was consistent: less agreeable individuals displayed reduced responsivity to social stimuli compared to non-social stimuli, while more agreeable subjects showed the opposite pattern. Source estimation procedures suggested the involvement of the right fronto-parietal areas

in the association between agreeableness and the ERP effects. Since this region represents a pivotal hub of the dorsal attention network, playing a crucial role in selecting relevant sensory information (Corbetta and Shulman, 2002; Ptak 2012; Sturm et al., 2016), these findings suggest that agreeableness can exert a modulatory influence on attentional processes, enhancing the focus and the priority on social stimuli.

What still needed to be addressed was whether the modulation of agreeableness occurred for both social and emotional stimuli processing, and whether it was affected by the complexity of the task and of the neural processes at study. Therefore, in Study 3 (Chapter 4), I investigated the modulation of agreeableness in decoding emotional cues by means of two tasks with different degrees of complexity spanning from the perception of basic emotional actions to the decoding of complex mental states. The influence of agreeableness on emotional information processing was found to be tied to the complexity of the tasks employed. While basic emotional perception processes did not reveal to be significantly modulated by agreeableness, higher-order processes such as mental state decoding did. Specifically, results showed that, relative to the control condition, higher agreeableness scores were associated with better accuracy in recognizing mental states, and with a modulation of ERP signals in a later time window, after the end of the stimulus presentation (~ 700-900 ms). Interestingly, sex differences similar to those of Study 2 were found since the modulation of agreeableness on the neural response in females was negligible while a later and larger cluster emerged in the male sample. Both as a whole and in the male sample, higher agreeableness scores exhibited a significantly higher difference between the ERPs associated with the emotional and control conditions, particularly driven by a heightened neural responsivity to emotional trials compared to non-emotional trials, which was reduced in less agreeable participants. This was indicated by an increased late slow wave (LSW) observed during mental state decoding. While the LSW has been typically associated with cognitive aspects of ToM (see Sabbagh, 2013 for a review), this study highlights its involvement in affective aspects as well. Furthermore, the LSW has been associated with emotional encoding by facilitating the integration

of emotional and cognitive information, emotional intensity (Hajcak et al., 2009; Olofsson et al., 2008; Schupp et al., 2000; Tesar et al., 2020; Yick et al., 2015; see Hajcak et al., 2010 for a review), and also used as an index of abnormal emotional responding (Dennis and Hajcak, 2009; Horan et al., 2010; Marissen et al., 2010). Notably, the vmPFC's involvement, identified through source estimation procedures, aligns with this interpretation since its connections with the amygdala and the temporoparietal junction are central to affective ToM abilities, including decoding others' emotional states. These findings suggest that the neural mechanisms underlying agreeableness might be more nuanced than previously believed. Furthermore, it underscores the importance of task selection and it emphasizes the need for tasks that can effectively tease apart the subtle neural modulations associated with individual differences in personality.

The abovementioned studies, hopefully, demonstrate that personality exerts an influence on how our brain encodes social stimuli. Specifically, I showed that agreeableness shapes social and emotional information processing associated with ToM, including the detection of social interactions and the decoding of complex mental states. This modulation goes through focused attention and higher responsivity to social stimuli, observed in particular at later latencies (Chapters 2-4). These results align with previous research demonstrating that social stimuli elicit larger ERP amplitudes compared to socially neutral stimuli (Ding et al., 2017; Kosonogov et al., 2019; Olofsson et al., 2008; Schacht and Vrtička, 2018), and this effect is smaller in populations with lower social skills and ToM difficulties (Benning et al., 2016; Cox et al., 2015; Jeste and Nelson, 2009; Leung et al., 2021; Stavropoulos and Carver, 2014). However, agreeableness modulation has been found to depend on the complexity of the neural processes considered. We could assume that lower-order processes might not be involved in personality, since they are more automatic and require fewer cognitive resources. These findings constitute an important advance in this particular research field.

Furthermore, this modulation seems to depend also on sex features. Indeed, in both EEG Studies 1-2, the impact of agreeableness was more prolonged and extensive among males than

females, suggesting that personality may play a role in processing social stimuli differently in males and in females. This result is in keeping with the extant literature on sex differences in the social domain whereby females are reported as being more socially oriented, often described as possessing inherent sensitivity to non-verbal social and emotional cues and being quick learners of such cues (Baron-Cohen et al., 2005; Hall and Halberstadt, 1981; Paletta et al., 2022; Pavlova et al., 2010; Proverbio, 2008, 2021; Stake and Eisele, 2010). This supposedly inherent sensitivity then would likely contribute to the lacking impact of personality on social-emotional processing observed in females. Therefore, our findings renew the critical importance of considering the sex of participants in studies investigating the neural mechanisms underlying inter-individual variations in personality traits.

While in the Studies (1-3) participants without specific challenges were recruited, in the last Study (4), I considered individuals with ASD without language and intellectual impairments. This investigation aimed at exploring the personality profile of young adults with this diagnosis and whether agreeableness might differentially modulate the same neural processes investigated in the previous two EEG studies in this population. Results underscored that while young adults with this diagnosis may exhibit lower scores in all the five personality traits from the FFM - except for openness - they do not uniformly fall into the extreme ends of all personality dimensions: on the contrary considerable variability persisted, reflecting a heterogeneity in personality traits even within the ASD group. Notably, agreeableness trait in particular turned out to be a key factor in distinguishing between the ASD and control groups. Indeed, ASD subjects overlapped with controls exhibiting lower agreeableness scores, while they differentiated from those with higher agreeableness scores. This evidence also emerged during mental state decoding, where individuals within the ASD spectrum exhibited comparable accuracy to less agreeable controls, even though their accuracy was significantly lower compared to controls with higher levels of agreeableness. At the neural level, and consistently to the literature about face processing in ASD (see Black et al., 2017, Samaey et al., 2020 for reviews), ERP results showed a delayed and reduced N170 during

stimulus processing for ASD participants compared with controls (Feuerriegel et al., 2014; Kappenman et al., 2021; Rossion et al., 2003; see Kang et al., 2018 for a meta-analysis). In addition, a reduced LSW only when decoding mental states was found. Specifically, the difference was more extensive and sustained over time for ASD vs. more agreeable controls than ASD vs. less agreeable ones. Overall, it emerged that even with the initial challenge posed by the N170, distinguishing ASD from controls, and the subsequent challenge tied to the LSW during mental state decoding, compensatory processes within ASD individuals might let them to achieve similar performance to less agreeable controls although without reaching the level of high agreeable ones. The present findings reinforce the role of agreeableness in these processes and align with results from previous EEG studies where agreeableness was found to improve behavioral performance along with higher neural responses during mental state decoding. Notably, no group differences emerged in the detection of social interaction and basic emotions either at the behavioral or neural level. The findings suggest that although individuals with ASD encounter difficulties in engaging in social interactions and conveying emotions, it should not be assumed that they lack the capacity to identify such stimuli. The apparent challenges in their social adeptness should not necessarily be interpreted as a comprehensive impairment in the sphere of social cognition. This is the rationale behind our methodology of breaking down this domain, assessing social and emotional facets individually, all while considering task complexity. This is particularly relevant for individuals with ASD, especially those without language and cognitive deficits, as they display a certain level of adaptability to environmental demands, including those relating to social and relational dimensions (Kamp-Becker et al., 2010; Liss et al., 2001; Renty et al., 2006). The results of this final study underscore the adaptive nature of the brain and highlight the importance of considering personality traits when studying clinical populations. Currently, there are different methodologies and models investigating autism and hopefully, the results provided from all these different approaches might lead to a better understanding of the critical mechanisms underlying the spectrum.

To sum up, the impact of personality on social-emotional processes of ToM unfolds at late latencies, becoming more evident post-stimulus presentation fostering the differentiation between social and neutral stimuli, and leading to a higher neural response towards the former. The observed influence appears to be contingent upon the complexity of the neural mechanisms investigated possibly because low-order processes remain unaltered by personality, given their automatic nature and minimal cognitive demands. Moreover, this modulation plays a role in processing social stimuli differently in males and females. Finally, personality turns out to be an important factor to consider even in clinical populations such as autism.

With my thesis I aimed at integrating models of social cognition with the Five Factor personality framework while also considering the mediating role of biological factors, such as sex, to create a cohesive narrative. My findings contribute to this integrative process as they allow to account for the influence of individual differences which to date have been poorly investigated. As I reminded in the introduction of the thesis, ToM is a wide domain, which involves a different processes, as well delineated by the hierarchical model of social cognition proposed by Schurz and colleagues (2014). Accordingly, in my investigation, I sought to deconstruct ToM, by focusing on specific mental processes involved in decoding social and emotional cues. Although all these processes fall predominantly within the intermediate and affective ToM (Schurt et al., 2014), they differ from one another, as they involve distinct social and emotional facets, and they also vary in the levels of complexity. The model I suggest encompasses personality, specifically agreeableness, and its modulatory influence on higher order mental processes for decoding social and emotional cues such as the detection of social interactions and the decoding of complex mental states. Moreover, as I observed sex differences, my thesis clearly suggests the importance of incorporating the biological underpinnings related to sex in the model. While the existing literature occasionally presents mixed findings, it frequently highlights that females tend to exhibit greater agreeableness (Costa t al., 2001; Soto and Tackett 2015; Tackett et al., 2019). On the other hand, within emotional and social processing studies, females are often reported as excelling or

displaying heightened sensitivity to these stimuli (Baron-Cohen et al., 2005; Hall and Halberstadt, 1981; Paletta et al., 2022; Pavlova et al., 2010; Proverbio, 2008, 2021; Stake and Eisele, 2010). Consequently, it is apparent that sex holds sway over both personality traits and ToM processes. However, my thesis has unveiled interesting nuances. When I controlled for sex as a factor, the influence of agreeableness on diverges significantly between males and females, compelling us to introduce the mediating role of sex in the complex relationship between agreeableness and ToM processes. Figure 6.1 shows a graphical representation of the proposed model.

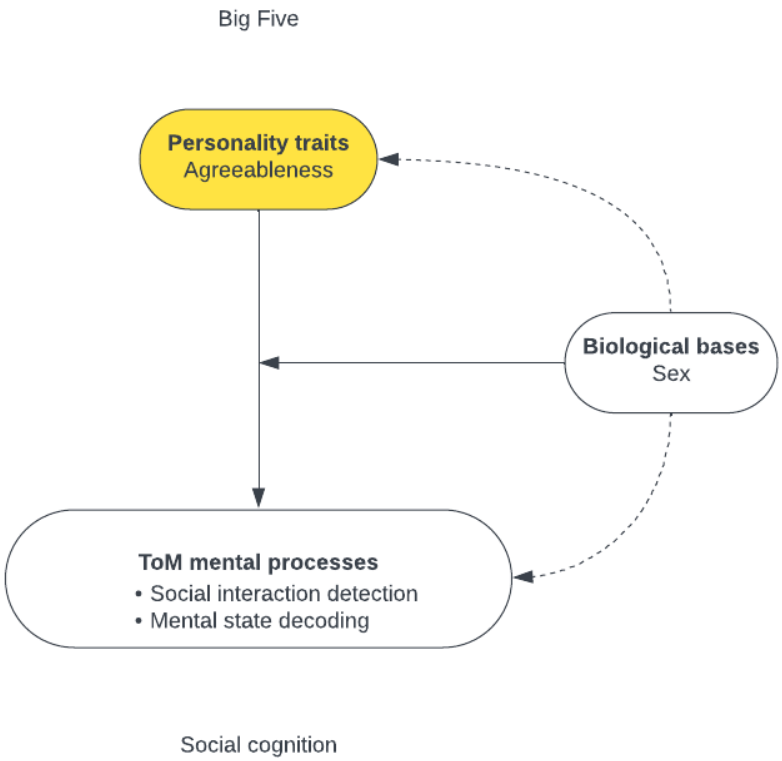


Figure 6.1. Graphical representation of the model proposed. Dashed lines represent relationships already establisehd outside the model.

Several promising avenues emerge from the current thesis or further inquiry and advancement. First, the lack of comparative studies necessitates the replication of these findings, acting as a cornerstone for establishing their robustness and reliability. Therefore, the model proposed is necessarily thought to be preliminary and requires validation. Furthermore, to deepen

our understanding, a more refined exploration is warranted, dissecting diverse forms of social interaction such as joint attention and mutual gaze (Guellaï et al., 2020; Mundy, 2018), together with parsing distinct emotional categories, particularly within basic emotions (Celeghin et al., 2017). This approach could unveil nuanced variations that might be obscured within broader assessments.

Another potential trajectory would be delving into more advanced components of ToM, such as those influencing its cognitive domain like *false belief* comprehension (Sabbagh and Bowman, 2018; Schurz et al., 2021). To foster a comprehensive grasp, expanding research using both EEG and fMRI is crucial, as each method contributes unique insights into the complex interplay between neural processes and behavioral manifestations. The investigation could further encompass disentangling the aspects of agreeableness, namely *compassion* and *politeness* (Allen and DeYoung, 2017), alongside the core trait employing tools like the Big Five Aspects Scale (DeYoung et al., 2007) together with the Big Five Inventory (Ubbiali et al., 2013).

In my thesis I considered the clinical population of individuals within the ASD, however it would be valuable to extend this approach to diverse patient groups. For instance, individuals with Williams syndrome are characterized, differently from those with ASD, by a hypersocial disposition (see Barak and Feng, 2016 for a review); however we do not know to which extent personality plays a role in this clinical population. Furthermore, future explorations within clinical populations should also take into account potential interindividual differences based on sex.

Overall the findings in this thesis have broader implications for the field of personality and cognitive neuroscience, as they demonstrate the significance of accounting for personality traits as modulating factors in neural processing, also by including sex as a valuable variable. Moreover, I believe that this thesis can provide a valuable contribution to the identification of the neural underpinnings of personality, thanks to the methodology employed. Indeed, I used different neuroimaging techniques, along with a range of analyses methods including multivariate and univariate analyses with robust statistical techniques, and I encompassed a spectrum from carefully

chosen samples of healthy individuals to clinical ones.

As agreeableness is closely connected to individual differences in interpersonal behavior (DeYoung et al., 2013; Graziano & Eisenberg, 1997; Koole et al., 2001), a deeper comprehension of it and the cognitive mechanisms that interact with it, encompassing social cognitive functions such as ToM, can potentially improve our predictive capabilities and deepen our insights into the diversity of interpersonal behaviors and relationship dynamics.

Finally, the insights garnered from the study involving individuals with ASD underline the importance of probing further into the role that personality occupies within the brain's compensatory mechanisms and the need for a nuanced approach in studying clinical populations. I am strongly in favour of going beyond diagnostic labels in order to formulate effective treatments and tailored interventions, that consider the unique strengths and challenges that each patient might possess, even within the context of the same diagnosis. To reach this goal embedding personality assessment within the clinical realm holds promise since it offers a comprehensive lens across individuals. All together these future avenues would help in bridging the gap between research and clinical practice.

In conclusion, inter-individual differences wield significant relevance and warrant constant consideration not only from researchers engaged in the realm of personality neuroscience since their exploration entails wide implications across economic, occupational, educational, and clinical domains, driven by the correlation of personality with a multitude of neural processes. Therefore, when delving into diverse and targeted cognitive processes, regardless of their nature, and averaging across participant data and neural signals to construct a comprehensive overview of the process's overall functioning, I believe it is important to acknowledge that such functioning retains its individualistic variance among subjects.

REFERENCES

Abram, S. V., Wisner, K. M., Fox, J. M., Barch, D. M., Wang, L., Csernansky, J. G., MacDonald III, A. W., & Smith, M. J. (2016). Fronto-temporal connectivity predicts cognitive empathy deficits and experiential negative symptoms in schizophrenia. *Human Brain Mapping*, *38*(3), 1111–1124. https://doi.org/10.1002/hbm. 23439

Adelstein, J. S., Shehzad, Z., Mennes, M., Deyoung, C. G., Zuo, X. N., Kelly, C., Margulies, D. S., Bloomfield, A., Gray, J. R., Castellanos, F. X., & Milham, M. P. (2011). Personality is reflected in the brain's intrinsic functional architecture. *PloS one*, *6*(11), e27633. https://doi.org/10.1371/journal.pone.0027633

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences*, *3*(12), 469–479. https://doi.org/10.1016/S1364-6613(99)01399-6

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *3*(3), 165–178. https://doi.org/10.1038/nrn1056

Agam, Y., & Sekuler, R. (2007). Interactions between working memory and visual perception: An ERP/EEG study. *NeuroImage*, *36*(3), 933–942. https://doi.org/10.1016/j.NeuroImage.2007.04.014

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., & Wenderoth, N. (2011). Action and emotion recognition from point light displays: An investigation of gender differences. *PLoS ONE*, *6*(6). https://doi.org/10.1371/journal.pone.0020989

Allen, T. A., & DeYoung, C. G. (2017). Personality Neuroscience and the Five Factor Model. In T. A. Widiger (Ed.). *The Oxford Handbook of the Five Factor Model* (pp. 319-349). Oxford University Press. https://doi.org/10.1093/oxfordhb

Allen, T. A., Rueter, A. R., Abram, S. V, Brown, J. S., & Deyoung, C. G. (2017). Personality and Neural Correlates of Mentalizing Ability. *European Journal of Personality*, *31*(6), 599–613. https://doi.org/10.1002/per.2133

American Psychiatric Association (2013) *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Pub.

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*, *20*(2), 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

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

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*(6), 717–746. https://doi.org/10.1068/p5096

Atkinson, A. P., Vuong, Q. C., & Smithson, H. E. (2012). Modulation of the face- and body-selective visual regions by the motion and emotion of point-light face and body stimuli. *Neuroimage*, *59*(2), 1700–1712. https://doi.org/10.1016/j.neuroimage.2011.08.073

Austin, E. J. (2005) Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Differences, 38*(2), 451–460. https://doi.org/10.1016/j.paid.2004.04.022

Auyeung, B., Baron-Cohen, S., Chapman, E., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Foetal testosterone and the child systemizing quotient. *European Journal of Endocrinology*, *155*(suppl_1), S123–S130. https://doi.org/10.1530/eje.1.02260

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labelling of elderly and neurodegenerative brain. Med. Image Anal. *12*(1), 26–41. https://doi.org/10.1016/J.MEDIA.2007.06.004

Avinun, R., Israel, S., Knodt, A. R., & Hariri, A. R. (2020). Little evidence for associations between the big five personality traits and variability in brain gray or white matter. *NeuroImage*, *220*, 117092. https://doi.org/10.1016/J.NEUROIMAGE.2020.117092

Bachevalier, J., & Loveland, K. A. (2006). The orbitofrontal-amygdala circuit and self-regulation of social-emotional behavior in autism. *Neuroscience and biobehavioral reviews*, *30*(1), 97–117. https://doi.org/10.1016/j.neubiorev.2005.07.002

Barak, B., & Feng, G. (2016). Neurobiology of social behavior abnormalities in autism and Williams syndrome. *Nature neuroscience*, *19*(6), 647–655. https://doi.org/10.1038/nn.4276

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hart- ley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A.Z., & Van Essen, D. C. (2013). Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189. https://doi.org/10.1016/J.NEUROIMAGE.2013.05.033

Barnhill, G. P. (2007). Outcomes in adults with Asperger syndrome. *Focus on Autism and Other Developmental Disabilities*, *22*(2), 116–126. https://doi.org/10.1177/10883576070220020301

Baron-Cohen, S. (1995). *Mindblindness: an Essay on Autism and Theory of Mind*. MIT Press.

Baron-Cohen S. (2009). Autism: the empathizing-systemizing (E-S) theory. *Annals of the New York Academy of Sciences*, *1156*, 68–80. https://doi.org/10.1111/j.1749-6632.2009.04467.x

Baron-Cohen, S. (2011). *Zero Degrees of Empathy: A New Theory of Human Cruelty*. Allen Lane.

Baron-Cohen, S., Auyeung, B., Nørgaard-Pedersen, B., Hougaard, D. M., Abdallah, M. W., Melgaard, L., Cohen, A. S., Chakrabarti, B., Ruta, L., & Lombardo, M. V. (2015). Elevated fetal steroidogenic activity in autism. *Molecular psychiatry*, *20*(3), 369–376. https://doi.org/10.1038/mp.2014.48

Baron-Cohen, S., Cassidy, S., Auyeung, B., Allison, C., Achoukhi, M., Robertson, S., Pohl, A., & Lai, M. C. (2014). Attenuation of typical sex differences in 800 adults with autism vs. 3,900 controls. *PloS one*, *9*(7), e102251. https://doi.org/10.1371/journal.pone.0102251

Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, M. K. (2005). Sex differences in the brain: Implications for explaining autism. *Science*, *310*(5749), 819–823. https://doi.org/10.1126/science.1115455

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., & Williams, S. C. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *The European journal of neuroscience*, *11*(6), 1891–1898. https://doi.org/10.1046/j.1460-9568.1999.00621.x

Baron-Cohen, S., Tager-Flusberg, H., & Lombardo, M. (2013). *Understanding other minds: Perspectives from developmental social neuroscience*. OUP Oxford.

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, and allied disciplines*, *42*(2), 241–251.

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends in cognitive sciences*, *15*(7), 301–309. https://doi.org/10.1016/j.tics.2011.05.002

Bayer, M., Rossi, V., Vanlessen, N., Grass, A., Schacht, A., & Pourtois, G. (2017). Independent effects of motivation and spatial attention in the human visual cortex. *Social Cognitive and Affective Neuroscience*, *12*(1), 146–156. https://doi.org/10.1093/scan/nsw162

Beacher, F. D., Minati, L., Baron-Cohen, S., Lombardo, M. V., Lai, M. C., Gray, M. A., Harrison, N. A., & Critchley, H. D. (2012). Autism attenuates sex differences in brain structure: a combined voxel-based morphometry and diffusion tensor imaging study. *AJNR. American journal of neuroradiology*, *33*(1), 83–89. https://doi.org/10.3174/ajnr.A2880

Beacher, F. D., Radulescu, E., Minati, L., Baron-Cohen, S., Lombardo, M. V., Lai, M. C., Walker, A., Howard, D., Gray, M. A., Harrison, N. A., & Critchley, H. D. (2012). Sex differences and autism: brain function during verbal fluency and mental rotation. *PloS one*, *7*(6), e38355. https://doi.org/10.1371/journal.pone.0038355

Beall, P.M., Moody, E.J., McIntosh, D.N., Hepburn, S.L., Reed, C.L. (2008). Rapid facial reactions to emotional facial expressions in typically developing children and children with autism spectrum disorder. *Journal of Experimental Child Psychology*, *101*(3), 206–23. https://doi.org/10.1016/j.jecp.2008.04.004

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

Benning, S. D., Kovac, M., Campbell, A., et al. (2016). Late Positive Potential ERP Responses to Social and Nonsocial Stimuli in Youth with Autism Spectrum Disorder. *Journal of Autism*

and Developmental Disorders, *46*(9), 3068–3077. https://doi.org/10.1007/s10803-016-2845-y

Bertelsen, N., Landi, I., Bethlehem, R., Seidlitz, J., Busuoli, E. M., Mandelli, V., … Lombardo, M. V. (2021). Imbalanced social-communicative and restricted repetitive behavior subtypes of autism spectrum disorder exhibit different neural circuitry. *Communications biology*, *4*(1), 574. https://doi.org/10.1038/s42003-021-02015-2

Best, C. S., Moffat, V. J., Power, M. J., Owens, D. G. C., & Johnstone, E. C. (2008). The boundaries of the cognitive phenotype of autism: Theory of mind, central coherence and ambiguous figure perception in young people with autistic traits. *Journal of Autism and Developmental Disorders, 38*(5), 840–847. https://doi.org/10.1007/ s10803-007-0451-8

Bilker, W. B., Hansen, J.A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, *19*(3), 354–369. https://doi.org/10.1177/1073191112446655

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

Bjørnebekk, A., Fjell, A. M., Walhovd, K. B., Grydeland, H., Torgersen, S., & Westlye, L. T. (2013). Neuronal correlates of the five factor model (FFM) of human personality: Multimodal imaging in a large healthy sample. *NeuroImage, 65*, 194–208. https://doi.org/10.1016/j.neuroimage.2012. 10.009

Black, M. H., Chen, N. T. M., Iyer, K. K., Lipp, O. V., Bölte, S., Falkmer, M., Tan, T., & Girdler, S. (2017). Mechanisms of facial emotion recognition in autism spectrum disorders: Insights from eye tracking and electroencephalography. *Neuroscience and biobehavioral reviews, 80*,

488–515. https://doi.org/10.1016/j.neubiorev.2017.06.016

Blain, S. D., Peterman, J. S., & Park, S. (2017). Subtle cues missed: Impaired perception of emotion from gait in relation to schizotypy and autism spectrum traits. *Schizophrenia research*, *183*, 157–160. https://doi.org/10.1016/j.schres.2016.11.003

Bora, E. (2020). Theory of mind and schizotypy: A meta-analysis. *Schizophrenia Research, 222*, 97–103. https://doi.org/10.1016/j. schres.2020.04.024

Brighenti, S., Schintu, S., Liloia, D., & Keller, R. (2018). Neuropsychological aspects of Asperger Syndrome in adults: a review. *Neuropsychological Trends*, *24*, 63–95 http://dx.doi.org/10.7358/neur-2018-024-brig

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science (New York, N.Y.)*, *307*(5712), 1118–1121. https://doi.org/10.1126/science.1105783

Brownell, H., Griffin, R., Winner, E., Friedman, O., & Happe´, F. (2000). Cerebral lateralization and theory of mind. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.). *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience* (pp. 311–338). Oxford University Press.

Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mind--evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience and biobehavioral reviews*, *30*(4), 437–455. https://doi.org/10.1016/j.neubiorev.2005.08.001

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science*, *15*, 118–121. https://doi.org/10.1111/j.0963-7214.2006.00419.x

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, *7*(6), 477–484. https://doi.org/10.1038/nrn1909

Cai, H., Zhu, J., & Yu, Y. (2020). Robust prediction of individual personality from brain functional connectome. *Social Cognitive and Affective Neuroscience*, *15*(3), 359–369. https://doi.org/10.1093/scan/nsaa044

Canli, T., Sivers, H., Whitfield, S. L., Gotlib, I. H., & Gabrieli, J. D. E. (2002). Amygdala response to happy faces as a function of extraversion. *Science*, *296*(5576), 2191. https://doi.org/10.1126/science.1068749

Canli, T., Zhao, Z., Desmond, J. E., Kang, E., Gross, J., & Gabrieli, J. D. E. (2001). An fMRI study of personality influences on brain reactivity to emotional stimuli. *Behavioral Neuroscience*, *115*(1), 33–42. https://doi.org/10.1037/0735-7044.115.1.33

Caprara, G. V., Alessandri, G., Giunta, L. D., Panerai, L., &Eisenberg, N. (2010). The contribution of agreeableness and self-efficacy beliefs to prosociality. *European Journal of Personality*, *24*(1), 36–55. https://doi.org/10.1002/per.739

Carlson, T., Goddard, E., Kaplan, D. M., Klein, C., & Ritchie, J. B. (2018). Ghosts in machine learning for cognitive neuroscience: moving from data to theory. *NeuroImage*, *180*, 88–100. https://doi.org/10.1016/J.NEUROIMAGE.2017.08.019

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns. *NeuroImage*, *12*(3), 314–325. https://doi.org/10.1006/NIMG.2000.0612

Celeghin, A., Diano, M., Bagnis, A., Viola, M., & Tamietto, M. (2017). Basic Emotions in Human Neuroscience: Neuroimaging and Beyond. *Frontiers in psychology*, *8*, 1432. https://doi.org/10.3389/fpsyg.2017.01432

Ceponiene, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., & Yaguchi, K. (2003). Speech-sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5567–5572. https://doi.org/10.1073/pnas.0835631100

Chakrabarti, B., Bullmore, E., & Baron-Cohen, S. (2006). Empathizing with basic emotions: common and discrete neural substrates. *Social neuroscience*, *1*(3-4), 364–384. https://doi.org/10.1080/17470910601041317

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: evidence from the empathy quotient (EQ) and the "reading the mind in the eyes" test. *Social neuroscience*, *1*(2), 135–148. https://doi.org/10.1080/17470910600992239

Chen, Y., & Canli, T. (2022). "Nothing to see here": No structural brain differences as a function of the Big Five personality traits from a systematic review and meta-analysis. *Personality Neuroscience, 5*, E8. https://doi.org/10.1017/pen.2021.5

Cherskov, A., Pohl, A., Allison, C., Zhang, H., Payne, R. A., & Baron-Cohen, S. (2018). Polycystic ovary syndrome and autism: A test of the prenatal sex steroid theory. *Translational psychiatry*, *8*(1), 136. https://doi.org/10.1038/s41398-018-0186-7

Chevallier, C., Huguet, P., Happé, F., George, N., & Conty, L. (2013). Salient social cues are prioritized in autism spectrum disorders despite overall decrease in social attention. *Journal of Autism and Developmental Disorders*, *43*(7), 1642–1651. https://doi.org/10.1007/s10803-012-1710-x

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The Social Motivation Theory of Autism. *Trends in Cognitive Sciences*, *16*(4), 231–239.

https://doi.org/10.1016/j.tics.2012.02.007

Coben, R., Chabot, R. J., & Hirshberg, L. (2013). EEG analyses in the assessment of autistic disorders. In M., Casanova, A., El-Baz, & J., Suri (Eds). *Imaging the brain in autism* (pp. 349-370). New York: Springer. https://doi.org/10.1007/978-1-4614-6843-1_12

Codispoti, M., Ferrari, V., & Bradley, M. M. (2007). Repetition and Event-related Potentials: Distinguishing Early and Late Processes in Affective Picture Perception. *Journal of Cognitive Neuroscience*, *19*(4), 577–586. https://doi.org/10.1162/jocn.2007.19.4.577

Cohen, M. X., Young, J., Baek, J.-M., Kessler, C., & Ranganath, C. (2005). Individual differences in extraversion and dopamine genetics predict neural reward responses. *Cognitive Brain Research, 25*(3), 851–861. https://doi.org/10.1016/j.cogbrainres.2005.09.018

Cools, R., Calder, A. J., Lawrence, A. D., Clark, L., Bullmore, E., & Robbins, T. W. (2005). Individual differences in threat sensitivity predict serotonergic modulation of amygdala response to fearful faces. *Psychopharmacology*, *180*(4), 670–679. https://doi.org/10.1007/s00213-005-2215-5

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental psychology*, *34*(1), 28–38. https://doi.org/10.1037/0012-1649.34.1.28

Corr, P. J., DeYoung, C. G., & McNaughton, N. (2013). Motivation and personality: a neuropsychological perspective. *Social and Personality Psychology Compass, 7*(3), 158–175. https://doi.org/10.1111/spc3.12016

Corradi-Dell'Acqua, C., Hofstetter, C., & Vuilleumier, P. (2014). Cognitive and affective theory

of mind share the same local patterns of activity in posterior temporal but not medial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *9*(8), 1175–1184. https://doi.org/10.1093/scan/nst097

Costa, P. T., & McCrae, R. R. (1992a). Four ways five factors are basic. *Personality and Individual Differences*, *13*(6), 653–665. https://doi.org/10.1016/0191-8869(92)90236-I

Costa, P. T., McCrae, R. R. (1992b). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, *6*(4), 343–359. https://doi.org/10.1521/pedi.1992.6.4.343

Costa, P. T., McCrae, R. R., & Löckenhoff, C. E. (2019). Personality Across the Life Span. *Annual review of psychology*, *70*, 423–448. https://doi.org/10.1146/annurev-psych-010418-103244

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of personality and social psychology*, *81*(2), 322–331. https://doi.org/10.1037/0022-3514.81.2.322

Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance NeuroImages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Cox, A., Kohls, G., Naples, A. J., Mukerji, C. E., Coffman, M. C., Rutherford, H. J. V, Mayes, L. C., & Mcpartland, J. C. (2015). Diminished social reward anticipation in the broad autism phenotype as revealed by event-related brain potentials. *Social Cognitive and Affective Neuroscience*, *10*(10), 1357–1364. https://doi.org/10.1093/scan/nsv024

D'Entremont, B., Hains, S. M. J., Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Develoment, 20*, 569–572. https://doi.org/10.1016/S0163-6383(97)90048-5

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/NIMG.1998.0395

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron*, *26*(1), 55–67. https://doi.org/10.1016/s0896-6273(00)81138-1

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of autism and developmental disorders*, *28*(6), 479–485. https://doi.org/10.1023/a:1026043926488

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology*, *27*(3), 403–424. https://doi.org/10.1207/s15326942dn2703_6

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews. Neuroscience*, *7*(3), 242–9. https://doi.org/10.1038/nrn1872

de Haas, B. (2018). How to enhance the power to detect brain–behavior correlations with limited resources. *Frontiers in Human Neuroscience, 12*, 421. https://doi.org/10.3389/fnhum.2018.00421

De Pauw, S. S., Mervielde, I., Van Leeuwen, K. G., & De Clercq, B. J. (2011). How temperament and personality contribute to the maladjustment of children with autism. *Journal of autism and developmental disorders, 41*(2), 196–212. https://doi.org/10.1007/s10803-010-1043-6

Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental psychology*,

36(4), 511–523.

Deckersbach, T., Miller, K. K., Klibanski, A., Fischman, A., Dougherty, D. D., Blais, M. A., Herzog, D. B., & Rauch, S. L. (2006). Regional cerebral brain metabolism correlates of neuroticism and extraversion. *Depression and anxiety*, *23*(3), 133–138. https://doi.org/10.1002/da.20152

Decuyper, M., De Pauw, S., De Fruyt, F., De Bolle, M., & De Clercq, B. J. (2009). A meta- analysis of psychopathy-, antisocial PD- and FFM associations. *European Journal of Personality, 23*, 531–565. https://doi.org/10.1002/per.729

Del Vecchio, M., Caruana, F., Sartori, I., Pelliccia, V., Zauli, F. M., Lo Russo, G., Rizzolatti, G., & Avanzini, P. (2020). Action execution and action observation elicit mirror responses with the same temporal profile in human SII. *Communications biology*, *3*(1), 80. https://doi.org/10.1038/s42003-020-0793-8

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/J.JNEUMETH.2003.10.009

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, *34*(4), 1443–1449. https://doi.org/10.1016/j.neuroimage.2006.11.004

Dennis, T. A., & Hajcak, G. (2009). The late positive potential: a neurophysiological marker for emotion regulation in children. *Journal of child psychology and psychiatry, and allied disciplines*, *50*(11), 1373–1383. https://doi.org/10.1111/j.1469-7610.2009.02168.x

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences, 22*(3),

491–517. https://doi.org/10.1017/s0140525x99002046

DeYoung, C. G. (2010). Personality Neuroscience and the Biology of Traits. *Social and Personality Psychology Compass*, *4*(12), 1165–1180. https://doi.org/10.1111/j.1751-9004.2010.00327.x

DeYoung, C. G. (2020). Intelligence and personality. In R. J., Sternberg & S. B., Kaufman (Eds.). *The Cambridge handbook of intelligence* (2nd ed., pp. 1011–1047). Cambridge University Press. https://doi.org/10.1017/CBO9780511977244.036

DeYoung, C., & Blain, S. (2020). Personality Neuroscience. In P. Corr & G. Matthews (Eds.). *The Cambridge Handbook of Personality Psychology* (pp. 273–292). Cambridge University Press. https://doi.org/10.1017/9781108264822.026

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of personality and social psychology*, *93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

DeYoung, C. G., Quilty, L. C., Peterson, J. B., Gray, J. R. (2014). Openness to experience, intellect, and cognitive ability. *Journal of Personality Assessment, 96,* 46–52. https://doi.org/10.1080/00223891.2013.806327

DeYoung, C. G., Shamosh, N. A., Green, A. E., Braver, T. S., & Gray, J. R. (2009). Intellect as distinct from Openness: Differences revealed by fMRI of working memory. *Journal of Personality and Social Psychology, 97,* 883–892. https://doi.org/10.1037/a0016615

DeYoung, C. G., Weisberg, Y. J., Quilty, L. C., & Peterson, J. B. (2013). Unifying the aspects of the Big Five, the interpersonal circumplex, and trait affiliation. *Journal of Personality*, *81*(5), 465–475

Di Tella, M., Miti, F., Ardito, R. B., & Adenzato, M. (2020). Social cognition and sex: Are men and women really different? *Personality and Individual Differences, 162*. https://doi.org/10.1016/j.paid.2020.110045

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*(6), 1246–1256. https://doi.org/10.1037//0022-3514.73.6.1246

Ding, R., Li, P., Wang, W., & Luo, W. (2017). Emotion Processing by ERP Combined with Development and Plasticity. *Neural Plasticity, 2017*. https://doi.org/10.1155/2017/5282670

Dittrich, W.H., Troscianko, T., Lea, S.E., Morgan, D. (1996). Perception ofemotion fromdynamic point-light displays represented in dance. *Perception*, *25*(6), 727–38. https://doi.org/10.1068/p250727

Dubois, J., Galdi, P., Han, Y., Paul, L. K., & Adolphs, R. (2018). Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. *Personality Neuroscience, 1*, e6. https://doi.org/10.1017/pen.2018.8

Dungan, J. A., Stepanovic, M., & Young, L. (2016). Theory of mind for processing unexpected events across contexts. *Social Cognitive and Affective Neuroscience, 11*(8), 1183–1192. https://doi.org/10.1093/scan/nsw032

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *Journal of autism and developmental disorders*, *38*(3), 464–473. https://doi.org/10.1007/s10803-007-0486-x

Eisenberger, N. I., Lieberman, M. D., & Satpute, A. B. (2005). Personality from a controlled processing perspective: an fMRI study of neuroticism, extraversion, and self-consciousness. *Cognitive, affective & behavioral neuroscience, 5*(2), 169–181.

https://doi.org/10.3758/cabn.5.2.169

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE, 12*(9), e0184661. https://doi.org/10.1371/journal.pone.0184661

Esteban, O., Blair, R. W., Nielson, D. M., Varada, J. C., Marrett, S., Thomas, A. G., Pol- drack, R. A., & Gorgolewski, K. J. (2019). Crowdsourced MRI quality metrics and ex- pert quality annotations for training of humans and machines. *Scientific Data, 6*(1), 30. https://doi.org/10.1038/s41597-019-0035-4

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods, 16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Etkin, A., Klemenhagen, K. C., Dudman, J. T., Rogan, M. T., Hen, R., Kandel, E. R., & Hirsch, J. (2004). Individual differences in trait anxiety predict the response of the basolateral amygdala to unconsciously processed fearful faces. *Neuron, 44*(6), 1043–1055. https://doi.org/10.1016/j.neuron.2004.12.006

Etzel, J. A., Courtney, Y., Carey, C. E., Gehred, M. Z., Agrawal, A., & Braver, T. S. (2020). Pattern similarity analyses of frontoparietal task coding: individual variation and genetic influences. *Cerebral Cortex, 30*(5), 3167–3183. https://doi.org/10.1093/cercor/bhz301

Fan, Y. T., Chen, C., Chen, S. C., Decety, J., & Cheng, Y. (2014). Empathic arousal and social understanding in individuals with autism: evidence from fMRI and ERP measurements. *Social cognitive and affective neuroscience, 9*(8), 1203–1213. https://doi.org/10.1093/scan/nst101

Farris, C., Treat, T. A., Viken, R. J., & McFall, R. M. (2008). Perceptual mechanisms that

characterize gender differences in decoding women's sexual intent. *Psychological Science, 19*(4), 348–354. https://doi.org/10.1111/j.1467-9280.2008.02092.x

Fellrath, J., Manuel, A. L., & Ptak, R. (2014). Task relevance effects in electrophysiological brain activity: Early, but not first. *NeuroImage, 101*, 68–75. https://doi.org/10.1016/j.NeuroImage.2014.06.059

Feuerriegel, D., Churches, O., Hofmann, J., & Keage, H. A. D. (2015). The N170 and face perception in psychiatric and neurological disorders: A systematic review. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, *126*(6), 1141–1158. https://doi.org/10.1016/j.clinph.2014.09.015

Fine, C. (2005). *Delusions of gender: The real science behind sex differences*. Icon Books Ltd.

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

Floris, D. L., Lai, M. C., Nath, T., Milham, M. P., & Di Martino, A. (2018). Network-specific sex differentiation of intrinsic brain function in males with autism. *Molecular autism, 9,* 17. https://doi.org/10.1186/s13229-018-0192-x

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

Fossati, A., Somma, A., Krueger, R. F., Markon, K. E., & Borroni, S. (2017). On the relationships between DSM-5 dysfunctional personality traits and social cognition deficits: A study in a sample of consecutively admitted Italian psychotherapy patients. *Clinical Psychology & Psychotherapy, 24*, 1421–1434. https://doi.org/10.1002/cpp.2091

Fox, N.A., Bakermans-Kranenburg, M.J., Yoo, K.H., et al. (2016). Assessing human mirror activity with EEG mu rhythm: a meta-analysis. *Psychological Bulletin*, *142*(3), 291–313. https://doi.org/10.1037/bul0000031

Frith, U., & Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *365*(1537), 165–176. https://doi.org/10.1098/rstb.2009.0160

Furnham, A., & Cheng, H. (2015). Early indicators of adult trait Agreeableness. *Personality and Individual Differences*, *73*, 67–71. https://doi.org/10.1016/j.paid.2014.09.025

Gallese, V., Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501. https://doi.org/10.1016/s1364-6613(98)01262-5

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current biology: CB*, *16*(18), 1824–1829. https://doi.org/10.1016/j.cub.2006.07.072

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., Barthélémy, C., Brunelle, F., Samson, Y., & Zilbovicius, M. (2004). Abnormal cortical voice processing in autism. *Nature neuroscience*, *7*(8), 801–802. https://doi.org/10.1038/nn1291

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., & Burgess, P. W. (2006). Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. *Journal of cognitive neuroscience*, *18*(6), 932–948. https://doi.org/10.1162/jocn.2006.18.6.932

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy*, *14*(5), 550–562. https://doi.org/10.1080/15250000903144199

Goldberg, L. R. (1990). An alternative “description of personality”: the big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037//0022-3514.59.6.1216

Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S.S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics, 5*, 13. https://doi.org/10.3389/fn- inf.2011.00013

Gorgolewski, K. J., Esteban, O., Ellis, D. G., Notter, M. P., Ziegler, E., Johnson, H., Hamalainen, C., Yvernault, B., Burns, C., Manhães-Savio, A., Jarecka, D., Markiewicz, C. J., Salo, T., Clark, D., Waskom, M., Wong, J., Modat, M., Dewey, B. E., Clark, M. G., & Ghosh, S. (2017). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. 0.13.1. Zenodo. https://doi.org/10.5281/ZENODO.581704

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V. V., Nichols, T. E., Poldrack, R. A., Poline, J. B., Yarkoni, T., & Margulies, D. S. (2015). NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics, 9*, 8. https://doi.org/10.3389/fn-inf.2015.00008

Gramfort, A., Papadopoulo, T., Olivi, E., & Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomedical Engineering Online*, *9*, 45. https://doi.org/10.1186/1475-925X-9-45

Gray, J. A. (1982). *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System*. Clarendon Press/Oxford University Press.

Graziano, W.G., & Eisenberg, N. (1997). Agreeableness: A dimension of personality. In R. Hogan,

J. A. Johnson, & S. R. Briggs (Eds.). *Handbook of Personality Psychology* (pp. 795–824). Academic Press.

Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, Empathy, and Helping: A Person × Situation Perspective. *Journal of Personality and Social Psychology*, *93*(4), 583–599. https://doi.org/10.1037/0022-3514.93.4.583

Graziano, W. G., & Tobin, R. M. (2016). Agreeableness and the five factor model. In T. A. Widiger (Ed.). *The Oxford Handbook of the Five Factor Model* (pp. 105–132). Oxford University Press. https://doi.org/10.1093/ox- fordhb/9780199352487.013.17

Green, A. E., Fugelsang, J. A., Kraemer, D. J., Shamosh, N. A., & Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain research*, *1096*(1), 125–137. https://doi.org/10.1016/j.brainres.2006.04.024

Greenberg, D. M., Warrier, V., Allison, C., & Baron-Cohen, S. (2018). Testing the Empathizing-Systemizing theory of sex differences and the Extreme Male Brain theory of autism in half a million people. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(48), 12152–12157. https://doi.org/10.1073/pnas.1811032115

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

Groppe, D.M., Urbach, T.P., Kutas, M., (2011a). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology*, *48*(12), 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Groppe, D.M., Urbach, T.P., Kutas, M., (2011b). Mass univariate analysis of event-related brain potentials/fields II: simulation studies. *Psychophysiology*, *48*(12), 1726–1737. https://doi.org/10.1111/j.1469-8986.2011.01272.x

Guellaï, B., Hausberger, M., Chopin, A., & Streri, A. (2020). Premises of social cognition: Newborns are sensitive to a direct versus a faraway gaze. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-66576-8

Gur, R. C., Gunning-Dixon, F., Bilker, W. B., & Gur, R. E. (2002). Sex differences in temporo-limbic and frontal brain volumes of healthy adults. *Cerebral Cortex*, *12*(9), 998–1003. https://doi.org/10.1093/cercor/12.9.998

Haas, B. W., & Miller, J. D. (2015). Borderline personality traits and brain activity during emotional perspective taking. *Personality Disorders: Theory, Research, and Treatment*, *6*(4), 315. https://doi.org/10.1037/per0000130

Haas, B. W., Omura, K., Constable, R. T., & Canli, T. (2007a). Emotional conflict and neuroticism: personality-dependent activation in the amygdala and subgenual anterior cingulate. *Behavioral neuroscience*, *121*(2), 249–256. https://doi.org/10.1037/0735-7044.121.2.249

Haas, B. W., Omura, K., Constable, R. T., & Canli, T. (2007b). Is automatic emotion regulation associated with agreeableness? A perspective using a social neuroscience approach. *Psychological science*, *18*(2), 130–132. https://doi.org/10.1111/j.1467-9280.2007.01861.x

Habashi, M. M., Graziano, W. G., & Hoover, A. E. (2016). Searching for the prosocial personality: a Big Five approach to linking personality and prosocial behavior. *Personality and Social Psychology Bulletin*, *42*(9), 1177–1192. https://doi.org/10.1177/0146167216652859

Hajcak, G., Dunning, J. P., & Foti, D. (2009). Motivated and controlled attention to emotion: time-course of the late positive potential. *Clinical neurophysiology : official journal of the*

International Federation of Clinical Neurophysiology, *120*(3), 505–510. https://doi.org/10.1016/j.clinph.2008.11.028

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: an integrative review. *Developmental neuropsychology*, *35*(2), 129–155. https://doi.org/10.1080/87565640903526504

Hall, J. A., & Halberstadt, A. G. (1981). Sex roles and nonverbal communication skills. *Sex Roles*, *7*(3), 273–287. https://doi.org/10.1007/BF00287542

Hansen, P. C. , Kringelbach, M. L. , & Salmelin, R. (2010). *MEG - An Introduction to Methods*. Oxford: University Press.

Hennig, J. (2004). Personality, serotonin, and noradrenaline. In R. M. Stelmack (Ed.). *On the Psychobiology of Personality: Essays in Honor of Marvin Zuckerman* (pp. 379–395). Elsevier Science. https://doi.org/10.1016/B978-008044209-9/50021-X

Hesselmark, E., Eriksson, J. M., Westerlund, J., & Bejerot, S. (2015). Autism Spectrum Disorders and Self-reports: Testing Validity and Reliability Using the NEO-PI-R. *Journal of autism and developmental disorders*, *45*(5), 1156–1166. https://doi.org/10.1007/s10803-014-2275-7

Hileman, C. M., Henderson, H., Mundy, P., Newell, L., & Jaime, M. (2011). Developmental and individual differences on the P1 and N170 ERP components in children with and without autism. *Developmental neuropsychology*, *36*(2), 214–236. https://doi.org/10.1080/87565641.2010.549870

Hobson, H. M., & Bishop, D. V. (2017). The interpretation of mu suppression as an index of mirror neuron activity: past, present and future. *Royal Society open science*, *4*(3), 160662. https://doi.org/10.1098/rsos.160662

Horan, W. P., Wynn, J. K., Kring, A. M., Simons, R. F., & Green, M. F. (2010). Electrophysiological correlates of emotional responding in schizophrenia. *Journal of abnormal psychology*, *119*(1), 18–30. https://doi.org/10.1037/a0017510

Hu, X., Erb, M., Ackermann, H., Martin, J. A., Grodd, W., & Reiterer, S. M. (2011). Voxel-based morphometry studies of personality: issue of statistical model specification--effect of nuisance covariates. *NeuroImage*, *54*(3), 1994–2005. https://doi.org/10.1016/j.neuroimage.2010.10.024

Huguet, P., Galvaing, M. P., Monteil, J. M., & Dumas, F. (1999). Social presence effects in the Stroop task: Further evidence for an attentional view of social facilitation. *Journal of Personality and Social Psychology*, *77*(5), 1011–1024. https://doi.org/10.1037/0022-3514.77.5.1011

Ibanez, A., Melloni, M., Huepe, D., Helgiu, E., Rivera-Rei, A., Canales-Johnson, A., Baker, P., & Moya, A. (2012). What event-related potentials (ERPs) bring to social neuroscience? *Social Neuroscience*, *7*(6), 632–649. https://doi.org/10.1080/17470919.2012.691078

Isernia, S., Sokolov, A. N., Fallgatter, A. J., & Pavlova, M. A. (2020). Untangling the Ties Between Social Cognition and Body Motion: Gender Impact. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00128

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

Jach, H. K., Feuerriegel, D., & Smillie, L. D. (2020). Decoding personality trait measures from resting EEG: An exploratory report. *Cortex; a journal devoted to the study of the nervous system and behavior*, *130*, 158–171. https://doi.org/10.1016/j.cortex.2020.05.013

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*, *14*(1), S103–9.

Jeannerod, M. (2006). *Motor Cognition: What Actions Tell the Self*, Oxford: Oxford University Press.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/NIMG.2002.1132

Jeste, S. S., & Nelson, C. A. (2009). Event related potentials in the understanding of autism spectrum disorders: An analytical review. *Journal of Autism and Developmental Disorders*, *39*(3), 495–510. https://doi.org/10.1007/s10803-008-0652-9

Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., Shefi, N., Pool, J., Urchs, S., Margulies, D. S., Liem, F., Hänggi, J., Jäncke, L., & Assaf, Y. (2015). Sex beyond the genitalia: The human brain mosaic. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(50), 15468–15473. https://doi.org/10.1073/pnas.1509654112

John, O., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy: history, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.). *Handbook of personality: Theory and research* (pp. 114–158). Guilford Press.

Kamp-Becker, I., Schröder, J., Remschmidt, H., & Bachmann, C. J. (2010). Health-related quality of life in adolescents and young adults with high functioning autism-spectrum disorder. *Psycho-social medicine*, *7*, Doc03. https://doi.org/10.3205/psm000065

Kang, E., Keifer, C. M., Levy, E. J., Foss-Feig, J. H., McPartland, J. C., & Lerner, M. D. (2018). Atypicality of the N170 Event-Related Potential in Autism Spectrum Disorder: A Meta-analysis. *Biological psychiatry. Cognitive neuroscience and neuroimaging*, *3*(8), 657–666. https://doi.org/10.1016/j.bpsc.2017.11.003

Kaplan, J. T., & Iacoboni, M. (2006). Getting a grip on other minds: mirror neurons, intention understanding, and cognitive empathy. *Social neuroscience*, *1*(3-4), 175–183. https://doi.org/10.1080/17470910600985605

Kapogiannis, D., Sutin, A., Davatzikos, C., Costa, P., Jr, & Resnick, S. (2013). The five factors of personality and regional cortical variability in the Baltimore longitudinal study of aging. *Human brain mapping*, *34*(11), 2829–2840. https://doi.org/10.1002/hbm.22108

Kappenman, E. S., Farrens, J. L., Zhang, W., Stewart, A. X., & Luck, S. J. (2021). ERP CORE: An open resource for human event-related potential research. *NeuroImage*, *225*, 117465. https://doi.org/10.1016/j.neuroimage.2020.117465

Keightley, M. L., Seminowicz, D. A., Bagby, R. M., Costa, P. T., Fossati, P., & Mayberg, H. S. (2003). Personality influences limbic-cortical interactions during sad mood induction. *NeuroImage*, *20*(4), 2031–2039. https://doi.org/10.1016/j.neuroimage.2003.08.022

Kerr, N., Dunbar, R. I. M., & Bentall, R. P. (2003). Theory of mind deficits in bipolar affective disorder. *Journal of Affective Disorders*, *73*(3), 253–259. https://doi.org/10.1016/s0165-0327(02)00008-3

Kikuchi, Y., Senju, A., Tojo, Y., Osanai, H., & Hasegawa, T. (2009). Faces Do Not Capture Special Attention In Children With Autism Spectrum Disorder: A Change Blindness Study. *Child Development*, *80*, 1421–1433. https://doi.org/10.1111/j.1467-8624.2009.01342.x

Kilner, J. M., Marchant, J. L., & Frith, C. D. (2006). Modulation of the mirror system by social relevance. *Social cognitive and affective neuroscience*, *1*(2), 143–148. https://doi.org/10.1093/scan/nsl017

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762–1776.

https://doi.org/10.1016/j.visres.2005.10.002

Kirkland, R. A., Peterson, E., Baker, C. A., Miller, S., & Pulos, S. (2013). Meta-analysis reveals adult female superiority in "Reading the Mind in the Eyes" Test. *North American Journal of Psychology*, *15*(1), 121–146.

Klados, M. A., Konstantinidi, P., Dacosta-Aguayo, R., Kostaridou, V. D., Vinciarelli, A., & Zervakis, M. (2020). Automatic Recognition of Personality Profiles Using EEG Functional Connectivity During Emotional Processing. *Brain sciences*, *10*(5), 278. https://doi.org/10.3390/brainsci10050278

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B. Reuter, M., Chaibub Neto, E., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Klem, G. H., Lüders, H., Jasper, H. H., & Elger, C. E. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology. Supplement*, *52*, 3–6.

Klin A. (1991). Young autistic children's listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *Journal of autism and developmental disorders*, *21*(1), 29–42. https://doi.org/10.1007/BF02206995

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry*, *59*(9), 809–816. https://doi.org/10.1001/archpsyc.59.9.809

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. https://doi.org/10.1016/j.NeuroImage.2008.03.059

Kogler, L., Müller, V. I., Werminghausen, E., Eickhoff, S. B., & Derntl, B. (2020). Do I feel or do I know? Neuroimaging meta-analyses on the multiple facets of empathy. *Cortex*, *129*, 341–355. https://doi.org/10.1016/j.cortex.2020.04.031

Koole, S. L., Jager,W., van den Berg, A. E., Vlek, C. A. J., &Hofstee, W. K. B. (2001). On the social nature of personality: Effects of Extraversion, Agreeableness, and feedback about collective re- source use on cooperation in a resource dilemma. *Personality and Social Psychology Bulletin*, *27*(3), 289–301. https://doi.org/10. 1177/0146167201273003

Kosonogov, V., Martinez-Selva, J. M., Carrillo-Verdejo, E., Torrente, G., Carretié, L., & Sanchez-Navarro, J. P. (2019). Effects of social and affective content on exogenous attention as revealed by event-related potentials. *Cognition and Emotion*, *33*(4), 683–695. https://doi.org/10.1080/02699931.2018.1486287

Koster-Hale, J., & Saxe, R. (2013). Functional neuroimaging of theory of mind. In S. Baron-Cohen, H. Tager-Flusberg, & M. V. Lombardo (Eds.). *Understanding other minds: Perspectives from developmental social neuroscience* (pp. 132–163). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199692972.003.0009

Kozel, B. A., Barak, B., Kim, C. A., Mervis, C. B., Osborne, L. R., Porter, M., & Pober, B. R. (2021). Williams syndrome. *Nature reviews. Disease primers*, *7*(1), 42. https://doi.org/10.1038/s41572-021-00276-z

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Developmental science*, *8*(1), F1–F12. https://doi.org/10.1111/j.1467-7687.2004.00384.x

Kuper, N., Käckenmester, W., & Wacker, J. (2019). Resting frontal EEG asymmetry and personality traits: a meta-analysis. *European Journal of Personality*, *33*(2), 154–175. https://doi.org/10.1002/per.2197

Lai, M. C., Lombardo, M. V., Suckling, J., Ruigrok, A. N., Chakrabarti, B., Ecker, C., Deoni, S. C., Craig, M. C., Murphy, D. G., Bullmore, E. T., MRC AIMS Consortium, & Baron-Cohen, S. (2013). Biological sex affects the neurobiology of autism. *Brain : a journal of neurology*, *136*(Pt 9), 2799–2815. https://doi.org/10.1093/brain/awt216

Lerner, M. D., McPartland, J. C., & Morris, J. P. (2013). Multimodal emotion processing in autism spectrum disorders: an event-related potential study. *Developmental cognitive neuroscience*, *3*, 11–21. https://doi.org/10.1016/j.dcn.2012.08.005

Leung, C., Lei, K. S., Wang, S. M., & Lam, B. Y. H. (2021). Theory of mind in schizotypy: A behavioral and event-related potential (ERP) study. *Schizophrenia Research: Cognition*, *23*, 100190. https://doi.org/10.1016/j.scog.2020.100190

Lewis, G. J., Dickie, D. A., Cox, S. R., Karama, S., Evans, A. C., Starr, J.M., Bastin, M. E., Wardlaw, J. M., Deary, I. J. (2018). Widespread associations between trait conscientiousness and thickness of brain cortical regions. *NeuroImage, 176*, 22–28. https://doi.org/10.1016/j.NeuroImage.2018.04.033

Li, G., Chen, Y., Wang, W., Dhingra, I., Zhornitsky, S., Tang, X., & Li, C. S. R. (2020). Sex Differences in Neural Responses to the Perception of Social Interactions. *Frontiers in Human*

Neuroscience, *14*, 565132. https://doi.org/10.3389/fnhum.2020.565132

Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): causal, multivariate, and reverse inference evidence. *Neuroscience & Biobehavioral Reviews*, *99*, 311–328. https://doi.org/10.1016/j.neubiorev.2018.12.021

Liss, M., Harel, B., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., Waterhouse, L., & Rapin, I. (2001). Predictors and correlates of adaptive functioning in children with developmental disorders. *Journal of autism and developmental disorders*, *31*(2), 219–230. https://doi.org/10.1023/a:1010707417274

Liu, D., Sabbagh, M.A., Gehring, W.J., & Wellman, H.M. (2005). An ERP study of 5-year-olds theory of mind. *Journal of Cognitive Neuroscience*, 190–190 .

Liu, W., Kohn, N., & Fernández, G. (2019). Intersubject similarity of personality is associated with intersubject similarity of brain connectivity patterns. *NeuroImage*, *186*, 56–69. https://doi.org/10.1016/J.NEUROIMAGE.2018.10.062

Liu, W.Y., Weber, B., Reuter, M., Markett, S., Chu, W.-C., & Montag, C. (2013). The Big Five of personality and structural imaging revisited: A VBM-DARTEL study. *Neuroreport*, *24*, 375–380. https://doi.org/10.1097/WNR.0b013e328360dad7

Lodi-Smith, J., Rodgers, J.D., Cunningham, S., Lopata, C., & Thomeer, M.L. (2019). Meta-analysis of Big Five personality traits in autism spectrum disorder. *Autism*, *23*, 556–565. https://doi.org/10.1177/1362361318766571

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E. J. H., Jones, R. M., Pickles, A., State, M. W., Taylor, J. L., & Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature reviews. Disease primers*, *6*(1), 5. https://doi.org/10.1038/s41572-

019-0138-4

Ludwig-Walz, H., Dannheim, I., Pfadenhauer, L. M., Fegert, J. M., & Bujard, M. (2023). Anxiety increased among children and adolescents during pandemic-related school closures in Europe: a systematic review and meta-analysis. *Child and adolescent psychiatry and mental health*, *17*(1), 74. https://doi.org/10.1186/s13034-023-00612-z

Macmillan, N. A., & Creelman, D. C. (1990). Response Bias: Characteristics of Detection Theory, Threshold Theory, and "Nonparametric" Indexes. *Psychological Bulletin*, *107*(3), 401–413. https://doi.org/10.1037/0033-2909.107.3.401

Makeig, S., Jung, T.-P., Bell, A. J., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In: D. Touretzky, M. C. Mozer, & M. Hasselmo (Eds.). *Advances in Neural Information Processing Systems* (Vol. 8, pp. 145-151). Cambridge: MIT Press.

Malhi, P., & Singhi, P. (2015). Adaptive Behavior Functioning in Children with Autism. *Indian journal of pediatrics*, *82*(8), 677–681. https://doi.org/10.1007/s12098-015-1692-8

Malouff, J. M., Thorsteinsson, E. B., Schutte, N. S., Bhullar, N., & Rooke, S. E. (2010). The five-factor model of personality and relationship satisfaction of intimate partners: A meta-analysis. *Journal of Research in Personality, 44*, 124–127. https://doi.org/10.1016/j. jrp.2009.09.004

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/J.JNEUMETH.2007.03.024

Marissen, M. A., Meuleman, L., & Franken, I. H. (2010). Altered emotional information processing in borderline personality disorder: an electrophysiological study. *Psychiatry research*, *181*(3), 226–232. https://doi.org/10.1016/j.pscychresns.2009.10.006

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology, 88*(1), 139–157. https://doi.org/10.1037/0022-3514.88.1.139

Masten, C. L., Colich, N. L., Rudie, J. D., Bookheimer, S. Y., Eisenberger, N. I., & Dapretto, M. (2011). An fMRI investigation of responses to peer rejection in adolescents with autism spectrum disorders. *Developmental cognitive neuroscience, 1*(3), 260–270. https://doi.org/10.1016/j.dcn.2011.01.004

McAbee, S. T., & Oswald, F. L. (2013). The criterion-related validity of personality measures for predicting GPA: A meta-analytic validity competition. *Psychological Assessment, 25*, 523–544. https://doi.org/10.1037/a0031748

Meier, B. P., Robinson, M. D., & Wilkowski, B. M. (2006). Turning the other cheek. Agreeableness and the regulation of aggression-related primes. *Psychological Science, 17*(2), 136–142.

Meinhardt, J., Sodian, B., Th oermer, C., Dohnel, K., & Sommer, M. (2011). True and false belief reasoning in children and adults: An event related potential study of theory of mind. *Developmental Cognitive Neuroscience, 1*(1), 67–76. https://doi.org/10.1016/j.dcn.2010.08.001

Meinhardt-Injac, B., Daum, M. M., Meinhardt, G., & Persike, M. (2018). The two-systems account of theory of mind: Testing the links to social-perceptual and cognitive abilities. *Frontiers in Human Neuroscience, 12*, 25. https://doi.org/10.3389/fnhum.2018.00025

Miller, J. D., & Lynman, D. (2001). Structural models of personality and their relation to antisocial behavior: A meta-analytic review. *Criminology, 39*, 765–798. https://doi.org/10.1111/j.1745-9125.2001.tb00940.x

Mobbs, D., Hagan, C. C., Azim, E., Menon, V., & Reiss, A. L. (2005). Personality predicts activity in reward and emotional regions associated with humor. *Proceedings of the National Academy*

of Sciences of the United States of America, *102*(45), 16502–16506. https://doi.org/10.1073/pnas.0408457102

Modi, M. E., & Young, L. J. (2012). The oxytocin system in drug discovery for autism: animal models and novel therapeutic strategies. *Hormones and behavior*, *61*(3), 340–350. https://doi.org/10.1016/j.yhbeh.2011.12.010

Moessnang, C., Otto, K., Bilek, E., Schäfer, A., Baumeister, S., Hohmann, S., Poustka, L., Brandeis, D., Banaschewski, T., Tost, H., & Meyer-Lindenberg, A. (2017). Differential responses of the dorsomedial prefrontal cortex and right posterior superior temporal sulcus to spontaneous mentalizing. *Human Brain Mapping*, *38*(8), 3791–3803. https://doi.org/10.1002/hbm.23626

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Montagne, B., Kessels, R. P. C., Frigerio, E., De Haan, E. H. F., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive Processing*, *6*, 135–141. https://doi.org/10.1007/s10339-005-0050-6

Montoya, E. R., Terburg, D., Bos, P. A., & van Honk, J. (2012). Testosterone, cortisol, and serotonin as key regulators of social aggression: A review and theoretical perspective. *Motivation and emotion*, *36*(1), 65–73. https://doi.org/10.1007/s11031-011-9264-3

Moore, C. (1999). Gaze following and the control of attention. In P. Rochat (Ed.). *Early Social Cognition: Understanding Others in the First Month of Life* (pp. 241–256). Erlbaum.

Moriguchi, Y., Ohnishi, T., Lane, R.D., Maeda, M., Mori, T., Nemoto, K., Mat- suda, H., Komaki, G., 2006. Impaired self-awareness and theory of mind: an fMRI study of mentalizing in

alexithymia. *NeuroImage, 32*(3), 1472–1482. https://doi.org/10.1016/J.NEUROIMAGE.2006.04.186

Mottron, L. (2020). Asperger syndrome. *The Wiley Encyclopedia of Health Psychology*, 187–196.

Mulders, P., Llera, A., Tendolkar, I., van Eijndhoven, P., & Beckmann, C. (2018). Personality profiles are associated with functional brain networks related to cognition and emotion. *Scientific Reports, 8*(1), 13874. https://doi.org/10.1038/s41598-018-32248-x

Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*, *47*(6), 497–514. https://doi.org/10.1111/ejn.13720

Musa, S., Dergaa, I., Bachiller, V., & Saad, H. B. (2023). Global Implications of COVID-19 Pandemic on Adults' Lifestyle Behavior: The Invisible Pandemic of Noncommunicable Disease. *International journal of preventive medicine, 14*, 15. https://doi.org/10.4103/ijpvm.ijpvm_157_21

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., et al. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society B: Biological Sciences, 277*(1696), 2935–2943. https://doi.org/10.1098/rspb.2010.0587.

Naselaris, T., & Kay, K. N. (2015). Resolving ambiguities of MVPA using explicit models of representation. *Trends in Cognitive Sciences, 19* (10), 551–554. https://doi.org/10.1016/J.TICS.2015.07.005

Netter, P. (2004). Personality and hormones. In R. M. Stelmack (Ed.). *On the Psychobiology of personality: Essays in Honor of Marvin Zuckerman* (pp. 353–377). Elsevier Science. https://doi.org/10.1016/B978-008044209-9/50020-8

Nettle, D., & Liddle, B. (2008). Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind. *European Journal of Personality*, *22*(4), 323–335. https://doi.org/10.1002/PER.672

Niedermeyer, E., da Silva, F.L. (2005) *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams &Wilkins.

Nostro, A. D., Müller, V. I., Varikuti, D. P., Pläschke, R. N., Hoffstaedter, F., Langner, R., Patil, K. R., & Eickhoff, S. B. (2018). Predicting personality from network-based resting-state functional connectivity. *Brain Structure and Function*, *223*(6), 2699–2719. https://doi.org/10.1007/s00429-018-1651-z

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, *43*, 971–990. https://doi.org/10.1016/j.paid.2007.03.017

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: an integrative review of ERP findings. *Biological psychology*, *77*(3), 247–265. https://doi.org/10.1016/j.biopsycho.2007.11.006

Omura, K., Todd Constable, R., & Canli, T. (2005). Amygdala gray matter concentration is associated with extraversion and neuroticism. *Neuroreport*, *16*(17), 1905–1908. https://doi.org/10.1097/01.wnr.0000186596.64458.76

Owens, M. M., Hyatt, C. S., Gray, J. C., Carter, N. T., MacKillop, J., Miller, J. D., & Sweet, L. H. (2019). Cortical morphometry of the five-factor model of personality: findings from the human connectome project full sample. *Social Cognitive and Affective Neuroscience*, *14*(4), 381–395. https://doi.org/10.1093/scan/nsz017

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential

outcomes. *Annual review of psychology, 57*, 401–421. https://doi.org/10.1146/annurev.psych.57.102904.190127

Paletta P., Bass, N., Aspesi, D., & Choleris, E. (2022). Sex Differences in Social Cognition. In: C. Gibson & L. A. M. Galea (Eds.). *Sex Differences in Brain Function and Dysfunction* (Vol. 62, pp. 207–234). New York: Springer International Publishing. https://doi.org/10.1007/7854_2022_325

Pardini, D. A., Raine, A., Erickson, K., & Loeber, R. (2014). Lower amygdala volume in men is associated with childhood aggression, early psychopathic traits, and future violence. *Biological psychiatry*, *75*(1), 73–80. https://doi.org/10.1016/j.biopsych.2013.04.003

Pavlova, M., Guerreschi, M., Lutzenberger, W., Sokolov, A. N., & Krägeloh-Mann, I. (2010). Cortical response to social interaction is affected by gender. *NeuroImage, 50*(3), 1327–1332. https://doi.org/10.1016/j.neuroimage.2009.12.096

Pavlova, M. A., Sokolov, A. N., and Bidet-Ildei, C. (2015). Sex differences in the neuromagnetic response to biological motion. *Cerebral Cortex, 25*, 3468–3474. https://doi.org/10.1093/cercor/bhu175

Pedersen, A., Koelkebeck, K., Brandt, M., Wee, M., Kueppers, K. A., Kugel, H., Kohl, W., Bauer, J., & Ohrmann, P. (2012). Theory of mind in patients with schizophrenia: is mentalizing delayed?. *Schizophrenia research*, *137*(1-3), 224–229. https://doi.org/10.1016/j.schres.2012.02.022

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *The Journal of Neuroscience, 30*(30), 10127–10134. https://doi.org/10.1523/JNEUROSCI.2161-10.2010

Penner, L. A., Fritzsche, B. A., Craiger, J. P., & Freifeld, T. S. (1995). Measuring the prosocial

personality. In J. N. Butcher & C. D. Spielberger (Eds.). *Advances in Personality Assessment*, (Vol. 10, pp. 147–163). Lawrence Erlbaum Associates, Inc.

Pernet, C. R., Chauveau, N., Gaspar, C., & Rousselet, G. A. (2011). LIMO EEG: A toolbox for hierarchical linear modeling of electroencephalographic data. *Computational Intelligence and Neuroscience, 2011,* 831409. https://doi.org/10.1155/2011/831409

Pernet, C. R., Latinus, M., Nichols, T. E., & Rousselet, G. A. (2015). Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. *Journal of Neuroscience Methods*, *250*, 85–93. https://doi.org/10.1016/j.jneumeth.2014.08.003

Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2013). Robust correlation analyses: False positive and power validation using a new open source matlab toolbox. *Frontiers in Psychology*, *3*, 606. https://doi.org/10.3389/fpsyg.2012.00606

Perrin, J. S., Leonard, G., Perron, M., et al. (2009). Sex differences in the growth of white matter during adolescence. *NeuroImage*, *45*(4), 1055–1066. https://doi.org/10.1016/j.NeuroImage.2009.01.023

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*(2), 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

Perry, A., Mankuta, D., & Shamay-Tsoory, S. G. (2015). OT promotes closer interpersonal distance among highly empathic individuals. *Social cognitive and affective neuroscience*, *10*(1), 3–9. https://doi.org/10.1093/scan/nsu017

Pineda, J. A., & Hecht, E. (2009). Mirroring and mu rhythm involvement in social cognition: are there dissociable subcomponents of theory of mind?. *Biological psychology*, *80*(3), 306–314.

https://doi.org/10.1016/j.biopsycho.2008.11.003

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Pisanu, E., Arbula, S., & Rumiati, R. I. (under review). Personality and sex differences in social interaction perception: an ERP and source imaging study

Pisanu, E., Arbula, S., & Rumiati, R. I. (under review). Agreeableness modulates mental state decoding: Electrophysiological evidence.

Pisanu, E., Di Benedetto, A., Infurna, M. R., & Rumiati, R. I. (2022). Psychological Impact in Healthcare Workers During Emergencies: The Italian Experience With COVID-19 First Wave. *Frontiers in psychiatry*, *13*, 818674. https://doi.org/10.3389/fpsyt.2022.818674

Pohl, A., Cassidy, S., Auyeung, B., & Baron-Cohen, S. (2014). Uncovering steroidopathy in women with autism: a latent class analysis. *Molecular autism*, *5*, 27. https://doi.org/10.1186/2040-2392-5-27

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

Proverbio, A. M. (2021). Sex differences in the social brain and in social cognition. *Journal of Neuroscience Research*, *101*(5), 730–738. https://doi.org/10.1002/jnr.24787

Proverbio, A. M., Zani, A., & Adorni, R. (2008). Neural markers of a greater female responsiveness to social stimuli. *BMC Neuroscience*, *9*, 56. https://doi.org/10.1186/1471-2202-9-56

Ptak, R. (2012). The frontoparietal attention network of the human brain: Action, saliency, and a priority map of the environment. *Neuroscientist*, *18*(5), 502–515. https://doi.org/10.1177/1073858411409051

Puce, A., Rossi, A. Parada, F. (2015) Biological motion. *Brain Mapping: An Encyclopedic Reference*, *3*, 125–130. https://10.1016/B978-0-12-397025-1.00166-4

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

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature reviews. Neuroscience*, *5*(3), 184–194. https://doi.org/10.1038/nrn1343

Ratner, K. & Burrow, A. L. (2017) Autistic features in the general population: Implications for sensing purpose in life. *Journal of Positive Psychology, 13*(5), 494–501. https://doi.org/10.1080/17439760.2017.1315647

Rauch, S. L., Milad, M. R., Orr, S. P., Quinn, B. T., Fischl, B., & Pitman, R. K. (2005). Orbitofrontal thickness, retention of fear extinction, and extraversion. *Neuroreport*, *16*(17), 1909–1912. https://doi.org/10.1097/01.wnr.0000186599.66243.50

Renty, J. O., & Roeyers, H. (2006). Quality of life in high-functioning adults with autism spectrum disorder: The predictive value of disability and support characteristics. *Autism, 10*(5), 511–524. https://doi.org/10.1177/1362361306066604

Reuter, M., Stark, R., Hennig, J., Walter, B., Kirsch, P., Schienle, A., & Vaitl, D. (2004). Personality and emotion: test of Gray's personality theory by means of an fMRI study. *Behavioral neuroscience*, *118*(3), 462–469. https://doi.org/10.1037/0735-7044.118.3.462

Riby, D. M., & Hancock, P. J. (2008). Viewing it differently: social scene perception in Williams syndrome and autism. *Neuropsychologia*, *46*(11), 2855–2860. https://doi.org/10.1016/j.neuropsychologia.2008.05.003

Riccelli, R., Toschi, N., Nigro, S., Terracciano, A., & Passamonti, L. (2017). Surface-based morphometry reveals the neuroanatomical basis of the five-factor model of personality. *Social Cognitive and Affective Neuroscience*, *12*(4), 671–684. https://doi.org/10.1093/scan/nsw175

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. https://doi.org/10.1016/s0028-3932(02)00175-6

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*(5695), 443–447. https://doi.org/10.1126/science.1100301

Ritchie, J. B., Kaplan, D. M., & Klein, C. (2019). Decoding the brain: neural representation and the limits of multivariate pattern analysis in cognitive neuroscience. *The British Journal for the Philosophy of Science*, *70*(2), 581–607. https://doi.org/10.1093/bjps/axx023

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual review of neuroscience*, *27*, 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: a detection advantage in the flicker paradigm. *Psychological Science*, *12*(1), 94–99. https://doi.org/10.1111/1467-9280.00317

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, *143*, 117–14. https://doi.org/10.1037/bul0000088

Robinson, E., Hull, L., & Petrides, K. V. (2020). Big Five model and trait emotional intelligence in camouflaging behaviours in autism. *Personality and Individual Differences*, *152*(April 2019), 109565. https://doi.org/10.1016/j.paid.2019.109565

Rodgers, J. D., Lodi-Smith, J., Hill, P. L., Spain, S. M., Lopata, C., & Thomeer, M. L. (2018). Brief Report: Personality Mediates the Relationship between Autism Quotient and Well-Being: A Conceptual Replication using Self-Report. *Journal of autism and developmental disorders*, *48*(1), 307–315. https://doi.org/10.1007/s10803-017-3290-2

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage*, *20*(3), 1609–1624. https://doi.org/10.1016/j.neuroimage.2003.07.010

Rueda, P. Fernández-Berrocal, P. & Baron-Cohen, S. (2015) Dissociation between cognitive and affective empathy in youth with Asperger Syndrome, *European Journal of Developmental Psychology*, *12*(1), 85–98. https://doi.org/10.1080/17405629.2014.950221

Sabbagh, M. A. (1999). Communicative intentions and language: evidence from right-hemisphere damage and autism. *Brain and language*, *70*(1), 29–69. https://doi.org/10.1006/brln.1999.2139

Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: implications for autism. *Brain and cognition*, *55*(1), 209–219. https://doi.org/10.1016/j.bandc.2003.04.002

Sabbagh, M. A. (2013). Brain electrophysiological studies of theory of mind. S., In Baron-Cohen, Tager-Flusberg, H., & Lombardo, M. (Eds.). *Understanding other minds: Perspectives from developmental social neuroscience* (pp. 119-131). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199692972.003.0008

Sabbagh, M. A., & Bowman, L. C. (2018). Theory of mind. In: S. Ghetti & J. Wixted (Eds.). *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (4th ed., pp. 1–34). Wiley. https://doi.org/10.1002/9781119170174.epcn408

Sabbagh, M. A., & Flynn, J. (2006). Mid-frontal EEG alpha asymmetries predict individual differences in one aspect of theory of mind: mental state decoding. *Social neuroscience*, *1*(3-4), 299–308. https://doi.org/10.1080/17470910601029163

Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural correlates of mental state decoding in human adults: an event-related potential study. *Journal of cognitive neuroscience*, *16*(3), 415–426. https://doi.org/10.1162/089892904322926755

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind. A comparison of Chinese and U.S. preschoolers. *Psychological science*, *17*(1), 74–81. https://doi.org/10.1111/j.1467-9280.2005.01667.x

Sabbagh, M.A., & Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: An event-related potential study. *Psychological Science, 11*(1), 46–50. https://doi.org/10.1111/1467-9280.00213

Samaey, C., Van der Donck, S., van Winkel, R., & Boets, B. (2020). Facial Expression Processing Across the Autism-Psychosis Spectra: A Review of Neural Findings and Associations With Adverse Childhood Events. *Frontiers in psychiatry*, *11*, 592937. https://doi.org/10.3389/fpsyt.2020.592937

Sampaio, A., Soares, J. M., Coutinho, J., Sousa, N., & Gonçalves, Ó. F. (2014). The Big Five default brain: functional evidence. *Brain structure & function*, *219*(6), 1913–1922. https://doi.org/10.1007/s00429-013-0610-y

Sato, W., Kochiyama, T., Yoshikawa, S., & Matsumura, M. (2001). Emotional expression boosts

early visual processing of the face: ERP recording and its decomposition by independent component analysis. *Neuroreport*, *12*(4), 709–714. https://doi.org/10.1097/00001756-200103260-00019

Saxe, R., & Baron-Cohen, S. (2006). The neuroscience of theory of mind. *Social Neuroscience*, *1*(3–4), 1–9. https://doi.org/10.1080/17470910601117463

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

Schacht, A., & Vrtička, P. (2018). Spatiotemporal pattern of appraising social and emotional relevance: Evidence from event-related brain potentials. *Cognitive, Affective, & Behavioral Neuroscience Volume*, *18*, 1172–1187. https://doi.org/10.3758/s13415-018-0629-x

Schinka, J. A., Busch, R. M., & Robichaux-Keene, N. (2004). A meta-analysis of the association between the serotonin transporter gene polymorphism (5-HTTLPR) and trait anxiety. *Molecular psychiatry*, *9*(2), 197–202. https://doi.org/10.1038/sj.mp.4001405

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). E-Prime 2.0. Psychology Software Tools, Inc.

Schriber, R. A., Robins, R. W., & Solomon, M. (2014). Personality and self-insight in individuals with autism spectrum disorder. *Journal of personality and social psychology*, *106*(1), 112–130. https://doi.org/10.1037/a0034950

Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R., & Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *Journal of cognitive neuroscience*, *19*(8), 1354–1372. https://doi.org/10.1162/jocn.2007.19.8.1354

Schultz R. T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *International journal of developmental neuroscience : the official journal of the International Society for Developmental Neuroscience*, *23*(2-3), 125–141. https://doi.org/10.1016/j.ijdevneu.2004.12.012

Schultz, R. T., Gauthier, I., Klin, A., Fulbright, R. K., Anderson, A. W., Volkmar, F., Skudlarski, P., Lacadie, C., Cohen, D. J., & Gore, J. C. (2000). Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger syndrome. *Archives of general psychiatry*, *57*(4), 331–340. https://doi.org/10.1001/archpsyc.57.4.331

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., & Lang, P. J. (2000). Affective picture processing: the late positive potential is modulated by motivational relevance. *Psychophysiology*, *37*(2), 257–261. http://www.ub.uni-konstanz.de/kops/volltexte/2006/2105/

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Schurz, M., Radua, J., Tholen, M. G., et al. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin*, *147*(3), 293–327. https://doi.org/10.1037/bul0000303

Schwartzman, B. C., Wood, J. J., & Kapp, S. K. (2016). Can the Five Factor Model of Personality Account for the Variability of Autism Symptom Expression? Multivariate Approaches to Behavioral Phenotyping in Adult Autism Spectrum Disorder. *Journal of autism and developmental disorders*, *46*(1), 253–272. https://doi.org/10.1007/s10803-015-2571-x

Schwarz, E., Guest, P. C., Rahmoune, H., Wang, L., Levin, Y., Ingudomnukul, E., Ruta, L., Kent,

L., Spain, M., Baron-Cohen, S., & Bahn, S. (2011). Sex-specific serum biomarker patterns in adults with Asperger's syndrome. *Molecular psychiatry*, *16*(12), 1213–1220. https://doi.org/10.1038/mp.2010.102

Scott-Van Zeeland, A. A., Dapretto, M., Ghahremani, D. G., Poldrack, R. A., & Bookheimer, S. Y. (2010). Reward processing in autism. *Autism research: official journal of the International Society for Autism Research*, *3*(2), 53–67. https://doi.org/10.1002/aur.122

Sen, S., Burmeister, M., & Ghosh, D. (2004). Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, *127B*(1), 85–89. https://doi.org/10.1002/ajmg.b.20158

Sibley, C. G., & Duckitt, J. (2008). Personality and prejudice: A meta- analysis and theoretical review. *Personality and Social Psychology Review*, *12*(3), 248–279. https://doi.org/10.1177/1088868308319226.

Siqi-Liu, A., Harris, A. M., Atkinson, A. P., & Reed, C. L. (2018). Dissociable processing of emotional and neutral body movements revealed by μ-alpha and beta rhythms. *Social Cognitive and Affective Neuroscience*, *13*(12), 1269–1279. https://doi.org/10.1093/scan/nsy094

Skerry, A. E., & Saxe, R. (2014). A common neural code for perceived and inferred emotion. *The Journal of Neuroscience*, *34*(48), 15997–16008. https://doi.org/10.1523/JNEUROSCI.1676-14.2014

Skerry, A. E., & Saxe, R. (2015). Neural representations of emotion are organized around abstract event features. *Current Biology*, *2*(15), 1945–1954.

https://doi.org/10.1016/J.CUB.2015.06.009

Sliwa, J., & Freiwald, W. A. (2017). A dedicated network for social interaction processing in the primate brain. *Science*, *356*(6339), 745–749. https://doi.org/10.1126/science.aam6383

Snowden, J. S., Gibbons, Z. C., Blackshaw, A., Doubleday, E., Thompson, J., Craufurd, D., Foster, J., Happé, F., & Neary, D. (2003). Social cognition in frontotemporal dementia and Huntington's disease. *Neuropsychologia, 41*(6), 688–701. https://doi.org/10.1016/s0028-3932(02)00221-x

Sokhadze, E. M., Frederick, J., Wang, Y., Kong, M., El-Baz, A. S., Tasman, A., & Casanova, M. F. (2015). Event-related potential (erp) study of facial expression processing deficits in autism. *Journal of Communications Research*, *7*(4), 391-412.

Solomon, M., Goodlin-Jones, B. L., & Anders, T. F. (2004). A social adjustment enhancement intervention for high functioning autism, Asperger's syndrome, and pervasive developmental disorder NOS. *Journal of autism and developmental disorders*, *34*(6), 649–668. https://doi.org/10.1007/s10803-004-5286-y

Soto, C. J., & Tackett, J. L. (2015). Personality traits in childhood and adolescence: Structure, development, and outcomes. *Current Directions in Psychological Science*, *24*(5), 358-362. https://doi.org/10.1177/0963721415589

Spreng, R. N., & Andrews-Hanna, J. R. (2015). The default network and social cognition. *Brain Mapping: An Encyclopedic Reference, 3,* 165–169. https://doi.org/10.1016/B978-0-12-397025-1.00173-1

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of Cognitive Neuroscience, 21*(3), 489–510.

https://doi.org/10.1162/jocn.2008.21029

Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension ofobserved emotion- al behavior. *NeuroImage, 59*(3), 3050–3059. https://doi.org/10. 1016/j.neuroimage.2011.10.005

Stake, J. E., & Eisele, H. (2010). Gender and personality. In J. Chrisler & D. McCreary (Eds.). *Handbook of Gender Research in Psychology, Vol 2: Gender research in social and applied psychology* (pp. 19–40*)*. Springer. https://doi.org/10.1007/978-1-4419-1467-5

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 3,* 37–149. https://doi.org/10.3758/BF03207704

Stavropoulos, K. K. M., & Carver, L. J. (2014). Reward anticipation and processing of social versus nonsocial stimuli in children with and without autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 55*(12), 1398–1408. https://doi.org/10.1111/jcpp.12270

Stenroos, M., Hunold, A., & Haueisen, J. (2014). Comparison of three-shell and simplified volume conductor models in magnetoencephalography. *NeuroImage, 94,* 337–348. https://doi.org/10.1016/j.neuroimage.2014.01.006

Stiller, J., & Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks, 29,* 93–104. https://doi.org/10.1016/j.socnet.2006.04.001

Strang, J. F., Kenworthy, L., Dominska, A., Sokoloff, J., Kenealy, L. E., Berl, M., Walsh, K., Menvielle, E., Slesaransky-Poe, G., Kim, K. E., Luong-Tran, C., Meagher, H., & Wallace, G. L. (2014). Increased gender variance in autism spectrum disorders and attention deficit hyperactivity disorder. *Archives of sexual behavior*, *43*(8), 1525–1533. https://doi.org/10.1007/s10508-014-0285-3

Strunz, S., Westphal, L., Ritter, K., Heuser, I., Dziobek, I., & Roepke, S. (2015). Personality Pathology of Adults With Autism Spectrum Disorder Without Accompanying Intellectual Impairment in Comparison to Adults With Personality Disorders. *Journal of autism and developmental disorders*, *45*(12), 4026–4038. https://doi.org/10.1007/s10803-014-2183-x

Sturm, V. E., Haase, C. M., & Levenson, R. W. (2016). Emotional Dysfunction in Psychopathology and Neuropathology: Neural and Genetic Pathways. *Genomics, Circuits, and Pathways in Clinical Neuropsychiatry*, 345–364. https://doi.org/10.1016/B978-0-12-800105-9.00022-6

Stuss, D. T., Gallup, G. G., & Alexander, M. P. (2001). The frontal lobes are necessary for "theory of mind". *Brain*, *124*(2), 279–286. https://doi.org/10.1093/brain/124.2.279

Su, J., van Boxtel, J. J. A., & Lu, H. (2016). Social Interactions Receive Priority to Conscious Perception. *Plos One*, *11*(8), e0160468. https://doi.org/10.1371/journal.pone.0160468

Suh, J., Orinstein, A., Barton, M., Chen, C. M., Eigsti, I. M., Ramirez-Esparza, N., & Fein, D. (2016). Ratings of Broader Autism Phenotype and Personality Traits in Optimal Outcomes from Autism Spectrum Disorder. *Journal of autism and developmental disorders*, *46*(11), 3505–3518. https://doi.org/10.1007/s10803-016-2868-4

Tackett, J. L., Hernández, M. M., & Eisenberg, N. (2019). Agreeableness. In D. P. McAdams, R. L. Shiner, & J. L. Tackett (Eds.). *Handbook of personality development* (pp. 171–184). The Guilford Press.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 2011, 879716. https://doi.org/10.1155/2011/879716

Tager-Flusberg, H., & Sullivan, K. (2000). A componential view of theory of mind: evidence from Williams syndrome. *Cognition*, *76*(1), 59–90. https://doi.org/10.1016/s0010-0277(00)00069-x

Taki, Y., Thyreau, B., Kinomura, S., Sato, K., Goto, R., Wu, K., Kawashima, R., Fukuda, H., 2013. A longitudinal study of the relationship between personality traits and the annual rate of volume changes in regional gray matter in healthy adults. *Human Brain Mapping, 34*(12), 3347–3353. https://doi.org/10.1002/hbm.22145

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: rationality, social impact, and valence. *PNAS*, *113*(1), 194–199. https://doi.org/10.1073/pnas.1511905112

Tankersley, D., Stowe, C. J., & Huettel, S. A. (2007). Altruism is associated with an increased neural response to agency. *Nature neuroscience*, *10*(2), 150–151. https://doi.org/10.1038/nn1833

Tauscher, J., Bagby, M. R., Javanmard, M., Christensen, B. K., Kasper, S., & Kapur, S. (2001). Inverse relationship between serotonin 5-HT1a receptor binding and anxiety: A [11C]WAY-100635 PET investigation in healthy volunteers. *American Journal of Psychiatry, 158*, 1326–1328.

Tellegen, A., & Waller, N. G. (1981). Exploring personality through test construction: development of the multidimensional personality questionnaire. In G. J. Boyle, G. Matthews & D. H. Saklofske (Eds.). *The SAGE Handbook of Personality Theory and Assessment: Volume 2 — Personality Measurement and Testing* (pp. 261–292). SAGE Publications Ltd. https://doi.org/10.4135/9781849200479.n13

Tesar, B., Deckert, M., Schmoeger, M., & Willinger, U. (2020). Electrophysiological Correlates of Basic and Higher Order Cognitive and Affective Theory of Mind Processing in Emerging and Early Adulthood-An Explorative Event-Related Potentials Study to Investigate First-,

Second-, and Third-Order Theory of Mind Processing Based on Visual Cues. *Frontiers in human neuroscience*, *14*, 79. https://doi.org/10.3389/fnhum.2020.00079

Turan, B., Guo, J., Boggiano, M. M., & Bedgood, D. (2014). Dominant, cold, avoidant, and lonely: Basal testosterone as a biological marker for an interpersonal style. *Journal of Research in Personality*, *50*, 84–89. https://doi.org/10.1016/j.jrp.2014.03.008

Twite, M. D., Stenquist, S., & Ing, R. J. (2019). Williams syndrome. *Paediatric anaesthesia*, *29*(5), 483–490. https://doi.org/10.1111/pan.13620

Ubbiali, A., Chiorri, C., Hampton, P., & Donati, D. (2013). Italian big five inventory. Psychometric properties of the Italian adaptation of the big five inventory (BFI). *Applied Psychology Bulletin*, *59*(266), 37–48.

Udochi, A. L., Blain, S. D., Sassenberg, T. A., Burton, P. C., Medrano, L., & DeYoung, C. G. (2022). Activation of the default network during a theory of mind task predicts individual differences in agreeableness and social cognitive ability. *Cognitive, Affective and Behavioral Neuroscience*, *22*(2), 383–402. https://doi.org/10.3758/s13415-021-00955-0

Ulloa, E. R., & Pineda, J. A. (2007). Recognition of point-light biological motion: mu rhythms and mirror neuron activity. *Behavioural brain research*, *183*(2), 188–194. https://doi.org/10.1016/j.bbr.2007.06.007

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, *30*(3), 829–858. https://doi.org/10.1002/hbm.20547

Vedel, A. (2014). The Big Five and tertiary academic performance: A systematic review and meta-analysis. *Personality and Individual Differences*, *71*, 66–76. https://doi.org/10.1016/j.paid.2014.07.011

Vestner, T., Tipper, S., Hartley, T., Over, H., & Rueschemeyer, S. A. (2018). Bound Together: Social binding leads to faster processing, spatial distortion and enhanced memory of interacting partners. *Journal of Vision, 18*(10), 448. https://doi.org/10.1167/18.10.448

Vincent, A., Da Fonseca, D., Baumstarck, K., Charvin, I., Alcaraz-Mor, R., & Lehucher-Michel, M. P. (2020). The quality of life and the future of young adults with Asperger syndrome. *Disability and rehabilitation*, *42*(14), 1987-1994.

Wagner, D. D., Chavez, R.S., & Broom, T. W. (2018). Decoding the neural representation of self and person knowledge with multivariate pattern analysis and data-driven approaches. *Wiley interdisciplinary reviews. Cognitive science, 10*(1), e1482. https://doi.org/10.1002/wcs.1482

Wagner, D. D., Kelley, W. M., & Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cerebral Cortex, 21*(12), 2788–2796. https://doi.org/10.1093/cercor/bhr074

Wagner, D. D., Kelley, W. M., Haxby, J. V., & Heatherton, T. F. (2016). The dorsal medial pre-frontal cortex responds preferentially to social interactions during natural viewing. *The Journal of Neuroscience, 36*(26), 6917–6925. https://doi.org/10.1523/JNEUROSCI.4220-15.2016

Wang, F., Wang, X., Wang, F., Gao, L., Rao, H., & Pan, Y. (2019). Agreeableness modulates group member risky decision-making behavior and brain activity. *NeuroImage, 202*, 116100. https://doi.org/10.1016/j.neuroimage.2019.116100

Wang, Y.W., Liu, Y., Gao, Y.X., Chen, J., Zhang, W.X., & Lin, C.D. (2008). False belief reasoning in the brain: An ERP study. *Science in China. Series C, Life Sciences*, *51*(1), 72–9. https://doi.org/10.1007/s11427-008-0014-z

Warrier, V., Grasby, K. L., Uzefovsky, F., Toro, R., Smith, P., Chakrabarti, B., Khadake, J., Mawbey-Adamson, E., Litterman, N., Hottenga, J. J., Lubke, G., Boomsma, D. I., Martin, N. G., Hatemi, P. K., Medland, S. E., Hinds, D. A., Bourgeron, T., & Baron-Cohen, S. (2018). Genome-wide meta-analysis of cognitive empathy: heritability, and correlates with sex, neuropsychiatric conditions and cognition. *Molecular psychiatry*, *23*(6), 1402–1409. https://doi.org/10.1038/mp.2017.122

Warrier, V., Toro, R., Won, H., Leblond, C. S., Cliquet, F., Delorme, R., De Witte, W., Bralten, J., Chakrabarti, B., Børglum, A. D., Grove, J., Poelmans, G., Hinds, D. A., Bourgeron, T., & Baron-Cohen, S. (2019). Social and non-social autism symptoms and trait domains are genetically dissociable. *Communications biology*, *2*, 328. https://doi.org/10.1038/s42003-019-0558-4

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in psychology*, *2*, 178. https://doi.org/10.3389/fpsyg.2011.00178

Wellman, H. M. (1990*). The Child's Theory of Mind*. Cambridge. 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. https://doi.org/10.1111/1467-8624.00304

Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: distinct roles for the social network and mirror system. *Psychological Science, 18*(6), 469–474. https://doi.org/10.1111/j.1467-9280.2007.01923.x

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., Weil, L., & Wakabayashi, A. (2006). Predicting Autism Spectrum Quotient (AQ) from the Systemizing

Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain research*, *1079*(1), 47–56. https://doi.org/10.1016/j.brainres.2006.01.012

White, T. L., & Depue, R. A. (1999). Differential association of traits of fear and anxiety with norepinephrine- and dark-induced pupil reactivity. *Journal of personality and social psychology*, *77*(4), 863–877. https://doi.org/10.1037//0022-3514.77.4.863

Wilcox, R. R., Rousselet, G. A., & Pernet, C. R. (2018). Improved methods for making inferences about multiple skipped correlations. *Journal of Statistical Computation and Simulation*, *88*(16), 3116–3131. https://doi.org/10.1080/00949655.2018.1501051

Wilkowski, B. M., Robinson, M. D., & Meier, B. P. (2006). Agreeableness and the prolonged spatial processing of antisocial and prosocial information. *Journal of Research in Personality*, *40*(6), 1152–1168. https://doi.org/10.1016/j.jrp.2005.12.004

Williams, L. M., Brown, K. J., Palmer, D., Liddell, B. J., Kemp, A. H., Olivieri, G., Peduto, A., & Gordon, E. (2006). The mellow years?: neural basis of improving emotional stability over age. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *26*(24), 6422–6430. https://doi.org/10.1523/JNEUROSCI.0022-06.2006

Wong, T. K., Fung, P. C., Chua, S. E., & McAlonan, G. M. (2008). Abnormal spatiotemporal processing of emotional facial expressions in childhood autism: dipole source analysis of event-related potentials. *The European journal of neuroscience*, *28*(2), 407–416. https://doi.org/10.1111/j.1460-9568.2008.06328.x

Wong, T. Y., Sid, A., Wensing, T., Eickhoff, S. B., Habel, U., Gur, R. C., & Nickl-Jockschat, T. (2019). Neural networks of aggression: ALE meta-analyses on trait and elicited aggression. *Brain Structure and Function*, *224*, 133-148. https://doi.org/10.1007/s00429-018-1765-3

Woodbury-Smith, M. R., & Volkmar, F. R. (2009). Asperger syndrome. *European child & adolescent psychiatry*, *18*(1), 2–11. https://doi.org/10.1007/s00787-008-0701-0

Wynn, J. K., Lee, J., Horan, W. P., & Green, M. F. (2008). Using event related potentials to explore stages of facial affect recognition deficits in schizophrenia. *Schizophrenia Bulletin*, *34*(4), 679–687. https://doi.org/10.1093/schbul/sbn047

Yang, H. H., Savostyanov, A. N., Tsai, A. C., & Liou, M. (2011). Face recognition in Asperger syndrome: a study on EEG spectral power changes. *Neuroscience letters*, *492*(2), 84–88. https://doi.org/10.1016/j.neulet.2011.01.06

Yarkoni, T. (2015). Neurobiological substrates of personality: A critical overview. In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.). *APA handbook of personality and social psychology, Vol. 4. Personality processes and individual differences* (pp. 61–83). American Psychological Association. https://doi.org/10.1037/14343-003

Yick, Y. Y., Buratto, L. G., and Schaefer, A. (2015). The effects of negative emotion on encoding-related neural activity predicting item and source recognition. *Neuropsychologia*, *73*, 48–59. https://doi. org/10.1016/j.neuropsychologia.2015.04.030

Young, L., Dodell-Feder, D., & Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia, 48*, 2658–2664. https://doi. org/10.1016/j.neuropsychologia.2010.05.012.

Ypma, R. J., Moseley, R. L., Holt, R. J., Rughooputh, N., Floris, D. L., Chura, L. R., Spencer, M. D., Baron-Cohen, S., Suckling, J., Bullmore, E. T., & Rubinov, M. (2016). Default Mode Hypoconnectivity Underlies a Sex-Related Autism Spectrum. *Biological psychiatry. Cognitive neuroscience and neuroimaging*, *1*(4), 364–371. https://doi.org/10.1016/j.bpsc.2016.04.006

Zhang, Y., Brady, M., Smith, S., (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Zhang, T., Sha, W.J., Zheng, X.R., Ouyang, H.L., & Li, H. (2009). Inhibiting one's own knowledge in false belief reasoning: An ERP study. *Neuroscience Letters*, *467*(3), 194–8. https://doi.org/10.1016/j.neulet.2009.10.009

Zhao, G., Ge, Y., Shen, B., Wei, X., & Wang, H. (2017). Emotion analysis for personality inference from EEG signals. *IEEE transactions on affective computing*, *9*(3), 362–371.

Zuckerman, M. (2005). *Psychobiology of Personality* (Second Edition, Revised and Updated). Cambridge University Press. https://doi.org/10.1017/CBO9780511813733

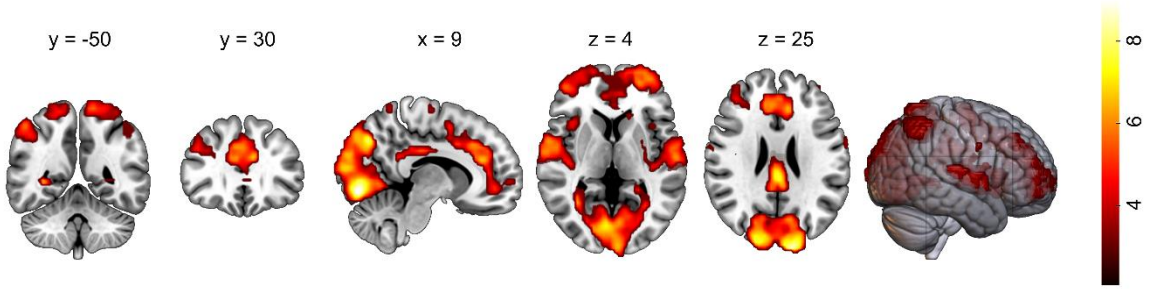
APPENDIX

Supplementary materials of Chapter 2

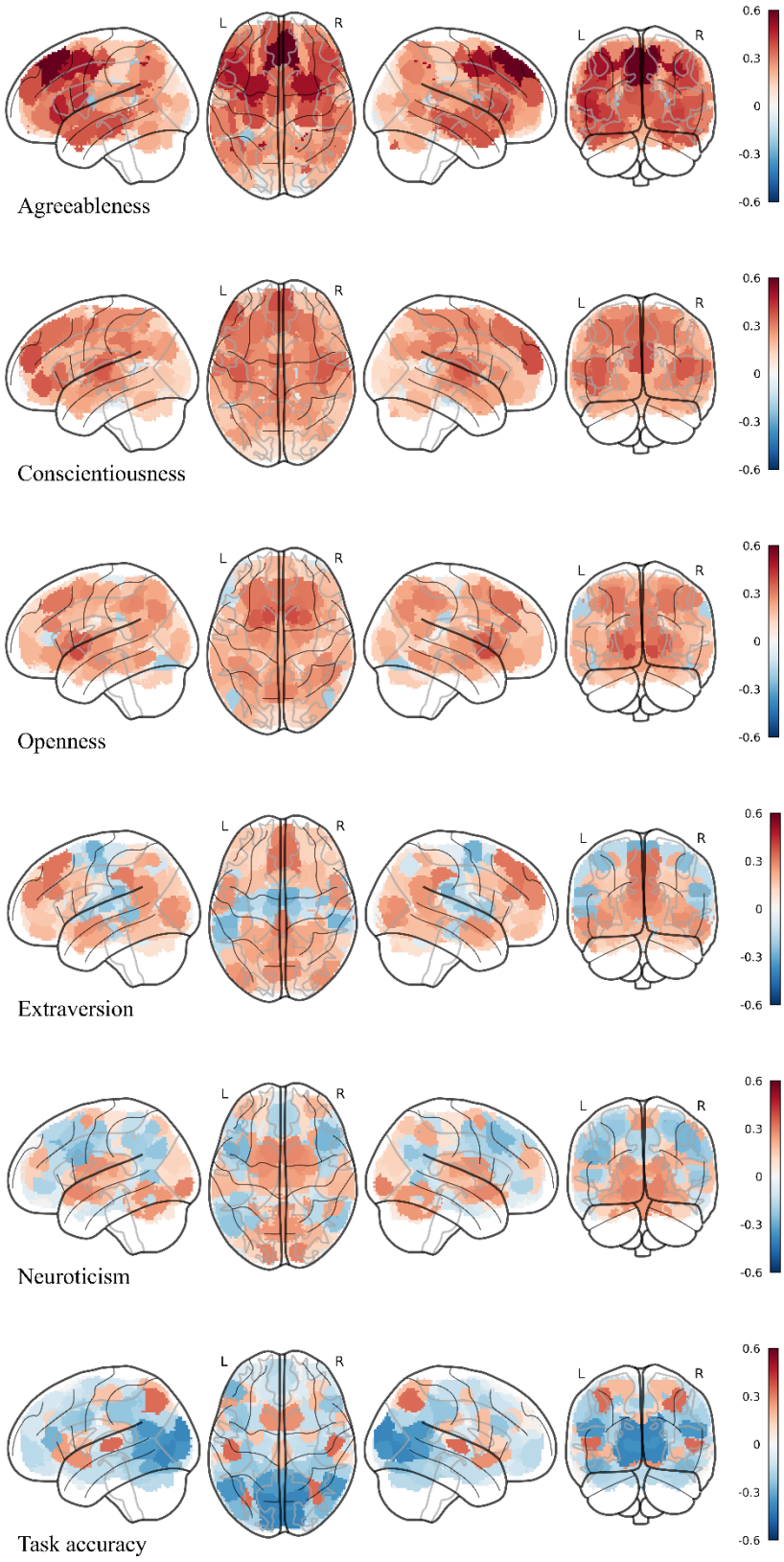
Supplementary table 2.1. Behavioral analyses results

Trait	Type of video	Accuracy[1]		Type of video × trait	RTs[2]		
		ρ	*p*		F	*η2*	*p*
Agreeableness	social	-0.056	0.704	main effect	0.71	0.017	0.404
	random	0.137	0.352	interaction	0.337	0.008	0.565
Conscientiousness	social	0.145	0.324	main effect	0.653	0.015	0.424
	random	0.141	0.341	interaction	0.001	0.000	0.988
Openness	social	0.009	0.948	main effect	2.979	0.066	0.091
	random	-0.01	0.946	interaction	0.516	0.012	0.476
Extraversion	social	-0.167	0.255	main effect	1.158	0.027	0.288
	random	-0.038	0.795	interaction	1.287	0.03	0.263
Neuroticism	social	0.11	0.422	main effect	1.225	0.028	0.275
	random	-0.066	0.654	interaction	0.316	0.007	0.577

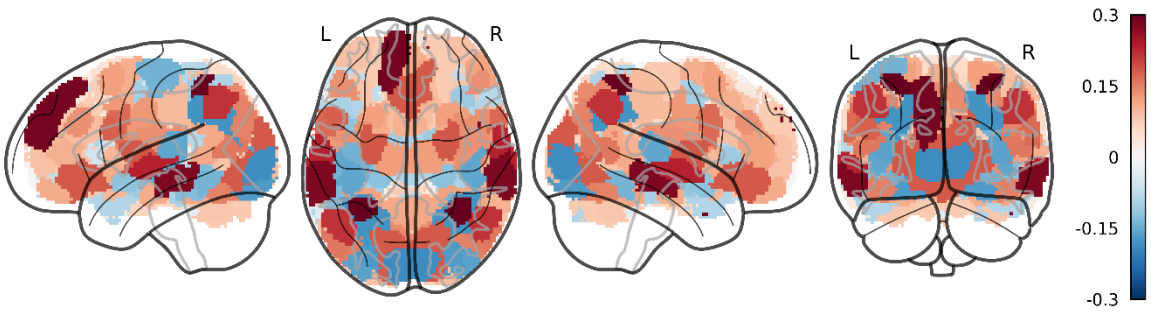
1. Spearman's rank correlation
2. Repeated measures ANCOVA



Supplementary figure 2.1. Random-social contrast activations. Brain activations from the whole brain voxelwise random - social content contrast. The statistical map was assessed with a cluster-based threshold of Z > 3.1, corrected at $p = 0.05$ (family-wise error correction) and projected onto a Montreal Neurological Institute (MNI) template. The color bar indicates Z values. The values above the slices indicate the coordinates in MNI space. Right is right.



Supplementary figure 2.2. Unthresholded correlational maps. Unthresholded results from the whole brain representational similarity analysis assessing the correlation between each personality score and task accuracy, and the dissimilarity in neural representations of social and random content. The color bar indicates the value of the Spearman's rho in each parcel.



Supplementary figure 2.3. Agreeableness correlation with social-random univariate contrast. Unthresholded results from the whole brain analysis assessing the correlation between the agreeableness scores and the univariate activation differences between social and random content. The color bar indicates the value of the Spearman's rho in each parcel.